DOI: 10.1002/anie.200805476

Protein Design

# De Novo Design of a βαβ Motif**

*Huanhuan Liang, Hao Chen, Keqiang Fan, Ping Wei, Xianrong Guo, Changwen Jin, Chen Zeng, Chao Tang, and Luhua Lai**

The design of molecules with a specific structure and function is a long-term goal. The de novo design of proteins can not only shed light on the process of protein folding, but can also generate potentially functional proteins.[1] Much knowledge has been accumulated for secondary-structure design,[2] but the de novo design of stable tertiary structures remains challenging.[3] A βαβ motif, which consists of two parallel β strands connected by an α helix, was chosen as our design target. This motif, like helix bundles and ββα motifs, is a versatile supersecondary structure in proteins. Natural α/β proteins contain continual βαβαβα structures; however, a stand-alone βαβ motif had never been observed. Derreumaux and co-workers had tried to design βαβαβ and βαβ folds, but failed to obtain stable structures.[4] Herein, we present the successful de novo design of a βαβ motif with only coded amino acids.

After a statistical analysis of the helix length in natural α/β proteins,[5] a length of 12 residues was chosen for the central helix in the designed βαβ motif. This length corresponds to the length of a five-residue β strand. The initial model was constructed according to known rules.[6] Standard secondary-structure geometrical restrictions were used to build the backbone structure. Binary patterns and the secondary-structure preferences of amino acids were considered for the sequence design. An amphipathic helix was designed with leucine and alanine on the hydrophobic face. On the basis of considerations of possible electrostatic interactions between side chains, glutamic acid and lysine were arranged alternately on adjacent helical turns. An Ncap motif (the helix-boundary motif at the N terminus) was chosen to stabilize the helix according to a statistical survey of α/β proteins.[10] For the parallel β sheets, isoleucine and valine residues were used to form a hydrophobic core, with leucine residues on the helix. The hydrophilic amino acids threonine and arginine were chosen for the exterior. Following the rational design of this structure, an automated program was used to rebuild the hydrophobic core.[5] Nine positions were selected for fixed-backbone sequence redesign with a backbone-dependent rotamer library. The designed peptide was expressed in *Escherichia coli* as a GST-fusion protein (GST = glutathione S-transferase) and purified on a GST-affinity column and then by reversed-phase HPLC. The proteins at this stage of the design process showed remarkable secondary structures in circular dichroism (CD) spectra, but were in molten globule states and aggregated significantly in solution (data not shown).

To obtain a stable monomeric βαβ motif, we then took two special measures. In comparison with the antiparallel β hairpin, which could be restricted by a tight turn, it is much harder to drive the parallel β sheets together, as they are connected by a longer sequence. We hypothesized that strong interactions, as found in βαβ motifs in natural proteins, should be designed between the β sheets. Tryptophan zippers (WW interactions) have proved very effective in stabilizing the tertiary structure of β hairpins.[7] We therefore tried to import a pair of tryptophan residues (W9/W34) into the parallel β sheets. On the other hand, nonspecific protein aggregation is intractable in de novo design. Negative design approaches have been used to solve this problem. Wang et al. introduced a lysine residue on the nonpolar face of β strands to make amyloid-like fibrils change into a monomeric β sheet.[8] Similarly, we designed two positively charged residues into the βαβ motif: K21 on the helix and R6 on the first strand. These residues were expected to cover the hydrophobic core with their long hydrophobic side chains to prevent aggregation by the terminal positive charges. The final sequence, named DS119, was obtained after an iterative design process (Table 1).

The CD spectra of DS119 showed good secondary structures that were not concentration-dependent (Figure 1). A distinct positive maximum was observed at 190 nm, and a broad negative maximum was observed between 208 and 222 nm. These values are typical for α/β

[*] Prof. L. Lai
State Key Laboratory for Structural Chemistry of Unstable and Stable Species, BNLMS, College of Chemistry and Molecular Engineering and Center for Theoretical Biology, Peking University
Beijing 100871 (China)
Fax: (+86) 10-6275-1725
E-mail: lhlai@pku.edu.cn
Homepage: http://mdl.ipc.pku.edu.cn
H. Liang, Dr. K. Fan, Dr. P. Wei
College of Chemistry and Molecular Engineering, Peking University
Beijing 100871 (China)
Dr. X. Guo, Prof. C. Jin
Beijing Nuclear Magnetic Resonance Center, Beijing 100871 (China)
Dr. H. Chen, Prof. C. Zeng
Department of Physics, The George Washington University
Washington, DC 20052 (USA)
Prof. C. Tang
Departments of Biopharmaceutical Sciences and Biochemistry and Biophysics
University of California, San Francisco, CA 94158 (USA)

[**] This research was supported, in part, by grants from the National Natural Science Foundation of China (No. 20640120446, No. 20228306, No. 90103029, and No. 10721403), the Ministry of Science and Technology of China, the Ministry of Education of China, and the US National Science Foundation (NSF) (DMR0313129; C.Z. and C.T.).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200805476.

  

***Table 1:*** Sequences of the designed protein and mutants.[a]

| | |
|---|---|
| DS119 | GSGQV RTIWV GGTPE ELKKL KEEAK KANIR VTFWG D |
| W9T | GSGQV RTITV GGTPE ELKKL KEEAK KANIR VTFWG D |
| W34T | GSGQV RTIWV GGTPE ELKKL KEEAK KANIR VTFTG D |
| W2T | GSGQV RTITV GGTPE ELKKL KEEAK KANIR VTFTG D |
| K21L | GSGQV RTIWV GGTPE ELKKL LEEAK KANIR VTFWG D |
| VRRV | GSGQR VTIWV GGTPE ELKKL KEEAK KANIR VTFWG D |

[a] Residues shown in light blue at the N terminus were introduced as a thrombin-cleavage site. Residues 4–36 are the real de novo designed sequences. Mutations are shown in red.

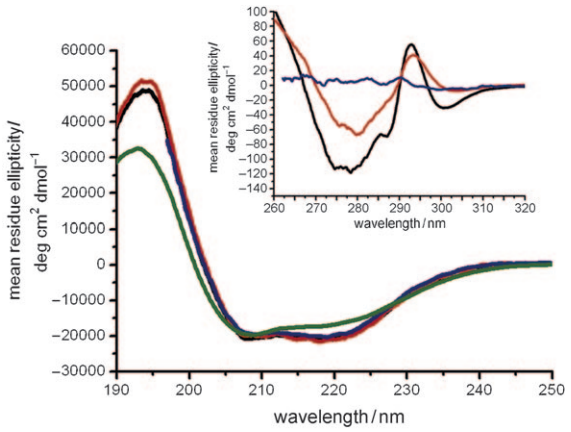


***Figure 1.*** CD spectra of DS119. Experiments were carried out at 25 °C, pH 7.3 in 20 mM phosphate buffer at different concentrations: 2 mg mL$^{-1}$ (——), 0.2 mg mL$^{-1}$ (——), and 0.02 mg mL$^{-1}$ (——). DS119 shows no concentration dependence and retains most of its secondary structure at 90 °C (——). The WW interaction was retained at 90 °C, as evident from the CD spectrum in the near-UV region (inset: 25 °C (——) and 90 °C (——); DS119 denatured with 4 M Gdn-HCl was used as a control (——).

