Selection and structural analysis of *de novo* proteins from an $\alpha 3\beta 3$ genetic library

Mariejoy Therese Jumawid,¹ Tsuyoshi Takahashi,^{1*} Toshimasa Yamazaki,² Hiroshi Ashigai,¹ and Hisakazu Mihara¹

¹Department of Bioengineering, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta, Yokohama 226-8501, Japan

²Protein Research Unit, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8602, Japan

Received 15 July 2008; Revised 19 November 2008; Accepted 24 November 2008

DOI: 10.1002/pro.41

Published online 5 January 2009 proteinscience.org

Abstract: The construction of novel functional proteins has been a key area of protein engineering. However, there are few reports of functional proteins constructed from artificial scaffolds. Here, we have constructed a genetic library encoding $\alpha 3\beta 3$ *de novo* proteins to generate novel scaffolds in smaller size using a binary combination of simplified hydrophobic and hydrophilic amino acid sets. To screen for folded *de novo* proteins, we used a GFP-based screening system and successfully obtained the proteins from the colonies emitting the very bright fluorescence as a similar intensity of GFP. Proteins isolated from the very bright colonies (vTAJ) and bright colonies (wTAJ) were analyzed by circular dichroism (CD), 8-anilino-1-naphthalenesulfonate (ANS) binding assay, and analytical size-exclusion chromatography (SEC). CD studies revealed that vTAJ and wTAJ proteins had both α -helix and β -sheet structures with thermal stabilities. Moreover, the selected proteins demonstrated a variety of association states existing as monomer, dimer, and oligomer formation. The SEC and ANS binding assays revealed that vTAJ proteins tend to be a characteristic of the folded protein, but not in a molten-globule state. A vTAJ protein, vTAJ13, which has a packed globular structure and exists as a monomer, was further analyzed by nuclear magnetic resonance. NOE connectivities between backbone signals of vTAJ13 suggested that the protein contains three α -helices and three β -strands as intended by its design. Thus, it would appear that artificially generated $\alpha 3\beta 3$ *de novo* proteins isolated from very bright colonies using the GFP fusion system exhibit excellent properties similar to folded proteins and would be available as artificial scaffolds to generate functional proteins with catalytic and ligand binding properties.

Keywords: protein design; $\alpha 3\beta 3$ *de novo* protein; GFP-based screening; NMR

Introduction

The diversity of unique functions seen in a variety of proteins has attracted the interests of chemistry and biology researchers. At the same time, as this upsurge of interest in proteins, a number of principles and techniques have been developed, which may further

our understanding of the nature and function of proteins. Protein engineering is not only focused on understanding the function of proteins, but also on designing and constructing novel proteins. Various attempts to construct novel proteins using combinatorial library technology and rational protein design have been reported. Along with combinatorial library technology, a variety of screening systems, such as phage display, ribosome display, etc., have been constructed.^{1–5}

Although a large library around 10^{12} can be prepared using phage and mRNA display systems, the number of all sequences of proteins using 20 amino acids at each position is too large to be covered. For example, the number of possible protein sequences

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Ministry of Education, Science, Culture and Sports, Japan.

*Correspondence to: T. Takahashi, Department of Bioengineering, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta, Yokohama 226-8501, Japan.
E-mail: ttakahas@bio.titech.ac.jp

resulting from 100 amino acid residues is 10^{130} . Moreover, protein function is closely related to highly ordered structure. The possibility of constructs comprising well-folded proteins is low and functional proteins are rarely found, even when libraries are fully randomized. Therefore, in order to obtain the novel functional proteins with native-like packing, protein in libraries should be specified to fold. Recently, P.G. Schultz and coworkers have been reported that the selection of folded proteins from a large library composed of secondary structural elements derived from *E. coli* proteins using a green fluorescent protein (GFP) fusion system.⁶ Some proteins obtained were found to have unique properties.

In addition to the library method, rational protein design has been attempted primarily using computational techniques.^{7–10} Several groups have accomplished the rational design of proteins, of which W.F. DeGrado and co-workers successfully designed catalytic metalloproteins capable of catalyzing a phenol oxidase reaction.^{11,12} These metalloproteins composed of a four-helix bundle structure were well-packed, and the catalytic residues were well organized in the hydrophobic core. M.H. Hecht *et al.* combined the combinatorial approach with rational techniques, and constructed a *de novo* protein library using “binary code” strategy.^{13–18} The designed α -helix and β -strand were generated by arranging hydrophobic and hydrophilic amino acids in an amphiphilic manner. Four-helix bundle proteins and amyloid proteins were constructed using α -helix and β -sheet scaffolds, respectively. Of note, several proteins in the helix bundle proteins possessed some native-like characteristics, and nuclear magnetic resonance (NMR) measurements indicated that the second-generation proteins fold into structures containing many tertiary contacts. D. Hilvert and coworkers have successfully constructed an active enzyme using a binary code and simplified amino acid sets.¹⁹ These studies suggest that a combinatorial approach using the binary code and simplified amino acid sets can be used to construct artificial proteins from a reduced library size. Although protein design using α -helices have been successful, proteins constructed from artificial scaffolds containing both α -helices and β -strands have scarcely been reported, despite the variety of combinations of α -helices and β -strands seen in many naturally occurring proteins. It has been difficult to design proteins with β -sheet structures, because the β -sheet proteins tend to aggregate into amyloid-like fibrils. Therefore, we considered that the library proteins with both α -helices and β -strands are constructed and folded proteins are selected using an *in vivo* screening system to remove improper structures of proteins.

