# A Miniature Protein Stabilized by a Cation−π Interaction Network

Timothy W. Craven,[†,§] Min-Kyu Cho,[†] Nathaniel J. Traaseth,[†] Richard Bonneau,*[,§,‡,#] and Kent Kirshenbaum*[,†]

[†]Department of Chemistry, New York University, 100 Washington Square East, New York, New York 10003, United States

[§]Department of Biology, Center for Genomics and Systems Biology, New York University, 12 Waverly Place, New York, New York 10003, United States

[‡]Department of Computer Science, Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, United States

[#]Simons Center for Data Analysis, New York, New York 10010, United States

**S** *Supporting Information*

**ABSTRACT:** The design of folded miniature proteins is predicated on establishing noncovalent interactions that direct the self-assembly of discrete thermostable tertiary structures. In this work, we describe how a network of cation−π interactions present in proteins containing "WSXWS motifs" can be emulated to stabilize the core of a miniature protein. This 19-residue protein sequence recapitulates a set of interdigitated arginine and tryptophan residues that stabilize a distinctive β-strand:loop:PPII-helix topology. Validation of the compact fold determined by NMR was carried out by mutagenesis of the cation−π network and by comparison to the corresponding disulfide-bridged structure. These results support the involvement of a coordinated set of cation−π interactions that stabilize the tertiary structure.
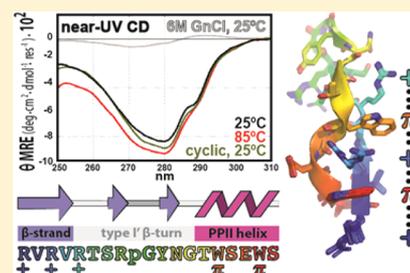
## INTRODUCTION

Miniature proteins afford an opportunity to isolate and evaluate noncovalent interactions that direct formation of the secondary and tertiary structures present in native proteins.[1−9] Previous studies of mini-proteins have elucidated thermodynamic and kinetic features of folding that may become convoluted within the context of more complex protein architectures. Within mini-proteins, the biophysical characteristics of particular interactions can therefore be elucidated more precisely. For example, the interactions between the guanidinium group of an arginine and the indole ring of tryptophan have now been extensively described.[10−19] These so-called cation−π interactions are representative of stabilizing noncovalent contacts between positively charged (Arg, Lys, His) and aromatic (Phe, Tyr, Trp) side-chain functionalities,[10] and provide distinctive contributions to protein structure and function.[11−15] Several studies have investigated the energetic contributions of binary cation−π interactions.[16−19] Their stabilizing strengths have been evaluated in the context of α-helices (−0.4 kcal/mol for residues positioned at $i$ and $i + 4$)[17] and β-sheets (−0.20 to −0.48 kcal/mol).[18,19] Computational studies have estimated that the free energy of cation−π interactions can vary up to −5.5 kcal/mol.[16] The strengths of cation−π interactions have been demonstrated to increase with temperature,[20] establishing thermoprotective influences in thermophilic organisms.[21]

Networks of cation−π interactions are evident in X-ray crystal structures of numerous proteins. One example is the "WSXWS motif" found in fibronectin type III (Fn3) domain containing proteins (Figure 1). The cation−π interactions in these motifs can be observed in X-ray structures stabilizing a relatively rare
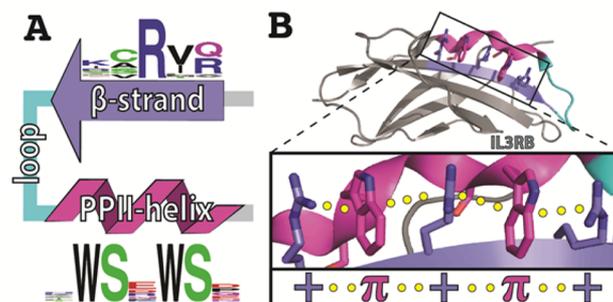


**Figure 1.** Cation−π interaction networks stabilize "WSXWS motifs" in native proteins. (A) Sequence logo[22] and cartoon representation of the conserved secondary structural elements observed in WSXWS motifs. (B) Cartoon representation of the Fibronectin type III domain of IL3RB in gray (PDBid: 1EGJ) with the WSXWS motif highlighted in blue (β-strand), cyan (loop), and purple (PPII-helix). (inset) The cation−π interaction network found in the X-ray structure of IL3RB. The dotted yellow lines highlight the connectivity of the cation−π interaction network formed between arginine (blue) and tryptophan (purple) side chains.

class of secondary structure, the left-handed poly proline type-II helix (PPII) (Figures 1, 2). In particular, an unusual PPII-helix devoid of proline residues has been observed in the sequences and structures of several Fn3 domains (Figures 1 and 2, Table 1). We sought to extract and replicate the β-strand:loop:PPII-helix
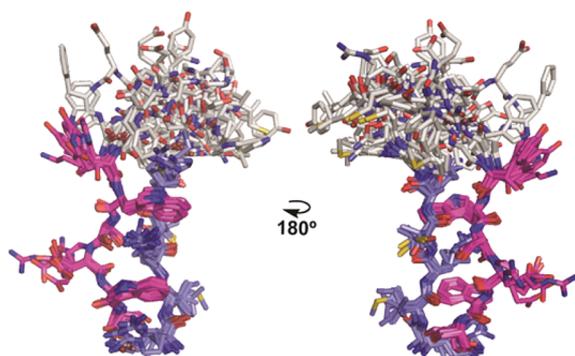
**Figure 2.** Structural alignment of 15 X-ray structures of WSXWS motifs.[23−37] β-strands, loops, and PPII-helical secondary structure elements can be seen in blue, gray, and purple respectively.