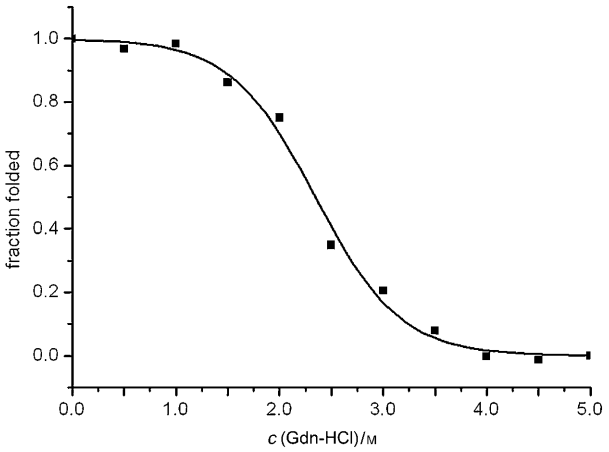


***Figure 2.*** Chemical denaturation of DS119. The mean CD residue ellipticity at 220 nm was monitored as a function of the concentration of Gdn-HCl.

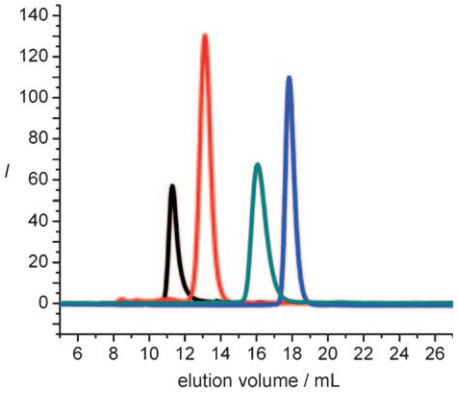


***Figure 3.*** Gel-filtration analysis of DS119. A prepacked superdex peptide 10/300 GL high-performance column was used to analyze the aggregation state of DS119 (4028 Da, ——). The molecular markers are cytochrome C (12.4 kDa, ——), aprotinin (6512 Da, ——), and vitamin $B_{12}$ (1355 Da, ——). The vertical axis is the UV intensity detected at 220 nm and the horizontal axis is the elution volume.

folds. The WW interaction could be detected in the CD spectra in the near-UV region. DS119 was found to be highly thermally stable and could not be denatured completely at 90 °C. This high thermal stability is exceptional for small proteins without disulfide bonds and unusual amino acids. Chemical denaturation was carried out by monitoring the mean residue ellipticity at 220 nm as a function of the concentration of guanidine hydrochloride (Gdn-HCl; Figure 2). The designed small protein underwent a typical two-state cooperative unfolding process. It started to denature from a Gdn-HCl concentration of about 1 M and became completely denatured at a Gdn-HCl concentration of about 4 M with a middle point at 2.5 M. The typical S-shaped steep transition curve expected for a monomeric single-domain protein was observed.

DS119 was also shown to be a monomer in an analytical gel-filtration experiment (Figure 3). The solution structure of the de novo designed protein was solved by homonuclear 2D $^1$H NMR spectroscopy. Double-quantum-filtered COSY (DQF-COSY), TOCSY, NOESY, and $^{13}$C HSQC (heteronuclear single quantum coherence) spectra were acquired to aid in signal assignment and structure determination. The secondary structures were assigned primarily from the chemical shifts of the $H_\alpha$ atoms. Unambiguous distance restraints derived from NOE signals indicated that residues 15–26 adopt an α-helical conformation, as demonstrated by the short-range $d_{NN(i,i+1)}$, medium-range $d_{NN(i,i+2)}$, $d_{\alpha N(i,i+3)}$, $d_{\alpha N(i,i+4)}$, and side chain–side chain $(i,i+3)$ NOEs. The strong cross-peaks produced by adjacent residues 6–10 and 30–34 (distance restraints $d_{\alpha N(i,i+1)} < 2.2$ Å) represent a parallel β sheet. Representative long-range NOEs indicated the presence of a hydrophobic core that was consistent with the desired tertiary structure.[5] The first 20 lowest-energy structures could be superimposed well with a backbone root-mean-square deviation (RMSD) of 0.46 Å for the secondary-structure region (Figure 4a).

The formation of the designed hydrophobic core was confirmed by NMR spectroscopy: Long-range NOE cross-peaks between the phenyl ring of F33 and other hydrophobic side chains, including I8, V10, L17, L20, I29, and V31, were well dispersed and could be assigned unambiguously. All side chains of the hydrophobic residues were confirmed by $^{13}$C HSQC. Cross-peaks between the core residues were

  

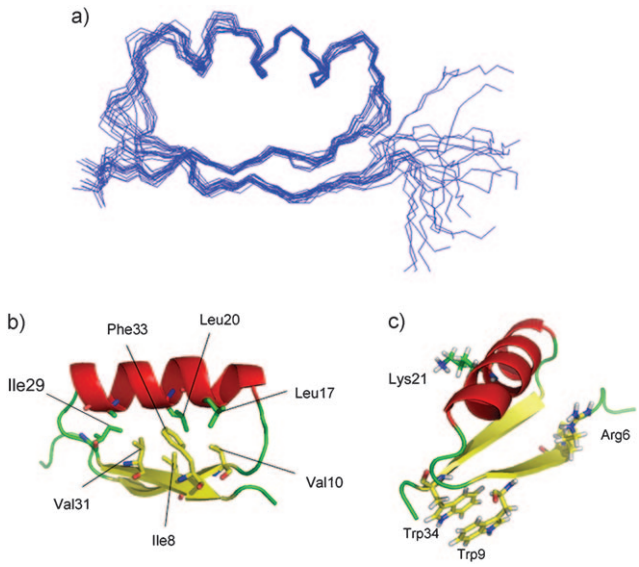


***Figure 4.*** NMR structures of DS119.[9] a) The best 20 structures obtained from structure calculations by NMR spectroscopy. The structures were superimposed by fitting the secondary-structure regions (residues 6–10, 15–26, 30–34). b) The average structure of the top 20 structures is shown with the hydrophobic core. The phenyl ring of F33 is locked by the side chains of I8, V10, L17, L20, and V31. c) Special design features in DS119. The interaction between W9 and W34, and the side chains of R6 and K21 are shown.

used to define the structure. The phenyl ring of F33 was located in the center of the hydrophobic core and interacted with the side chains of I8, V10, L17, L20, and V31 (Figure 4b).

From the structure established by NMR spectroscopy, we could also see that the two indole rings of W9 and W34 packed together. In contrast to the large upfield shift of the signals for the aromatic hydrogen atoms in the edge-to-face interaction observed in the Trp-zipper β-hairpin structure,[7] no upfield shifting of the resonances of the Trp aromatic hydrogen atoms was found for DS119. The NOE restraints favor a possible face-to-face interactions.[5] The positively charged residues, R6 and K21, covered the hydrophobic core as expected (Figure 4c). Several structures with mutations at these four positions were constructed to check the functions of these residues in DS119 (Table 1). All mutations caused the protein to aggregate in a gel-filtration study.[5] There may be different reasons for the importance of these four residues: The WW interaction may assist the formation of parallel β sheets by locking the two strands together and avoiding hydrophobic exposure. It may function as a stabilizer for the monomeric structure. R6 and K21 were used for the purpose of negative design. The positive charges on their side chains played a key role in preventing aggregation.

To confirm the novelty of the designed sequence, we conducted a BLAST (basic local alignment search tool) search of all known protein sequences. The designed sequence has very low similarity to any known natural protein sequences. We also compared the NMR structure with known βαβ motifs in large proteins in the PDB. The NMR structure does resemble many known βαβ motifs. The most similar structure is a βαβ motif from molybdopterin-biosynthesis MOEB protein (PDB code: 1WB, residues 122–154 in the B chain) with a backbone RMSD of 1.52 Å.[5]

In conclusion, a stand-alone βαβ motif was de novo designed with a stable monomeric tertiary structure and only coded amino acids. A tryptophan zipper on the parallel β sheets to stabilize the tertiary structure and prevent aggregation by locking the two β strands in place was crucial. No de novo designed stable structure with parallel β sheets has been reported previously. The designed small protein may provide a model system for a protein-folding study. As the designed protein is monomeric and highly thermally stable, the central helix might be modified further for a functional purpose, such as the inhibition of protein–protein interactions.

