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.

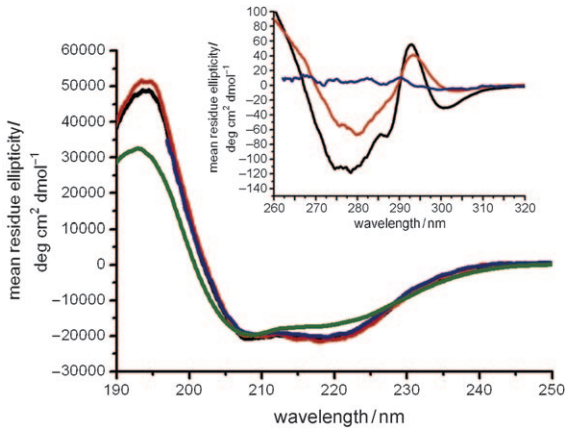

Figure 1. CD spectra of DS119. Experiments were carried out at 25 °C, pH 7.3 in 20 mM phosphate buffer at different concentrations: 2 mg mL^{-1} (——), 0.2 mg mL^{-1} (——), and 0.02 mg mL^{-1} (——). DS119 shows no concentration dependence and retains most of its secondary structure at 90 °C (——). The 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 (——).

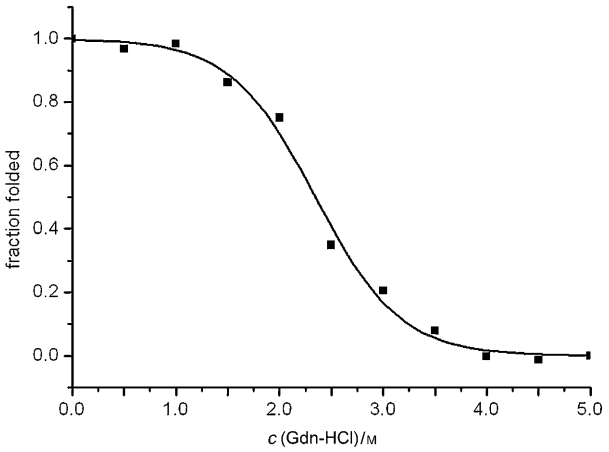

Figure 2. Chemical denaturation of DS119. The mean CD residue ellipticity at 220 nm was monitored as a function of the concentration of Gdn-HCl.

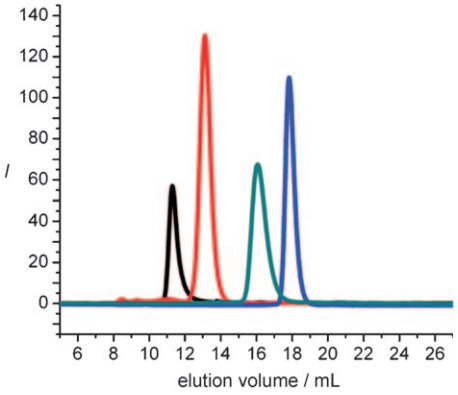

Figure 3. Gel-filtration analysis of DS119. A prepacked superdex peptide 10/300 GL high-performance column was used to analyze the aggregation state of DS119 (4028 Da, ——). The molecular markers are cytochrome C (12.4 kDa, ——), aprotinin (6512 Da, ——), and vitamin B_{12} (1355 Da, ——). The vertical axis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

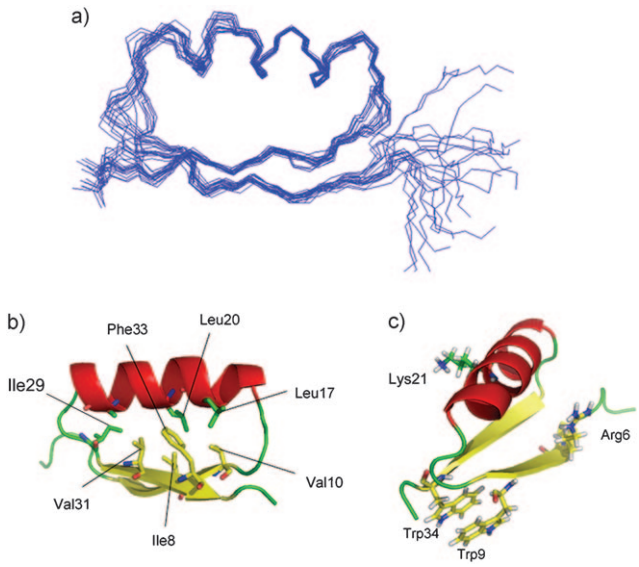

Figure 4. NMR structures of DS119.[9] a) The best 20 structures obtained from structure calculations by NMR spectroscopy. The structures were superimposed by fitting the secondary-structure regions (residues 6–10, 15–26, 30–34). b) The average structure of the top 20 structures is shown with the hydrophobic 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.