**Table 1. UniprotIDs and Multiple Sequence Alignment of WSXWS Motifs Found in X-ray Crystal Structures[23−37]**

| UniprotID | Multiple Sequence Alignment |
|---|---|
| P40189 | **R**IR**C**MKEDGKG....YW**S**D**WS**E |
| P14784 | QVR**V**K**PLQGEFT...TW**S**P**WS**Q |
| P31785 | **R**VR**S**RFNPLCGSAQ.HW**S**EW**S**H |
| P78552 | **R**IR**V**KTNKLCYEDDKLW**S**NW**S**Q |
| P40223 | QMR**C**IRSSLPG....FW**S**P**WS**P |
| P19235 | AVR**A**RMAEPSFGG..FW**S**EW**S**E |
| Q99062 | QIR**C**IRWPLPG....HW**S**D**WS**P |
| P42702 | **R**IR**C**STETFW.....KW**S**KW**S**N |
| P05710 | QTR**C**KPDHG......YW**S**RW**S**Q |
| Q14627 | VVR**S**KVNIYCSDDG.IW**S**EW**S**D |
| P08887 | QLR**A**QEEFGQG....EW**S**EW**S**P |
| Q01344 | QVR**A**AVSSMCREAG.LW**S**EW**S**Q |
| Q9HBE5 | QVR**A**GPMPGSSYQG.TW**S**EW**S**D |
| P48357 | QVR**C**KRLDGLG....YW**S**NW**S**N |
| P32927[(a)] | **R**VR**V**RTSRTGYNG..IW**S**EW**S**E |
| consensus | + + +       W**S** W**S** |

[a]The common name for P32927 is IL3RB. Note: Additional information about X-ray structures of WSXWS motifs including their PDB identifiers can be found in the Supporting Information (Figure S1).

topology observed in these motifs and determine whether this tertiary structure could exist in the absence of the parent Fn3 protein domain. This report details the successful design of a 19-residue mini-protein that emulates the tertiary structure of WSXWS motifs including a core network of cation−π interactions.

**WSXWS Motifs in Natural Proteins.** WSXWS motifs are found in the structures of proteins comprising an Fn3 domain and include a β-strand:loop:PPII topology (Figure 1) with a conserved "R...WSXWS" primary sequence (Table 1). The structures feature a conserved arginine side chain intercalated between the two tryptophan residues (R...WSXWS, Figure 3A). Additionally, the hydroxyl groups of the conserved serine residues (R...WSXWS) serve as hydrogen bond acceptors from an adjacent β-strand, as well as potential hydrogen bond donors to the i − 1 carbonyl oxygen (Figure 3B). The serine backbone amide HN groups also provide hydrogen bond donors to the adjacent β-strand. In some proteins, such as IL3RB (Figure 1, Table 1), the WSXWS motif is stabilized by a core network of cation−π interactions involving three arginine residues on a β-strand, which interdigitate with two tryptophan residues of an adjacent PPII-helix. The WSXWS motifs have been the focus of extensive biological studies, as they are found in a wide array of extracellular receptors. These motifs are also subject to
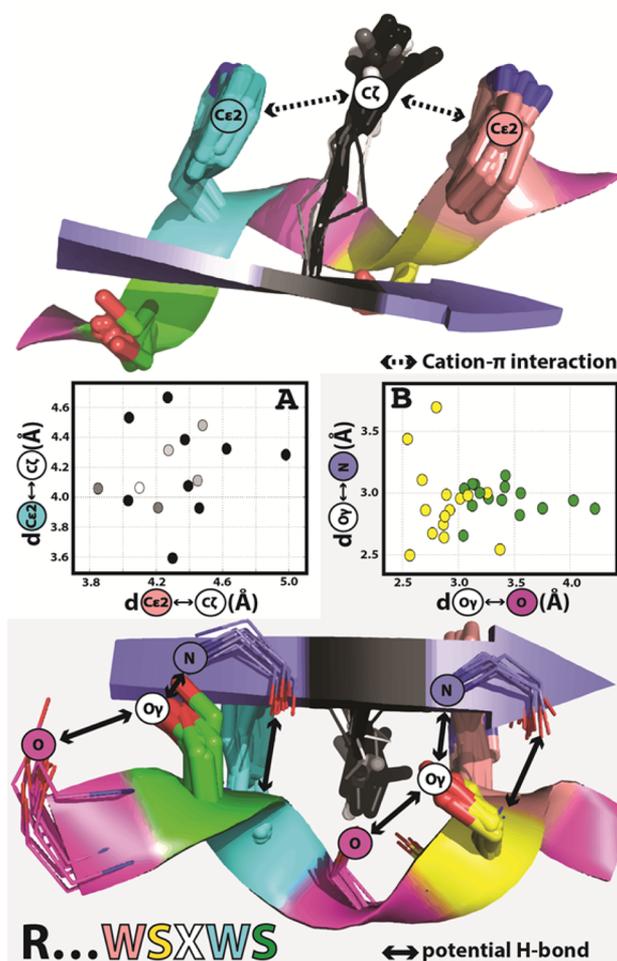


**Figure 3.** Characteristic cation−π interactions and hydrogen bonds in WSXWS motifs from proteins found in Table 1. (A) Plot of the distance between the Cε2 indole carbon atoms of each tryptophan to the Cζ of the central arginine guanidinium group highlighted with space-filled atoms in black → white. Different shades of black → white identify arginine rotamers having similar side-chain dihedral angles.[38] (Tables S5, S6). All values are <6.5 Å, a characteristic cutoff distance of cation−π interactions. (B) Plot of the distance between the serine side-chain Oγ atom and the i − 1 carbonyl oxygen (O) and the adjacent β-stand backbone amide nitrogen (N). Yellow and green circles correspond to the color-labeled serine residues found in the structural alignment. Black lines with arrows represent potential hydrogen bonds (H-bonds).

unique C-mannosylation of their conserved tryptophan residues (PDBid: 3TGX). In addition, a variety of genetic alterations within these sequences are associated with "X-linked severe combined immunodeficiency", (OMIMid: XSCID; a.k.a. "bubble boy syndrome") revealing the critical role played by the WSXWS motifs.

**PPII-Helices in Miniature Proteins.** Proline-free PPII-helices are difficult to design. Unlike other secondary structure types, the PPII-helix does not feature a propagating hydrogen bonded network. With three residues per helical turn, the cofacial i and i + 3 PPII positions are, on average, 8 Å apart (Cα → Cα), and are too distant to generate strong noncovalent interactions. Both the Trp-cage and the Avian Pancreatic Peptide (APP) mini-proteins[3,6] include a proline-rich PPII-helix packed against an α-helix generating an α-helix:loop:PPII topology. The structural rigidity engendered by multiple proline residues establishes an

underlying foundation for interfacing $\alpha$-helical elements. Thus, the rigidity of a PPII-helix has been demonstrated to augment the stability of proximal $\alpha$-helical secondary structures. Here we invert this relationship by stabilizing a proline-free PPII-helix through a set of stabilizing tertiary contacts with a neighboring peptide $\beta$-strand. This study presents a previously unexplored strategy for enforcing the left-handed PPII secondary structure type.