Received: November 10, 2008
Revised: January 7, 2009
Published online: April 3, 2009

**Keywords:** foldamers · peptides · protein design · tertiary structure · WW interactions

[1] a) J. Kaplan, W. F. DeGrado, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 11566; b) X. I. Ambroggio, B. Kuhlman, *Curr. Opin. Struct. Biol.* **2006**, *16*, 525; c) S. M. Butterfield, W. J. Cooper, M. L. Waters, *J. Am. Chem. Soc.* **2005**, *127*, 24; d) O. Iranzo, C. Cabello, V. L. Pecoraro, *Angew. Chem.* **2007**, *119*, 6808; *Angew. Chem. Int. Ed.* **2007**, *46*, 6688.

[2] a) J. Venkatraman, S. C. Shankaramma, P. Balaram, *Chem. Rev.* **2001**, *101*, 3131; b) L. Baltzer, H. Nilsson, J. Nilsson, *Chem. Rev.* **2001**, *101*, 3153; c) R. M. Hughes, M. L. Waters, *Curr. Opin. Struct. Biol.* **2006**, *16*, 514; d) R. B. Hill, D. P. Raleigh, A. Lombardi, W. F. DeGrado, *Acc. Chem. Res.* **2000**, *33*, 745.

[3] a) B. I. Dahiyat, A. Heitz, L. Chiche, P. Derreumaux, *Science* **1997**, *278*, 82; b) B. Kuhlman, *Science* **2003**, *302*, 1364.

[4] M. Coinçon, A. Heitz, L. Chiche, P. Derreumaux, *Proteins Struct. Funct. Genet.* **2005**, *60*, 740.

[5] See the Supporting Information.

[6] a) IUPAC, *Biochemistry* **1970**, *9*, 3471; b) P. Y. Chou, G. D. Fasman, *Annu. Rev. Biochem.* **1978**, *47*, 251; c) K. T. O'Neil, W. F. DeGrado, *Science* **1990**, *250*, 646; d) C. Toniolo, *Macromolecules* **1978**, *11*, 437; e) H. Xiong, B. L. Buckwalter, H. M. Shieh, M. H. Hecht, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6349; f) S. Marqusee, R. L. Baldwin, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8898; g) R. Aurora, G. D. Rose, *Protein Sci.* **1998**, *7*, 21.

[7] a) A. G. Cochran, N. J. Skelton, M. A. Starovasnik, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5578; b) M. Dhanasekaran, O. Prakash, Y. X. Gong, P. W. Baures, *Org. Biomol. Chem.* **2004**, *2*, 2071; c) W. W. Streicher, G. I. Makhatadze, *J. Am. Chem. Soc.* **2006**, *128*, 30.

[8] a) J. S. Richardson, D. C. Richardson, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2754; b) W. Wang, M. H. Hecht, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 2760.

[9] a) Figure 4a was created with MOLMOL: R. Koradi, M. Billeter, K. Wuthrich, *J. Mol. Graph.* **1996**, *14*, 51; the structures in Figure 4b,c were created with PyMOL: b) http://pymol.sourceforge.net/.

[10] Y. Qi et al., unpublished results.

  

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

# Supporting Information



69451 Weinheim, Germany

**Supporting Information**

# De Novo Design of a βαβ-Motif

Huanhuan Liang, Hao Chen, Keqiang Fan, Ping Wei, Xianrong Guo, Changwen Jin, Chen Zeng, Chao Tang, Luhua Lai*

## *Protein design Section*

### Rational design based on known rules

The βαβ-Motif was chosen as the design target. In order to build an initial model, a survey of sequence lengths of the α-helix and the β-strands in βαβ-motifs was done by analyzing α/β protein structures in PDB, as the secondary structures were length-dependent.[1, 2] Most regular helices in the βαβ-motif were found to contain 9-13 residues (Figure S1). So the central helix was chosen to be 12 residues in length, corresponding to a 5 residue β-strand in our design target. The backbone structure was built according to the standard α-helix and β-sheet geometrical restrictions. Known rules in *de novo* design was thoroughly taken into account. Binary patterns and the secondary structure preferences of amino acids were considered most important in our design process. Several amino acids were favorite for secondary structure design: Leu, Ala, Glu, Lys, Val, Ile, Thr and Arg. An amphipathic helix was designed using leucine and alanine on the hydrophobic face. Considering possible electrostatic interactions between side chains, glutamic acid and lysine were arranged alternately on adjacent helix turns. The helix Ncap motif in α/β proteins was statistical studied, and several candidates were tested to stabilize the helix, such as DPEE, NPEE and TPEE. TPEE was found to be the best based on experiment results. As successful parallel β-sheet had never been reported before, we tried to design it based on rules came from nature proteins and β-hairpin model systems. Isoleucine and valine were arranged on the hydrophobic face to form a hydrophobic core with leucines on the helix. The hydrophilic threonine and arginine were chosen for the exterior.

After the initial design, an automated program was used to rebuild the hydrophobic core. Nine core positions were selected for a fixed-backbone sequence redesign using a backbone-dependent rotamer library. F33 was chosen as the central residue in hydrophobic core. The calculation process was described in the follow section.

After this automatic design process, we got designed sequences with typical secondary structures, but all of them were in molten globule states and aggregated non-specifically in solution. No tertiary structure was obtained. We thought that this may come from the flexibility of the two β-strands as they are connected by a long helix compared to the tight turns in β-harpins. In order to drive the two β-strands together, we introduced the WW interaction on the two strands (W9 and W34). Stable tertiary structure was got at this step, but this protein existed as a dimer. We then tried to get a monomer by doing negative design, and finally, DS119 was obtained. An illustration of the design process was shown in Figure S2.

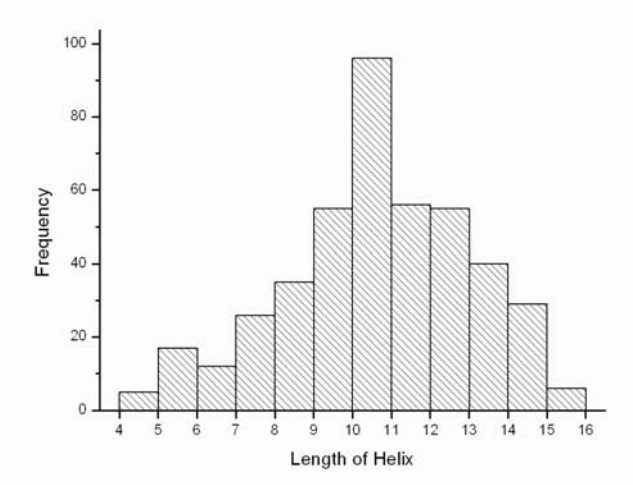


**Figure S1** Statistical analysis on the secondary structure lengths in known βαβ motifs in the PDB.

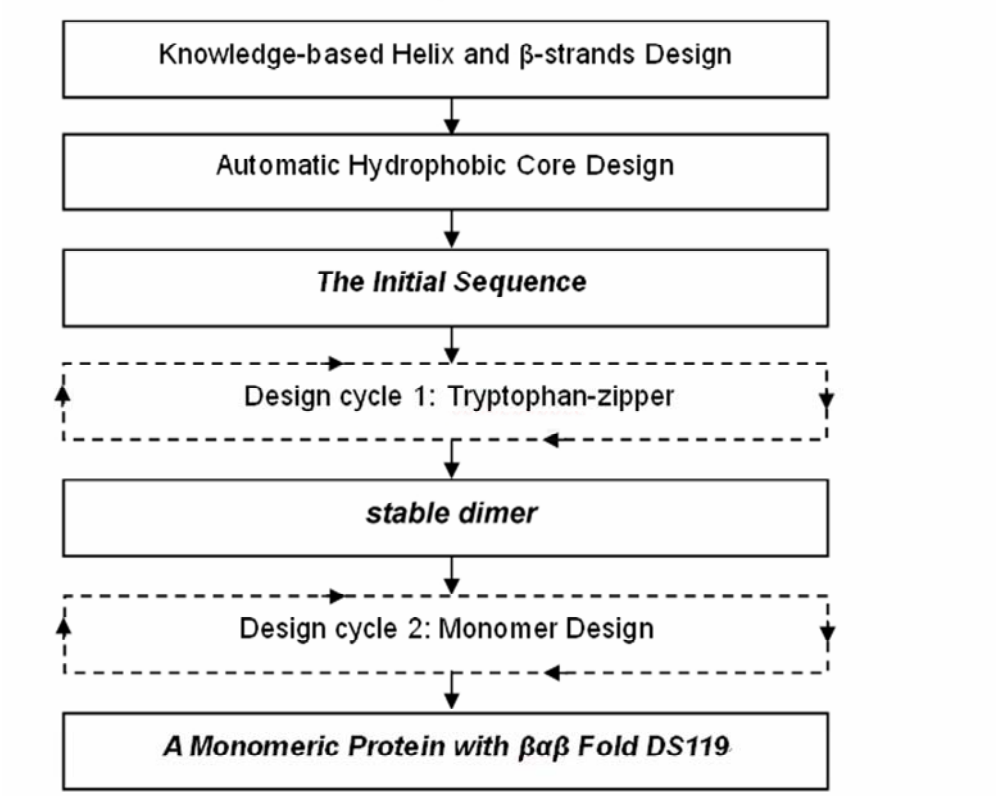


**Figure 2** The *de novo* design process. Helix and β-strands were designed first based on known rules, which were then combined to form the initial sequence by an automatic hydrophobic core design program. The design-test-design cycles were performed to optimize the sequence. In cycle 1, a pair of tryptophan-zipper was introduced at the C-terminal of the β-sheet. They were proved to be useful in stabilizing the tertiary structures. A negative design strategy was then used to prevent aggregation. Finally, a desired sequence that could fold into stable monomeric βαβ fold was obtained.

## Core Design Using the Dead-End-Elimination (DEE) Algorithm

### *Classification of core/boundary/surface residue position*

To reduce the size of the search space, it is natural to classify the backbone positions into core and surface sites. To do this we follow the scheme of S. L. Mayo *at al*[3], and use as our generic sidechain three spheres of radius $r_\beta$. If the distance from $C_\alpha$ to $C_\beta$ is $r_\beta$, then these three spheres are placed in the direction from $C_\alpha$ to $C_\beta$ at $r_\beta$ (the $C_\beta$ position), 2r, and 3r. We have taken r=1.53Å, and $r_\beta$=1.95Å. To calculate solvent accessible surface area, we have used a water add-on radius of 1.40Å, and the surface area is computed by placing uniformly 256 dots on each sphere.

The solvent accessible surface area for each residue is then calculated. After some experimentation, we have used a reasonable core cutoff at 24Å$^2$, and boundary cutoff at 120Å$^3$. Those positions in between the two cutoffs are then boundary sites. The amino acid choices at core sites are[4], Ala, Val, Leu, Ile, Phe, Thy, and Trp, at surface sites, Ala, Ser, Thr, His, Asp, Asn, Glu, Gln, Lys, and Arg, and the combined set of 16 residues are used for boundary sites. Glycine is used whenever the backbone dihedral angle $\varphi<0$.

### *Rotamer library*

We employ the widely-used libraries by Dunbrack and his coworkers[5]. For calculations we describe here, we use the backbone-dependent library, which gives a set of rotamers for backbone dihedral angles ($\varphi$, $\psi$) separated by 10 degrees. For a given backbone, the ($\varphi$, $\psi$) sequence is calculated, and rotamers for the closest angle combination in the rotamer library are retrieved and used for our calculation. We control the number of rotamers for each amino acid by specifying the minimum total probability (normally set to 90 percent) and the minimum number of choices (if there are that many, normally set at 10). At a given position on the backbone, for a given rotamer, with its dihedral angles, we then

computationally build the coordinates of the sidechain atoms including hydrogen.

### *Energy function*

We have explored our energy function by repeated feedback study and have settled down to the following four terms.

1. Electrostatic energy

First, electrostatic energy in the following form is used,

$$E_{elec} = 322.0637\left(\frac{Q_i Q_j}{\varepsilon R}\right)$$

Where R is the separation between two atoms and $Q_i$ and $Q_j$ are their charges taken from CHARMM. The dielectric constant is taken to be a distance-attenuated value 40R[6], with which $E_{elec}$ is significant for closely placed atoms only.

2. van der Waals energy

Second, the van der Waals energy is

$$E_{vdw} = D_0\left[\left(\frac{R_0}{R}\right)^{12} - 2\left(\frac{R_0}{R}\right)^{6}\right]$$

Where R is again the separation between two atoms and $R_0$ and $D_0$ are again taken from CHARMM. $E_{vdW}$ is not computed for two atoms separated by one or two covalent bonds, and for those separated by three bonds, $R_0$ and $D_0$ are scaled. Furthermore, the van der Waals radius $R_0$ is scaled by a factor of 0.9 to accommodate some sidechain and backbone flexibility[6].

3. Solvation energy

Third, as hydrophobicity is believed to be the driving force for folding, a solvation term that encourages the burial of nonpolar atoms is crucial. We have also included a term that discourages the exposure of nonpolar atoms so that large sidechains are not always favored by solvation energy. We have the following expression,

$$E_{sol} = \sigma_{np}^{buried} A_{np}^{buried} + \sigma_{np}^{exposed} A_{np}^{exposed}$$

where we have taken $\sigma_{np}^{buried}$ =-30cal/mol/Å$^2$ and $\sigma_{np}^{exposed}$ =100cal/mol/Å$^2$.[7]

One complication is that solvation is a cooperative effect, which means we need to know the rotamers for the entire chain before we can calculate solvent accessible surface area and therefore solvation energy. Yet, the dead-end elimination algorithm requires that the total energy is written with single rotamer and rotamer pair terms only. Street and Mayo have investigated ways to write total area in pairwise terms. We have used the following formula suggested by them.[8]

$$A = \sum_i A_i - s\sum_{i<j}(A_i + A_j - A_{ij})$$

Here A is the total solvent accessible area of the entire chain; $A_i$ is the solvent accesible area for residue at i with other sites empty and the area buried by the backbone excluded; and $A_{ij}$ is the solvent accesible area for twi residues at i and j with other sites empty and the area buried by the backbone excluded. $A_i$+$A_j$-$A_{ij}$ is then the area buried by the two residues at i and j but is individually accessible for i or j. It should be subtracted from $A_i$+$A_j$, yet, because there are three and four-body terms, etc., $A_i$+$A_j$-$A_{ij}$ overcounts the true area buried, and a factor s is placed in front of it. We have used s=0.70 after checking pairwise areas with the true total area for a close match.

4. Secondary structure propensities

For our design of βαβ we have one α-helix and two β-sheets. We have taken the experimentally determined values for α-helix and β-sheet.[9, 10]

## *Experimental Section*

### *Protein preparation*

The artificial genes of the designed peptides were carefully selected using optimal codons for *E. coli*. The genes were synthesized and cloned into plasmid pGEX-4T-1 (Invitrogen). The vector was transformed into *E. coli* strain XL1-blue for amplification and the DNA sequence was determined by Invitrogen Biotech Co. Ltd. The mutations were made using a PfuUltra$^{TM}$ High-Fidelity DNA polymerase (Stratagene).