## ■ EXPERIMENTAL METHODS

**Peptide Synthesis.** All peptides were synthesized using standard FMOC solid-phase chemistry on 2-chlorotrityl (2-ClTrt) resin (Anaspec, Fremont, CA) using PyBop (benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate) activating agent (Anaspec, Fremont, CA). All peptides were precipitated out of cold ether and purified by reverse-phase HPLC using preparative chromatography and eluted with $H_2O$−Acetonitrile (ACN) gradients in 0.1% trifluoroacetic acid (TFA). After purification, all proteins were lyophilized three times in 1:1:1 (v:v:v) $H_2O$:ACN:Acetic Acid to remove TFA salts and transferred into buffer with pH adjustment after protein addition. Protein stock solutions at 2.0 mM concentrations were stored at −20 °C. All peptides were characterized by mass analysis using MALDI-TOF MS or ESI MS; sequence and purity were evaluated by HPLC and NMR spectroscopy analysis. Disulfide-bridged peptides were synthesized, purified, and characterized in an analogous way but instead were cleaved from 2-ClTrt resin using a 90:5:3:2 (v:v:v:v) mixture of TFA:thioanisole:ethanedithiol:anisole.

**Circular Dichroism (CD).** CD spectra were collected using an Aviv (Lakewood, NJ) stopped flow CD Spectropolarimeter Model 202SF. The far-UV (190−260 nm) CD spectra were obtained using a 1 mm path length cell and in 10 mM PBS buffer at a pH of 7.5 at 100 $\mu$M protein concentrations. The near-UV (245−320 nm) CD spectra were obtained using a 1 cm path length cell and in 10 mM PBS buffer at a pH of 7.5 at 50 $\mu$M concentrations. The consensus CD spectra were established as the Gaussian weighted average of 40 scans with evaluation of ellipticity every 0.5 nm, each having an averaging time of 0.5 s. Proper "blank" spectra were subtracted from all multiple wavelength and single-wavelength spectra and thermal melts utilizing the equation $(\Theta_{sample} - \Theta_{"blank"})/(L \cdot c \cdot n)$ where $L$ is the path length, $c$ is the protein concentration, and $n$ is the number of amino acids in the sequence. CD spectra are presented unsmoothed. Concentrations of protein stock solutions for conversion of millidegrees to mean residue ellipticity (MRE) were determined by UV absorbance at 280 nm using the standard extinction coefficients for the tryptophan (5690 $M^{-1}$ $cm^{-1}$), tyrosine (1280 $M^{-1}$ $cm^{-1}$), and disulfide-bridged cysteine (125 $M^{-1}$ $cm^{-1}$) chromophores in 6 M GnCl. Thermal melts were conducted using a 1 °C/min temperature gradient, a 1 min equilibration time at each temperature step, and an average of 20 single wavelength measurements. Spectra for GnCl concentrations that spanned the folded to unfolded transition are presented in Figure 4B and their MRE values at 280 nm were converted to $\Delta G°$ at 25 °C using the equation $\Delta G° = -RT \ln Q$ where $Q = \chi/1 - \chi$ and $\chi$ is the fraction of folded protein at the given concentration of GnCl. $\Delta G°$ values were plotted versus the concentration of GnCl and least-squares fit to a linear equation in order to extrapolate the y-intercept value corresponding to the $\Delta G°$ in native buffer at 25 °C.

**NMR Spectroscopy.** NMR spectra were collected using a Bruker (Billerica, MA) AVANCE III-600 NMR spectrometer equipped with a TCI CryoProbe. All samples were prepared in $H_2O$ and $D_2O$ that had been purged of dissolved $O_2$ gas using the freeze−pump−thaw method. 1D and 2D $^1H$ NMR were collected at protein concentrations of 1.8 mM in 90% $H_2O$:10% $D_2O$ solutions at pH 3.8 (uncorrected). For determination of $^3J_{HN-H\alpha}$ values, 1D $^1H$ NMR spectra were obtained with 32k complex data points. Total correlated spectroscopy (TOCSY) spectra were collected utilizing a mixing time of 120 ms using Bruker's mlevesgpph pulse sequence. 2D $^1H$−$^1H$ nuclear Overhauser effect spectroscopy (NOESY) spectra were collected utilizing a 250 ms mixing time using Bruker's noesyesgpph pulse sequence. All chemical shifts are reported relative to a 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) internal standard.

**NMR Structure Ensemble.** NMR assignment and volume calculations were performed using the Sparky NMR package. Using the Rosetta Molecular Modeling package,[39] 50 000 structures were initialized with $\varphi$ values to one of the two values dictated by an empirically derived Karplus equation $(6.51 \cdot \cos^2(\varphi - 60) - 1.76 \cdot \cos(\varphi - 60) + 1.635)$[40] for $^3J_{HN-H\alpha}$ values >6.4 Hz (Table S4). Glycine residues were initialized and randomized into the full range of $\varphi/\psi$ values. Distance constraints were calculated using the tryptophan indole reference cross-peak volume from W15H$\eta$2 → W15H$\zeta$3 as this has a known reference distance of 2.46 Å. Other indole ring proton cross-peak volumes with similar distances were also found close to this reference volume ($V_{ref}$). The distance from proton $i$ to $j$ was calculated using the equation $d_{i,j} = 2.46 \cdot (V_{ref}/V_{i,j})^{(1/6)}$. Upon the basis of these calculated distances, NOEs were characterized as strong (1.8−2.5 Å), medium (1.8−3.3 Å), or weak (1.8−5.0 Å). Structures were minimized simultaneously with respect to energy and distance constraints (with "soft" 2.5 Å tolerances) using an Armijo type line-search algorithm within a Monte Carlo Metropolis acceptance criterion algorithm. Randomized backbone $\varphi$ moves were limited to values within the bounds dictated by the Karplus equation ±30°. The 25 lowest energy sequences were found to obey all distance constraints within tolerances and were aligned using their C$\alpha$ backbone atoms. The structure was further supported by conducting identical calculations using the Xplor-NIH software package.[41,42] Using a standard Xplor-NIH simulated annealing minimization of 10 000 structures, sequences were found to converge to the same backbone structure determined by the Rosetta calculations (Figure S7).