The plasmid was transformed into *E.coli* strain Rosetta <DE3>. Cells were cultured in LB medium at 37$^{o}$C with 100mg/L of Ampicillin and 34 mg/L Chloramphenicol. When the $OD_{600}$ reached 0.8, the culture was induced at 30$^{o}$C with 0.5mM isopropyl-thio-β-D-glactosidase (IPTG). The cells were harvested after 5 hours incubation by centrifugation at 5000 rpm for 10 min. The cells were suspended in lysate buffer (40mM sodium phosphate, pH 7.3, 100mM NaCl, 1mM Pheylmethylsulfonyl fluoride (PMSF), 5 mM ethylenediaminetetraacetic acid (EDTA)) and lysed by sonication. The lysate was separated by centrifugation at 17,000rpm for 30 min at 4$^{o}$C. The supernatant was filtered and purified by GST affinity column (GE Healthcare). The fusion protein was digested by thrombin in digestion buffer (40mM sodium phosphate, pH 7.3, 100mM NaCl) at room temperature for 10 hours. Further purification was done using $C_{18}$ column on reversed-phase high performance liquid chromatography (HPLC). The molecular weight was confirmed by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. All of the proteins dissolved well in water and most of them had a yield higher than 1mg/L.

### *Circular dichroism studies*

CD spectra data were collected on a Jasco J810 CD spectrometer at room temperature. Lyophilized peptide powders were dissolved in 20mM sodium phosphate buffer, pH 7.3. Most of the peptides concentrations were estimated from the calculated molar extinction coefficient at 280nm[11]. For the peptide W2T with no tryptophan, its concentration was calculated by the absorbance difference of $A_{225}$-$A_{215}$[12] . Cells with path length of 0.1mm, 1mm and 10mm were used in the concentration-dependent experiment and the corresponding peptide concentrations were 2mg/ml, 0.2mg/ml and 0.02mg/ml, respectively. The chemical denaturation experiments were prepared with different concentrations of guanidine hydrochloride in 20mM sodium phosphate buffer, pH 7.3. Cell with a path length of 0.1mm was used and the peptide concentrations are 1.5mg/ml. Peptides used in the thermal denaturation experiments had concentrations at 0.2mg/ml and the path length is 1mm. The temperature range was from 25$^{o}$C to 95$^{o}$C, with a step of 5$^{o}$C. 3 minutes equilibrium for the samples was setup at every temperature points. The spectra were collected and the data at 220nm were selected for analysis. The denaturation profiles were fitted to a two-state model using the equation: $y = y_u+(y_n-y_u)/(1+\exp((-\Delta G_{u(H2O)}-m*x)/RT))$.

### *Analytical gel filtration*

The aggregation states of the designed peptides were analyzed using pre-packed superdex high-performance columns, superdex peptide 10/300 GL (GE Healthcare), on an ÄKTA fast protein liquid chromatography equipment. Lyophilized peptides or marker proteins were dissolved into equilibration buffer (2mg/ml, 40mM sodium phosphate, 100mM NaCl at pH 7.3). After equilibration at room temperature for 30 minutes, 100µl samples were injected into the column and eluted with a flow rate of 0.5 ml/min. The eluted peaks were monitored at 280nm and 220nm. All of the marker proteins, cytochrome C for 1.24K, aprotinin for 6512 and Vitamin $B_{12}$ for 1355, were purchased from Sigma Corporation.

### *Structure Calculations*

For the NMR experiments, 1mM samples were prepared in 40mM sodium phosphate at pH 6.0 with 10% $D_2O$. DQF-COSY, TOCSY and NOESY spectra were acquired to accomplish resonance assignments and structure determination. They were collected on a Bruker Avance spectrometer at 298K. The $^{13}$C-HSQC spectrum used for $CH_2$ or $CH_3$ identification was collected on a 600 MHz spectrometer. NOESY spectra were recorded with mixing time of 60ms, 120ms, 160ms, and 300ms. NOESY with a 120ms mixing time were used to derive distance restraints. All the spectra were processed with NMRpipe[13] and analyzed using the program NMRView[14]. Crosspeaks intensities

were translated into upper limit distance constraints: 2.2 Å, 2.8 Å, 3.5 Å, 4.2 Å, 4.5 Å, 5.2 Å and 5.6 Å. The initial assignments were manually made according to the computational model. The secondary structure regions were judged by Chemical Shift Index (CSI) module in NMRView. Structure calculations were performed with CYANA[15] using standard protocols for distance geometry simulated annealing. The NOE-derived distance restraints and dihedral angles restraints were both considered. An automatic program SANE[16] were used to extend ambiguous NOE signals. Finally, 200 structures were calculated with good geometry and no distance restraint violations greater than 0.3 Å. The ensemble of the first 100 structures was optimized using Sander module of AMBER7 (http://amber.scripps.edu/). Restraints of distance and chirality were used in the optimization. The first 20 lowest-energy structures were selected and analyzed using the program PROCHECK_NMR[17]. An average structure was generated by averaging the coordinates of the 20 ensemble. It was presented as the final structure of DS119.

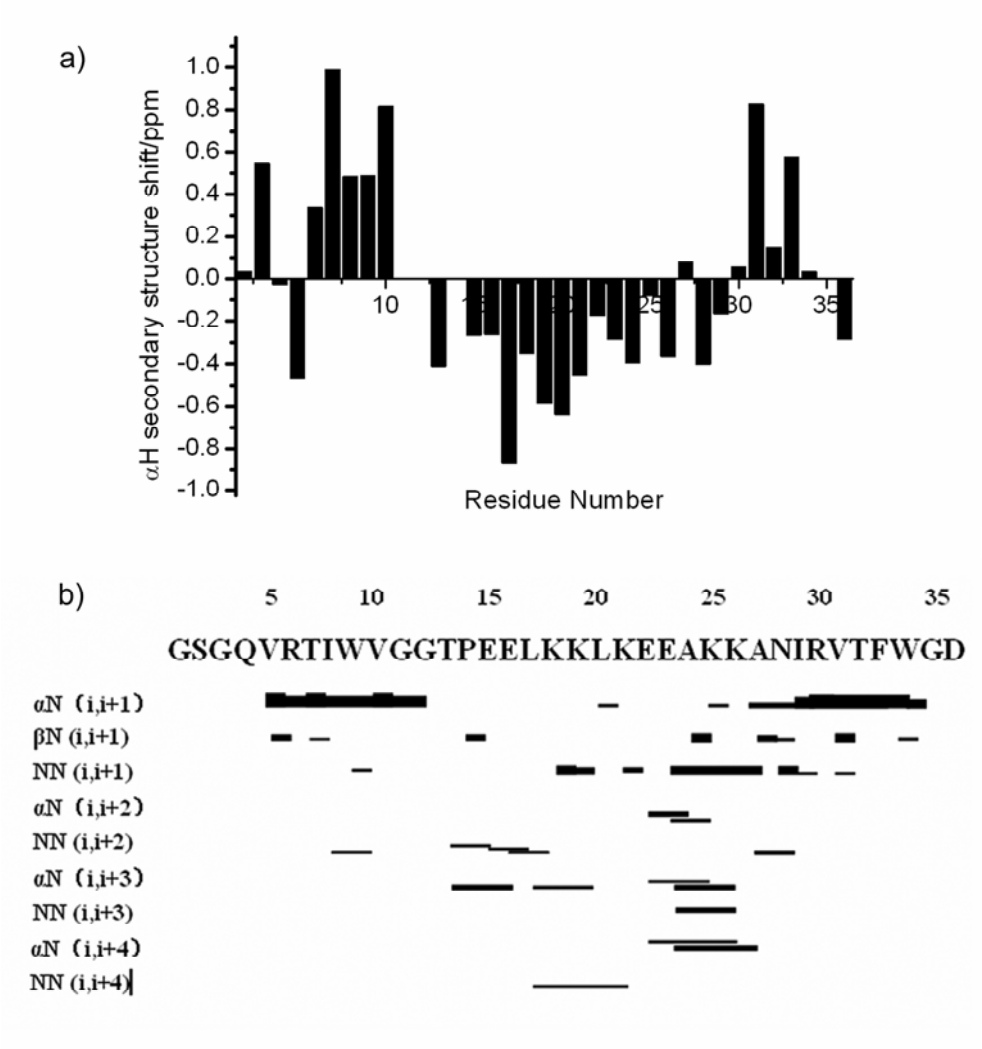


**Figure S3.** NMR structure assignment of DS119. (a)$\delta_{\alpha H}$ values of DS119 relative to random coil shifts. (b) The backbone NOEs of secondary structures. The different width of the lines represents the relative strength of the NOEs.

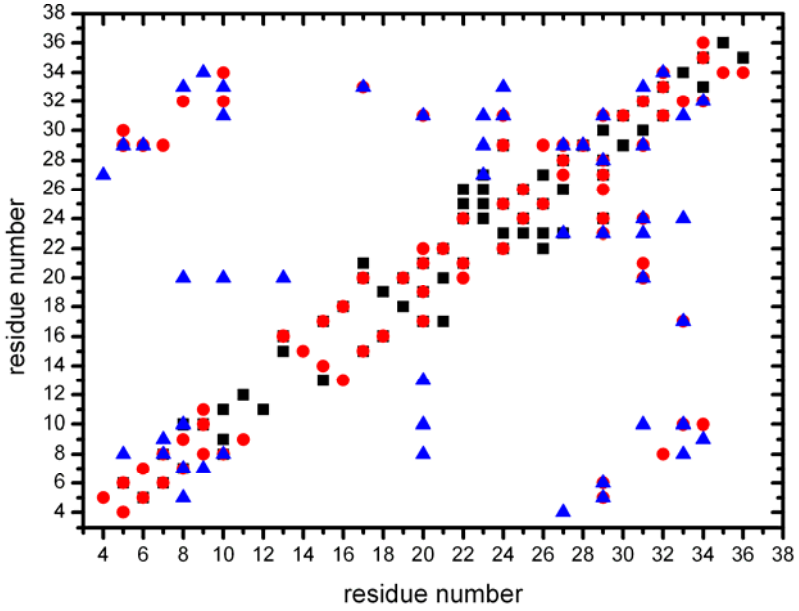


**Figure S4.** Two dimensional representation of inter-residue unambiguous NOE. Backbone-backbone, backbone-side chain, and sidechain-sidechain interactions are indicated by black squares, red diamonds and blue triangles, respectively.

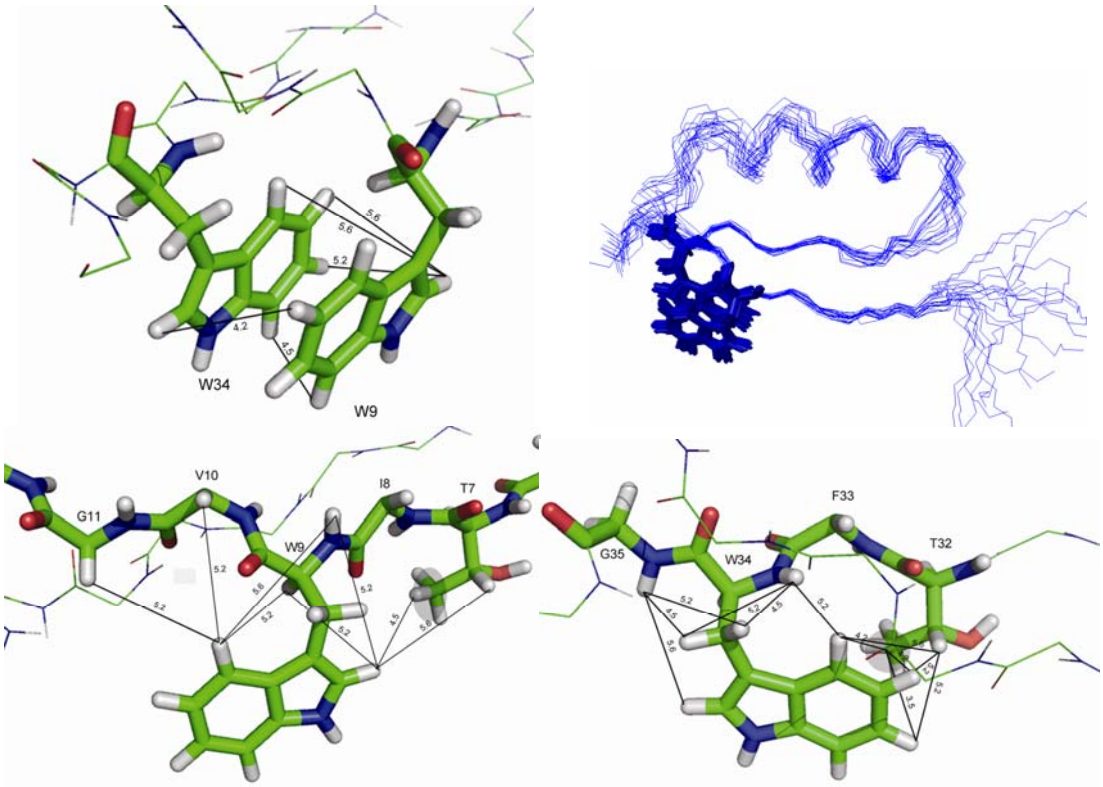


**Figure S5.** The WW interaction in DS119. The NOE distance restraints between indole rings were shown in detail. When the backbone of β-sheet regions (residues 6-10 and 30-34) were fitted, the tryptophans in the top 20 NMR structures superposed well.

**Figure S6**. Structure comparison of the designed peptide and the natural protein fragment. The average experimentally determined NMR structure is shown in blue and the 122-154AA in the B chain of 1JWB in green. Backbone RMSD of these two structures is 1.52Å.