**Sedimentation Equilibrium.** Sedimentation equilibrium measurements were carried out on a Beckman (Brea, CA) XL-A analytical ultracentrifuge equipped with absorption optics and an An-60 Ti rotor. Analysis was conduced at rotor speeds of 45 000 and 48 000 rpm at 4 °C. Data were acquired at two wavelengths per rotor speed setting and processed globally for the best fit to a single-species model of absorbance versus radial distance. Solvent density and protein partial specific volume were calculated according to solvent and protein composition, respectively.

**Fluorescence Spectroscopy.** Fluorescence spectra were obtained using a 1 cm by 2 mm path length cell (excitation path length versus emission path length) in 10 mM PBS buffer at a pH of 7.5 and 50 $\mu$M peptide concentrations. The fluorometer was configured with a calibrated Peltier temperature controller.

**Miscellaneous.** Least-squares fitting and graphical plotting was performed using the Scientific Python (scipy) scipy.optimize.curve_fit and matplotlib libraries. All 3-dimensional-cartoon protein graphics were created using the Schrodinger (New York, NY) PyMOL software package.

## ■ RESULTS AND DISCUSSION

**Design and Characterization of a Mini-Protein Stabilized by a Cation−$\pi$ Network.** The native sequence on which to initiate the mini-protein design was chosen based on two primary criteria. First, we sought a sequence containing a loop featuring a $\beta$-hairpin, which could be nucleated using a D-Pro-Gly motif.[2,7] Most of the loop regions observed in WSXWS motifs are lacking in canonical forms of secondary structure that could be enforced with rational peptide design (Figure 2). Second, we wanted a sequence containing three arginine residues on the $\beta$-strand portion of the fold composing the cation−$\pi$ interaction network. The only sequence meeting these two criteria was that of the WSXWS motif found in IL3RB (UniprotID: P32927). Thus, we began our investigation by synthesizing a modified 19-residue sequence from the WSXWS motif of IL3RB. The design included two modifications from the native sequence: D-Pro was incorporated at position 9 to promote a $\beta$-hairpin, and an Ile → Thr mutation was introduced at position 14 to enhance solubility. The resulting sequence, "RVRVRTSRpGYNGTWSEWS" is denoted as "TrpPlexus": a portmanteau of "tryptophan" and
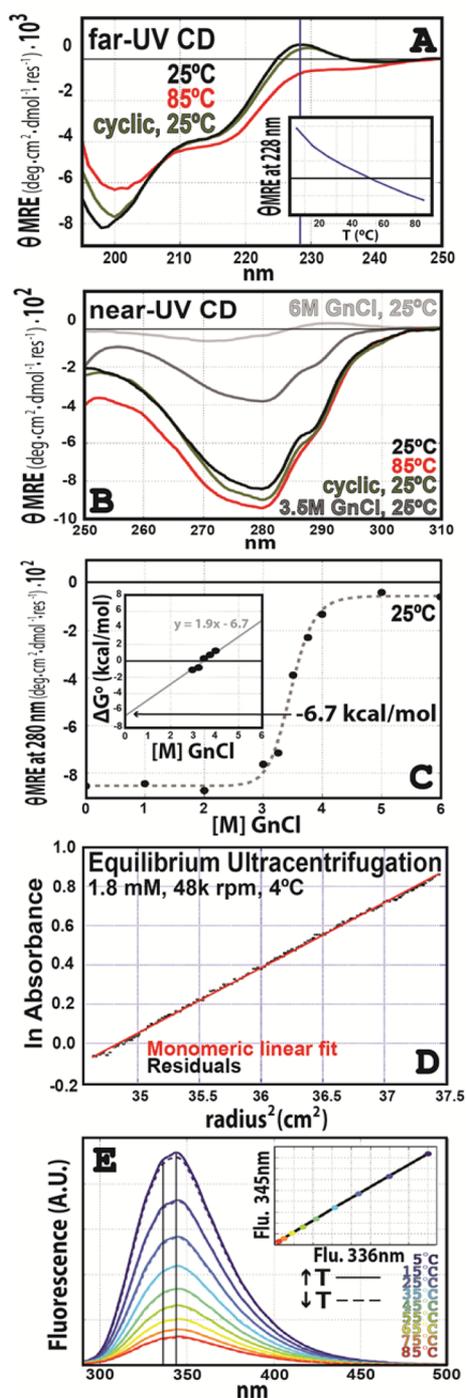
**Figure 4.** Spectroscopic evaluation of the TrpPlexus mini-protein. (A) Far-UV Circular Dichroism spectra (50 $\mu$M protein, 1 cm path length, 10 mM PBS buffer pH 7.5) in units of mean residue ellipticity (MRE). Inset shows the relationship of the positive lobe of ellipticity at 228 nm as a function of temperature. (B) Near-UV CD spectra (100 $\mu$M protein, 1 mm path length, 10 mM PBS buffer pH 7.5). (C) GnCl titrations (50 $\mu$M protein, 1 cm path length). (inset) MRE values converted to fraction folded at a given concentration of GnCl from a least-squares fit to a sigmoidal function. (D) Sedimentation equilibrium analysis. (E) Temperature dependent fluorescence spectra (excitation 280 nm) in Arbitrary Units (A.U.). Inset provides dual wavelength fluorescence intensity plot (336 nm vs 345 nm).

"plexus", signifying the network arrangement of stabilizing interactions (Figures 4 and 5).[43] TrpPlexus was found to be

monomeric at a concentration of 1.8 mM at 4 °C by sedimentation equilibrium analysis (Figure 4D).[44] The apparent molecular mass was within 10% of that calculated for an ideal monomer, with no systematic deviation of the residuals. The fluorescence emission spectrum (Figure 4E) for TrpPlexus was observed to have two major emission wavelengths of similar intensities, one at 336 and 345 nm. A two-wavelength emission intensity plot is provided (Figure 4E inset) as a proxy for the change in emission spectra shape as a function of temperature. A linear relationship between the intensities of the two emission maxima wavelengths was observed and the spectra were found to be completely reversible.

Secondary and tertiary structure content of the mini-protein was initially evaluated using circular dichroism (CD) spectroscopy (Figure 4A,B). In the near-UV (260−320 nm) CD spectrum, ellipticity can arise from indole and phenol absorbance bands when tryptophan or tyrosine side chains are present in an asymmetric environment. Strong CD signals in this region are exclusively observed for sequences containing significant tertiary structure.[45] Consequently, the near-UV CD signal is a stringent metric for the presence of tertiary structure. The strong near-UV CD signal for TrpPlexus was found to slightly increase in strength at elevated temperatures, indicating thermostability (Figure 4B, see discussion below). Conversely, the near-UV CD signal was observed to decrease in strength upon addition of the chaotropic agent guanidinium chloride (GnCl). The near-UV signal demonstrated two-state sigmoidal characteristics as a function of GnCl concentration. The variable GnCl concentration near-UV CD data correlate to a $\Delta G°$ for folding of approximately −6.7 kcal/mol in standard aqueous conditions at 25 °C (Figure 4C). The far-UV (180−260 nm) CD spectrum includes features indicating the presence of a PPII-helix, such as a lobe of positive ellipticity at 228 nm[46−48] that can be attributed to the $n \rightarrow \pi*$ transition of amide groups present in a PPII-helix, as well as minor contributions from aromatic side chains[49] (Figure 4A). The far-UV CD spectra also display similar temperature-dependent behavior observed in the spectra of other PPII-helical peptides.[50]

In order to obtain an atomic-resolution solution structure, nuclear magnetic resonance (NMR) experiments were carried out on TrpPlexus (Figure 5). Proton chemical shifts were assigned in the standard manner using $^1$H−$^1$H TOCSY experiments. Structural restraints to determine the three-dimensional structure were established using 2D $^1$H−$^1$H NOESY experiments. The assigned NOE cross-peaks were indicative of interdigitation of arginine and tryptophan residues and therefore complemented the near-UV CD data showing the presence of a highly ordered tertiary structure (Figure 4B). Using a set of unambiguous distance constraints from the NOE spectrum including 12 long-range restraints (Tables S1−S3), we determined the three-dimensional solution structure of TrpPlexus that depicts a cation−$\pi$ stabilized tertiary structure including a $\beta$-strand and a left-handed PPII-helix. The NMR ensemble structure is remarkably similar to the secondary and tertiary structure observed in the template WSXWS motif from the Fn3 domain of IL3RB with only minor differences. These small differences include alternate rotamers in the R8 and Y11 residues forming a more compact structure (Figure 5A), and small deviations of the backbone dihedral angles for the ensemble typically within 30° of the angles observed in the X-ray structure of IL3RB. The $\omega$ dihedral angle of D-Pro9 was determined to be in a predominantly *trans* amide conformation as evidenced by strong R8H$\alpha \rightarrow$ p9H$\delta$ NOE cross-peaks. Additionally, the presence of NOEs from G10HN $\rightarrow$ p9H$\delta$ supports the presence of a type I′ $\beta$-hairpin. In addition, the neighboring Asn-Gly (N12-G13)
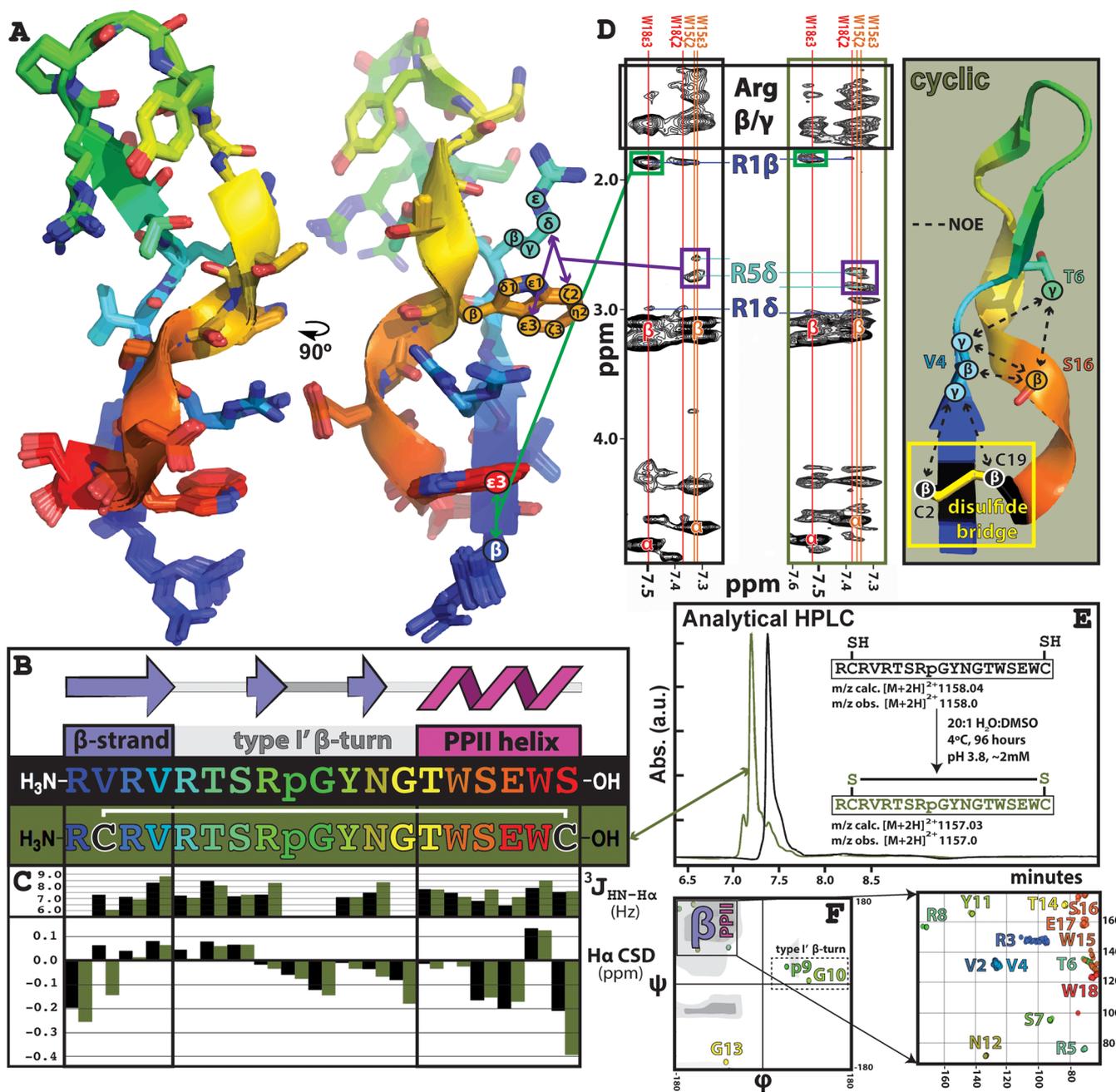
**Figure 5.** Structure determination of the TrpPlexus mini-protein. (A) NMR ensemble alignment of the 25 lowest energy structures satisfying distance constraints (Tables S1−S3). Highlighted regions with arrows show long-range NOE cross-peaks of Hε3 and Hζ2 of W15 and W18. Tryptophan indole ring and arginine side-chain atom descriptors are labeled on W15 and W18 as well as R1 and R5. (B) Cartoon of TrpPlexus secondary structure and color representations for each residue in panels A, D, and F. (C) $^3J_{HN-H\alpha}$ scalar couplings and Hα CSDs of the linear (black) and cyclic disulfide-bridged (green) sequences at 25 °C. (D) (left panels) NOESY spectra (250 ms mixing time, 1.8 mM protein, 10:90 (v:v) D$_2$O:H$_2$O, 4 °C) with highlighted long-range NOE interactions present in the spectra for both linear and cyclic structures. (right panel) Unambiguous NOEs supporting the model of the disulfide-bridged cyclic structure (Figure S9). (E) Analytical HPLC trace monitoring the cyclization of the disulfide-bridged sequence. (F) Ramachandran plot of the backbone dihedral angles of the NMR ensemble structures with labels corresponding to the φ and ψ angle domains of the secondary structures present. (inset) Highlights of the β-strand and PPII regions.