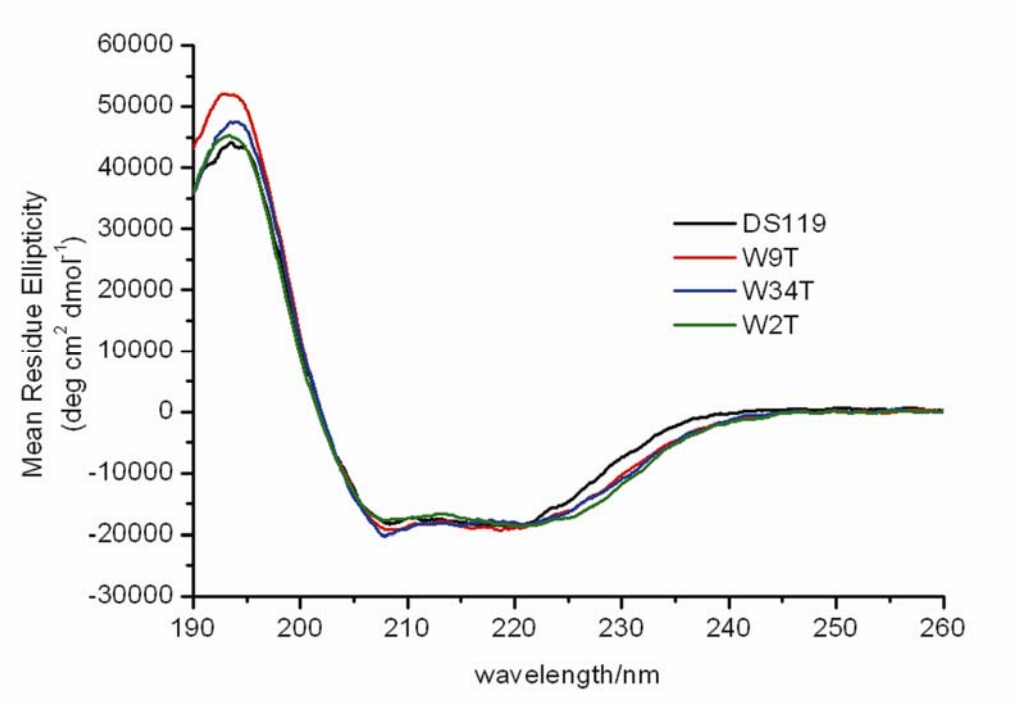


**Figure S7.** Comparison of the far-UV CD spectra of DS119, W9T, W34T and W2T.

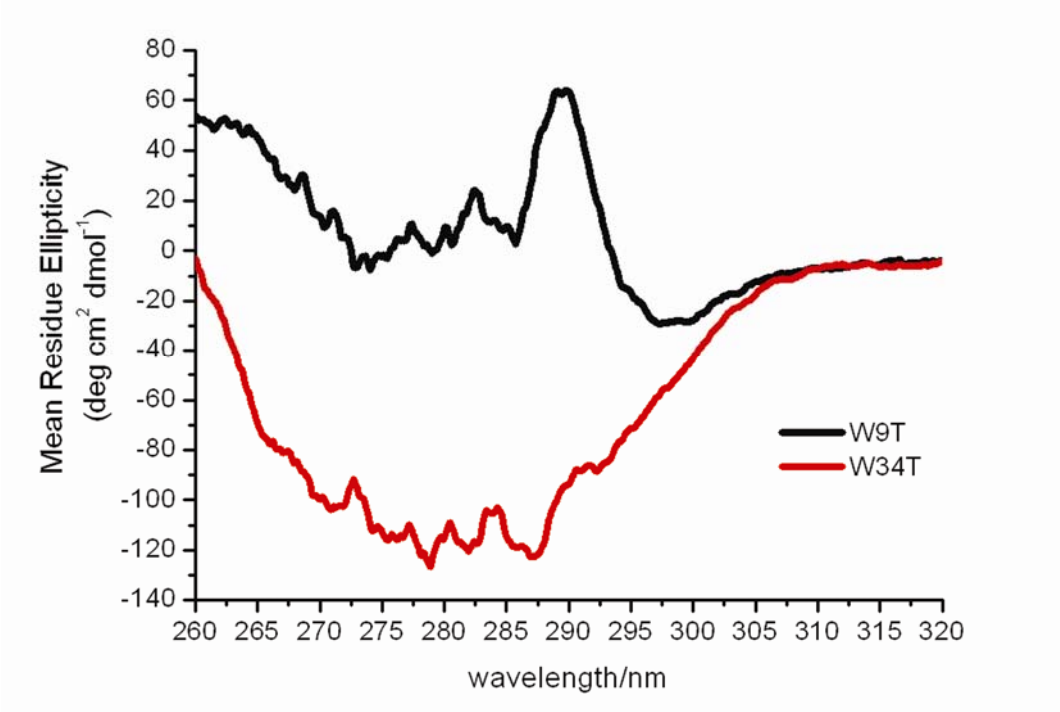


**Figure S8.** The near-UV CD spectra of W9T and W34T.

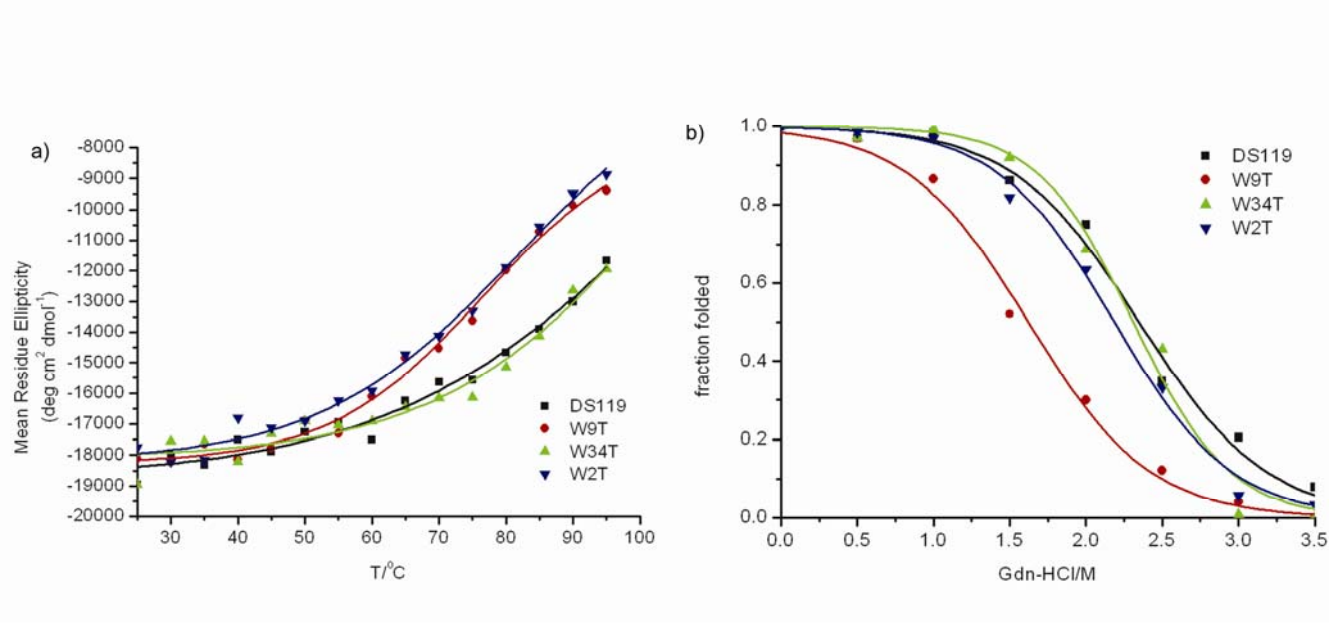


**Figure S9.** Stability studies of the designed peptides. (a) Temperature melting curves. (b) Gdn-HCl denaturation curves. The mean CD residue ellipticity at 220nm was monitored. Peptides were dissolved in 20mM phosphate buffer, pH 7.3. The concentration for temperature melting was 0.2mg/ml (a) and for Gdn-HCl denaturation was 1.5mg/ml (b).