motif[1,7,9] serves to terminate the nascent β-strand, allowing for a transition to a left-handed PPII-helix. The patterning and relative intensities of the inter-residue and sequential NOE cross-peaks from both tryptophan indole groups match those expected for a PPII-helix displaying the most probable tryptophan rotamer.[38,51]

The resulting Hα chemical shift deviations (CSDs) from "random-coil" values were consistent in strength with those found in the CSDs of the WSXWS motifs in published NMR solution structures of whole Fn3 domains (PDBid(s): 2LFG,

1GCF, Figure S4). However, relative to CSDs of the Hα protons observed for other mini-proteins, those found for TrpPlexus are relatively small. To obtain CD spectra and CSD values corresponding to a fully "folded control" sequence, a cyclic version of TrpPlexus was designed with a disulfide-bridge (V2 → C, S19 → C). The disulfide-bond formation of the cyclic TrpPlexus was monitored by analytical HPLC and ESI mass spectrometry (Figure 5E) to identify a shift in the HPLC retention time and a reduction in mass for the cysteine sulfhydryl protons upon

disulfide formation. The linear cysteine-containing sequence was observed to cyclize into two conformations. The major conformer was isolated using preparative HPLC and analyzed by NMR and CD spectroscopy. The long-range NOEs seen in Figure 5D, the measured $^3J_{HN-H\alpha}$ scalar couplings, and Hα CSDs (Figure 5C) were found to be in good agreement with the linear peptide. Furthermore, the CD characterizations of linear and disulfide-bridged structures were found to be nearly identical in strength and overall spectral shape (Figure 4A,B). The similarity of both CD and NMR results strongly validates the structure and stability of the linear TrpPlexus as depicted in Figure 5A.

To explore the influence of cationic groups on fold stability, mutants were designed that replaced particular guanidinyl (arginine) side chain groups with either amino (lysine) cations or charge-neutral urea (L-citrulline) groups (Figure 6). The near-UV CD signal monitored at 280 nm provides a metric of relative tertiary structure content (Figure 6B). The absence of a strong near-UV CD signal for the L-citrulline triple mutant supports the conclusion that cation−π, and not stacking interactions or hydrogen bonding, is a primary determinant of TrpPlexus's folding. The near-UV CD signals, as a function of temperature, for TrpPlexus ("RRR") and the "KRK" mutant demonstrate weakly sigmoidal transitions to stronger ellipticities at elevated temperatures (Figure 6). While being an important indicator of tertiary structure, the near-UV CD signal as a function of temperature is also dependent on local environmental factors. It is possible that the weak transitions are the result of small changes in the distributions of tryptophan and cationic side-chain rotamers at elevated temperatures. For the other cationic group mutants "RKR" and "KKK", the observed transitions in the ellipticities can similarly be attributed to fluctuations within rotamer populations since the cation−π interactions were stronger at elevated temperatures.[20,21] These data support a model in which a structure containing a β-hairpin is able to situate cationic and tryptophan residues on two adjacent peptide strands in proximity to one another. However, in order to achieve a discrete tertiary structure as opposed to a molten globule, a central arginine is required to interdigitate the two tryptophan residues, resulting in a gain of tertiary structure that enhances near-UV ellipticity. Thus, the central arginine residue is the critical determinant of tertiary structure as evidenced by near-UV CD signal strength.

## CONCLUSION

We demonstrate that establishing a network of solvent-exposed cation−π interactions can be an effective strategy in the design of a novel miniature protein tertiary structure. We also validate a strategy in which solvent-exposed cation−π interactions serve as an alternative to hydrophobic-core packing as a protein design paradigm.[16] Excising polypeptide sequences from protein structures having the following characteristics could prove valuable in the development of additional well-ordered mini-proteins: (1) rich in arginine and tryptophan side-chain tertiary contacts as well as contacts between the side chains of other residues containing cations (lysine and histidine) and π electron systems (phenylalanine and tyrosine), (2) secondary structure content supporting or enabling those contacts such as a β-hairpin, which can be stabilized and nucleated via canonical peptide design, and (3) an evolutionary conservation or homology between similar sequences maintaining a putative cation−π network. A plethora of small sequence motifs stabilized by cation−π interactions can be found in the Protein Databank. This study demonstrates a successful workflow for miniprotein discovery initiated by bioinformatic sequence/structure analysis, followed by canonical peptide design stabilizing small secondary structure motifs, and validation via thorough structural characterization.

This report details a new miniature protein offering several distinctive features, which present new opportunities for future investigation in the context of an isolated well-folded system. These features include (1) an extensive cation−π interaction network, as opposed to a binary isolated interaction, (2) a novel miniature protein topology composed of a β-strand:loop:PPII-helix, (3) an isolated β-strand in a monomeric protein, which could offer opportunities for modulating β-sheet formation, and (4) a left-handed poly proline type-II helix formed in the absence of proline residues. The PPII-helix is stabilized through interdigitated tertiary contacts. This mode of PPII-helix stabilization in the context of a miniature protein could afford a new scaffold to emulate the *i* and *i* + 3 PPII-helical motif positions recognized by SH3 and WW protein domains. The continued engineering of miniature protein structures to contain β-strand,[52] PPII,[53,54] or poly arginine[55−57] motifs may prove advantageous for developing a next generation of peptide therapeutics and biochemical tools.



**Figure 6.** Near-UV CD thermal melts for the TrpPlexus cation−π interation network mutants. (A) Schematic of the cation−π network of TrpPlexus detailing variable amino acid positions 1, 3, and 5 in the mutant sequences. (B) Near-UV CD thermal melts for TrpPlexus ("RRR") and each mutant monitoring mean residue ellipticity at 280 nm (50 μM protein, 1 cm path length, 10 mM PBS buffer pH 7.5). Inset compares side-chain functionalities of the arginine (R), lysine (K), and L-citrulline (Cit) residues.

## ASSOCIATED CONTENT

**Ⓢ Supporting Information**

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b10285.

NMR structure coordinates. (PDB)

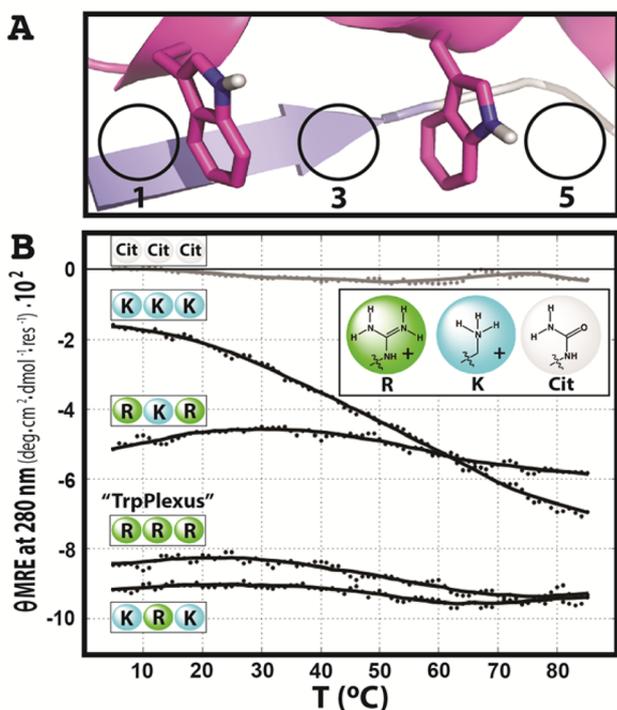Supplementary includes Tables S1−S11, Figures S1−S10. (PDF)

## ■ AUTHOR INFORMATION

**Corresponding Authors**

*rbonneau@nyu.edu
*kent@nyu.edu

**Notes**

The authors declare no competing financial interest.

## ■ REFERENCES

(1) Kortemme, T.; Ramirez-Alvarado, M.; Serrano, L. *Science* **1998**, *281*, 253−256.

(2) Struthers, M. D.; Cheng, R. P.; Imperiali, B. *Science* **1996**, *271*, 342−345.

(3) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425−430.

(4) Honda, S.; Yamasaki, K.; Sawada, Y.; Morii, H. *Structure* **2004**, *12*, 1507−1518.

(5) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 5578−5583.

(6) Hodges, A. M.; Schepartz, A. *J. Am. Chem. Soc.* **2007**, *129*, 11024−11025.

(7) Espinosa, J. F.; Gellman, S. H. *Angew. Chem., Int. Ed.* **2000**, *39*, 2330−2333.

(8) Ottesen, J. J.; Imperiali, B. *Nat. Struct. Biol.* **2001**, *8*, 535−539.

(9) Kier, B. L.; Andersen, N. H. *J. Am. Chem. Soc.* **2008**, *130*, 14675−14683.

(10) Dougherty, D. A. *Science* **1996**, *271*, 163−168.

(11) Gallivan, J. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9459−9464.

(12) Trotta, C. R.; Paushkin, S. V.; Patel, M.; Li, H.; Peltz, S. W. *Nature* **2006**, *441*, 375−377.

(13) Xiu, X.; Puskar, N. L.; Shanate, J. A. P.; Lester, H. A.; Dougherty, D. A. *Nature* **2009**, *458*, 534−537.

(14) Dagil, R.; Knudsen, M. J.; Olsen, J. G.; O'Shea, C.; Franzmann, M.; Goffin, V.; Teilum, K.; Breinholt, J.; Kragelund, B. B. *Structure* **2012**, *20*, 270−282.

(15) Hilton, D. J.; Watowich, S. S.; Katz, L.; Lodish, H. F. *J. Biol. Chem.* **1996**, *271*, 4699−4708.

(16) Gallivan, J. P.; Dougherty, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 870−874.

(17) Shi, Z.; Olson, C.; Kallenbach, N. *J. Am. Chem. Soc.* **2002**, *124*, 3284−3291.

(18) Tatko, C. D.; Waters, M. L. *Protein Sci.* **2003**, *12*, 2443−2452.

(19) Hughes, R. M.; Waters, M. L. *J. Am. Chem. Soc.* **2006**, *128*, 12735−12742.

(20) Prajapati, R. S.; Sirajuddin, M.; Durani, V.; Sreeramulu, S.; Varadarajan, R. *Biochemistry* **2006**, *45*, 15000−15010.

(21) Chakravarty, S.; Varadarajan, R. *Biochemistry* **2002**, *41*, 8152−8161.

(22) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. *Genome Res.* **2004**, *14*, 1188−1190.

(23) Bravo, J.; Staunton, D.; Heath, J.; Jones, E. *EMBO J.* **1998**, *17*, 1665−1674.

(24) Wang, X.; Rickert, M.; Garcia, K. *Science* **2005**, *310*, 1159−1163.