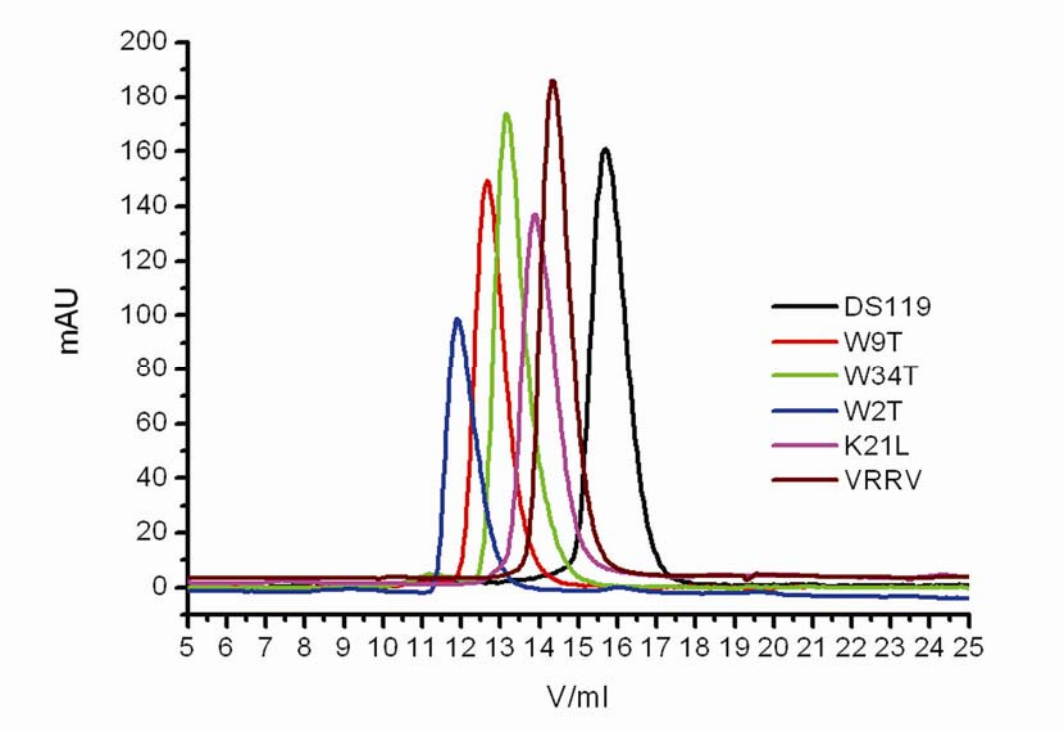


***Figure S10.*** Gel filtration studies of DS119 and other mutations. A pre-packed superdex peptide 10/300 GL high-performance column was used to analyze the aggregation state. The elution buffer was 40mM phosphate, pH 7.3, with 100mM NaCl.

**Table S1** NMR structure statistics. All statistics were over the first 20 low-energy structure of 100 structures. In the 20 accepted structures, there were no restraint violations >0.2Å.

| NMR structure statistics | | |
|---|---|---|
| *Distance restraints* | | |
| Intraresidue unambiguous NOEs | 297 | |
| Sequential unambiguous NOEs | 101 | |
| Medium range unambiguous NOEs | 97 | |
| Long range unambiguous NOEs | 103 | |
| Total unambiguous NOEs | 598 | |
| Total ambiguous NOEs | 294 | |
| Dihedral angles | 0 | |
| *Ramachandran stratistics* | | |
| Residues in most favored regions (%) | 86.0 | |
| Residues in additional allowed regions (%) | 13.6 | |
| Residues in generously allowed regions (%) | 0.2 | |
| Residues in disallowed regions (%) | 0.2 | |
| *Atomic r.m.s.deviations(A)* (4-34) | | |
| | Backbone | Heavy atoms |
| Between two structures | 1.127±0.238 | 1.749±0.207 |
| With mean structures | 0.777±0.138 | 1.205±0.121 |

**Table S2** Chemical shifts of all the assigned hydrogen in DS119.

| | | NH | αH | βH | Other H |
|---|---|---|---|---|---|
| 2 | Ser | 8.762 | 4.533 | 3.901 | |
| 3 | Gly | 7.862 | 4.515 | | |
| 4 | Gln | 8.223 | 4.343 | 1.980 | γH: 2.104 2.364; δNH：6.895 7.568 |
| 5 | Val | 8.055 | 3.823 | 1.912 | γH: 0.884 0.806 |
| 6 | Arg | 8.632 | 4.717 | 1.760 | γH: 1.863 1.611; δH：3.195; NH：7.342 |
| 7 | Thr | 8.126 | 5.338 | 3.792 | γH：0.955 |
| 8 | Ile | 9.142 | 4.714 | 1.076 | γ1H：2.243; γ2H：1.342; δH：0.806 |
| 9 | Trp | 8.117 | 5.187 | 2.179 | δH1：6.584; εH1: 9.260; ζH2: 7.206; ηH2: 6.998; ζH3: 6.888; εH3: 7.239 |
| 10 | Val | 8.809 | 4.993 | 1.596 | γH: 0.318 0.475 |
| 11 | Gly | 8.918 | 5.762 3.587 | | |
| 12 | Gly | 8.661 | 4.029 3.947 | | |
| 13 | Thr | 8.662 | 4.348 | 4.107 | γH：0.806 |
| 14 | Pro | | | 1.470 | |
| 15 | Glu | 8.089 | 4.025 | 2.067 2.196 | γH：2.352 2.485 |
| 16 | Glu | 8.722 | 4.028 | 1.973 | γH：2.144 2.362 |
| 17 | Leu | 7.703 | 3.507 | -0.212 0.863 | γH：1.470; δH: 2.446 1.243 |
| 18 | Lys | 7.520 | 4.010 | 1.886 | γH：1.569 1.451; δH: 1.688; εH: 2.932 |
| 19 | Lys | 7.943 | 3.774 | 1.856 | γH：1.432 1.283; δH: 1.642; εH: 2.920 |
| 20 | Leu | 8.339 | 3.740 | 1.434 1.793 | γH：1.035; δH: 0.517 0.706 |
| 21 | Lys | 8.211 | 3.926 | 2.205 | γH：1.440; δH: 1.032; εH: 3.369 |
| 22 | Glu | 8.672 | 4.117 | 2.094 1.977 | γH：2.210 2.360 |
| 23 | Glu | 7.528 | 4.004 | 1.958 1.731 | γH：2.575 2.346 |
| 24 | Ala | 8.356 | 3.954 | 1.645 | |
| 25 | Lys | 7.449 | 4.278 | 1.997 | γH：1.475 1.589; δH: 1.751; εH: 3.016 |
| 26 | Lys | 8.224 | 3.996 | 1.887 | γH：1.568 1.462; δH: 1.689 |
| 27 | Ala | 7.936 | 4.426 | 1.379 | |
| 28 | Asn | 7.871 | 4.348 | 3.143 2.713 | γNH: 7.529 6.778 |
| 29 | Ile | 8.498 | 4.064 | 0.808 | γ1H：0.737; γ2H：1.694; δH：0.946 |
| 30 | Arg | 9.120 | 4.436 | 1.724 1.842 | γH：1.633 1.557; δH: 3.167;NH: 7.152 |
| 31 | Val | 7.417 | 5.005 | 1.756 | γH：0.423 0.632 |
| 32 | Thr | 8.747 | 4.500 | 3.755 | γH：0.929 |
| 33 | Phe | 8.785 | 5.237 | 2.910 | δH：7.236; εH: 7.073; εH: 6.845 |
| 34 | Trp | 8.908 | 4.735 | 3.129 2.710 | δH1：7.031; εH1: 9.752; ζH2: 6.869; ηH2: 7.244; ζH3: 7.310; εH3: 7.047 |
| 35 | Gly | 8.423 | 4.314 3.798 | | |
| 36 | Asp | 8.053 | 4.477 | | |

## Reference


[1] C. A. Rohl, *Biochemistry* **1992**, *31*, 1263.
[2] H. E. Stanger, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 12015.
[3] S. A. Marshall, *J. Mol. Biol.* **2001**, *305*, 619.
[4] B. I. Dahiyat, *Science* **1997**, *278*, 82.
[5] R. L. Dunbrack, *J. Mol. Biol.* **1993**, *230*, 543.
[6] D. B. Gordon, *Curr. Opin. Struct. Biol.* **1999**, *9*, 509.
[7] B. I. Dahiyat, *Protein Sci.* **1996**, *5*, 895.
[8] A. G. Street, *Fold Des.* **1998**, *3*, 253.
[9] C. N. Pace, *Biophys. J.* **1998**, *75*, 422.
[10] A. G. Street, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9074.
[11] C. N. Pace, *Protein Sci.* **1995**, *4*, 2411.

[12] G. R. Grimsley, *Curr. Protoc. Protein. Sci.* **2004**, *Chapter 3*, Unit 3 1.
[13] F. Delaglio, *J. Biomol. NMR* **1995**, *6*, 277.
[14] B. A. Johnson, *Methods Mol. Biol.* **2004**, *278*, 313.
[15] P. Guntert, *J. Mol. Biol.* **1997**, *273*, 283.
[16] B. M. Duggan, *J. Biomol. NMR* **2001**, *19*, 321.
[17] R. A. Laskowski, *J. Biomol. NMR* **1996**, *8*, 477.