(25) Laporte, S. L.; Juo, Z. S.; Vaclavikova, J.; Colf, L. A.; Qi, X.; Heller, N. M.; Keegan, A. D.; Garcia, K. C. *Cell* **2008**, *132*, 259−272.

(26) Aritomi, M.; Kunishima, N.; Okamoto, T.; Kuroki, R.; Ota, Y.; Morikawa, K. *Nature* **1999**, *401*, 713−717.

(27) Syed, R. S.; Reid, S. W.; Li, C.; Cheetham, J. C.; Aoki, K. H.; Liu, B.; Zhan, H.; Osslund, T. D.; Chirino, A. J.; Zhang, J.; Finer-Moore, J.; Elliott, S.; Sitney, K.; Katz, B. A.; Matthews, D. J.; Wendoloski, J. J.; Egrie, J.; Stroud, R. M. *Nature* **1998**, *395*, 511−516.

(28) Tamada, T.; Honjo, E.; Maeda, Y.; Okamoto, T.; Ishibashi, M.; Tokunaga, M.; Kuroki, R. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 3135−3140.

(29) Skiniotis, G.; Lupardus, P.; Martick, M.; Walz, T.; Garcia, K. C. *Mol. Cell* **2008**, *31*, 737−748.

(30) Rossjohn, J.; McKinstry, W. J.; Woodcock, J. M.; McClure, B. J.; Hercus, T. R.; Parker, M. W.; Lopez, A. F.; Bagley, C. J. *Blood* **2000**, *95*, 2491−2498.

(31) Broutin, I.; Jomain, J.; Tallet, E.; Agthoven, J.; Raynal, B.; Hoos, S.; Kragelund, B. B.; Kelly, P. A.; Ducruix, A.; England, P.; Goffin, V. *J. Biol. Chem.* **2010**, *285*, 8422−8433.

(32) Lupardus, P. J.; Birnbaum, M.; Garcia, K. *Structure* **2010**, *18*, 332−342.

(33) Varghese, J. N.; Moritz, R. L.; Lou, M. Z.; Donkelaar, A.; Ji, H.; Ivancic, N.; Branson, K. M.; Hall, N. E.; Simpson, R. J. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 15959−15964.

(34) Huyton, T.; Zhang, J. G.; Luo, C. S.; Lou, M. Z.; Hilton, D. J.; Nicola, N. A.; Garrett, T. P. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 12737−12742.

(35) Patino, E.; Kotzsch, A.; Saremba, S.; Nickel, J.; Schmitz, W.; Sebald, W.; Mueller, T. D. *Structure* **2011**, *19*, 1864−1875.

(36) Hamming, O. J.; Kang, L.; Svensson, A.; Karlsen, J. L.; Rahbek-Nielsen, H.; Paludan, S. R.; Hjorth, S. A.; Bondensgaard, K.; Hartmann, R. *J. Biol. Chem.* **2012**, *287*, 9454−9460.

(37) Carpenter, B.; Hemsworth, G. R.; Wu, Z.; Maamra, M.; Strasburger, C. J.; Ross, R. J.; Artymiuk, P. J. *Structure* **2012**, *20*, 487−497.

(38) Dunbrack, R. L.; Karplus, M. A. *J. Mol. Biol.* **1993**, *230*, 543−574.

(39) Leaver-Fay, A.; et al. An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. In *Methods in Enzymology*; Johnson, M. L., Brand, L., Eds.; Computer Methods, Part C; Academic Press: New York, 2011; Vol. 487, pp 545−574.

(40) Vuister, G. W.; Bax, A. *J. Am. Chem. Soc.* **1993**, *115*, 7772−7777.

(41) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 66−74.

(42) Schwieters, C. D.; Kuszewski, J. J.; Clore, G. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 47−62.

(43) Within this report, the notation *xy* specifies a particular amino acid *x* at position number *y* in the 19 residue TrpPlexus sequence. Lower-case "p" denotes a D-Proline residue.

(44) Sedimentation equilibrium was conducted on the NMR sample used for solution structure determination.

(45) Kelly, S. M.; Price, N. C. *Curr. Protein Pept. Sci.* **2000**, *4*, 349−384.

(46) Horng, J. C.; Raines, R. T. *Protein Sci.* **2006**, *1*, 74−83.

(47) Chellgren, B. W.; Creamer, T. P. *Biochemistry* **2004**, *43*, 5864−5869.

(48) Krimm, S.; Tiffany, M. L. *Isr. J. Chem.* **1974**, *12*, 189−200.

(49) The lack of a significant exciton couplet in the far-UV CD spectra indicates that the indole rings of the tryptophan residues of TrpPlexus are not in a stacked or edge-to-face configuration.[5,9]

(50) Brown, A. M.; Zondlo, N. J. *Biochemistry* **2012**, *51*, 5041−5051.

(51) $p = 0.54$ for $\chi_1/\chi_2$ at $-66°/100°$ and $\varphi/\psi$ at $-90°/150°$. Strong W18Hδ1 → W18HN and W15Hδ1 → W15HN NOE cross-peaks were observed indicating the presence of this rotamer.

(52) Zheng, J.; Baghkhanian, A. M.; Nowick, J. S. *J. Am. Chem. Soc.* **2013**, *135*, 6846−6852.

(53) Nguyen, J. T.; Turck, C. W.; Cohen, F. E.; Zuckermann, R. N.; Lim, W. A. *Science* **1998**, *282*, 2088−2092.

(54) Bunagan, M. R.; Yang, X.; Saven, J. G.; Gai, F. *J. Phys. Chem. B* **2006**, *110*, 3759−3763.

(55) Appelbaum, J. M.; LaRochelle, J. R.; Smith, B. A.; Balkin, D. M.; Holub, J. M.; Schephartz, A. *Chem. Biol.* **2012**, *19*, 819−830.

(56) Smith, B.; Daniels, D. S.; Coplin, A.; Jordan, G.; McGregor, L.; Schepartz, A. *J. Am. Chem. Soc.* **2008**, *130*, 2948−2949.

(57) Li, M.; Tao, Y.; Shu, Y.; LaRochelle, J. R.; Steinauer, A.; Thompson, D.; Schepartz, A.; Chen, Z.-Y.; Liu, D. *J. Am. Chem. Soc.* **2015**, *137*, 14084−14093.