Here, we designed a *de novo* protein library encoding three α -helices and three β -strands ($\alpha 3\beta 3$) as an artificial scaffold using a binary code system with simplified amino acid sets. Of note, we employed an

amino acid Ala, which appears in both surface and buried areas of proteins, as a modulator in the hydrophobic interaction. Using both α -helices and β -strands, a number of topologies can be generated from the genetic library. We then screened folded proteins from the library using a GFP fusion system, and characterized the structural properties of the selected proteins. Proteins selected from the $\alpha 3\beta 3$ library show characteristics of folded proteins and are more likely to possess the $\alpha 3\beta 3$ structure as intended by our design [Fig. 1(A)].

Results and Discussion

Design of $\alpha 3\beta 3$ *de novo* protein library

To construct an artificial scaffold, we constructed a library of 103-residue proteins designed to fold into three α -helices and three β -strands ($\alpha\text{-}\alpha\text{-}\alpha\text{-}\beta\text{-}\beta\text{-}\beta$; $\alpha 3\beta 3$). Each α -helix was composed of 14 amino acids and each β -strand was composed of seven amino acids (see Fig. 1). Based on the pioneering work of M.H. Hecht *et al.*,^{13–18} the α -helix and β -strand were designed as linear sequences according to the following pattern of hydrophilic (○) and hydrophobic amino acids (●): ○●○●○●○●○●○●○●○ for the α -helix and ○●○●○●○ for the β -strand. Although Hecht *et al.* reported that longer helix was suitable for constructing the helix bundle protein using the combinatorial library, we used the 14-helix to generate smaller proteins. Seven different amino acids were used to construct the library: Lys, Glu, Thr, and Ala were positioned in the hydrophilic sites, and Leu, Ile, and Val were positioned in the hydrophobic sites. It has been predicted that Lys and Glu make electrostatic interaction at proper positions. Although proteins usually have aromatic amino acids, Trp, Phe, and Tyr, we omitted the aromatic amino acids in both α -helices and β -strands to know whether proteins can fold without these amino acids. Moreover, if the selected proteins can fold into specific structures without any aromatic side chain, the aromatic amino acids would be incorporated into the proteins to add functionality such as binding to target compounds. Therefore, we have focused on the availability of aliphatic amino acids, Leu, Ile, and Val. Ala is a nonpolar amino acid that has a small methyl group on its side chain. Thus, Ala can be represented both on the surface and within the interior of the protein structure. Based on this, we designed the mixed codons corresponding to their amino acids: RMR (R = A or G; M = A or C) for the hydrophilic amino acids and VTH (V = A, C, or G; H = A, C, or T) for the hydrophobic amino acids. Mixed codons for the hydrophilic residues of the α -helix were manipulated so that Lys and Glu were three times more likely to be presented than Thr and Ala, whereas codons for the hydrophobic residues were manipulated so that Leu and Ile were twice more likely to be presented than Val. Unlike the α -helix, codons were

Figure 1. (A) Schematic representation of designed α -helix and β -strand using a binary code pattern. $\circ$ and $\bullet$ depict the position of hydrophilic and hydrophobic amino acids, respectively. (B) DNA sequences corresponding to the designed α -helix (upper) and β -strand (lower). Possibilities presented are shown in the Table. (C) Schematic representation of the construction of the $\alpha\beta3$ genetic library. L, α , and β represent the linker, α -helix, and β -strand sequences, respectively.

designed so that amino acids were equally distributed within the β -strand. A linker sequence composed of Gly-Gly-His-Gly-Gly was positioned between each fragment and at the N- and C-terminal ends. This flexible linker was chosen so as not to prevent protein folding. His residues were incorporated into the linker region expecting that the structures are modulated by metal

binding in the future studies. Single-stranded DNA molecules encoding the α -helix or β -strand with the linker sequence were extended using a DNA polymerase, after which the double-stranded DNAs were digested at sequences corresponding to the linker region using the restriction enzymes *BbsI* and *BsaI* [Fig. 1(C)]. These digested DNAs were ligated using T4

Figure 2. Schematic representation of the selection of folded and well-expressed $\alpha 3 \beta 3$ *de novo* proteins using a GFP fusion system. When an $\alpha 3 \beta 3$ protein folds, GFP fused at the C-terminal of the protein can also fold. In contrast, when an $\alpha 3 \beta 3$ protein cannot fold, GFP cannot also fold.

DNA ligase. Finally, full-length DNAs encoding $\alpha 3 \beta 3$ *de novo* proteins were inserted into a vector containing a *gfp* gene.

Screening of the folded $\alpha 3 \beta 3$ *de novo* proteins from the library

To screen for the folded $\alpha 3 \beta 3$ *de novo* proteins from the constructed genetic library, we applied a system using a GFP as a folding reporter^{20–22} (see Fig. 2). Using this system, if a protein folds, the GFP fused to its C-terminal can emit green fluorescence. On the other hand, if a protein has a difficulty to fold into a specific structure, its solvent-exposed hydrophobic amino acids can disrupt GFP folding, producing less fluorescence. Using this system, 201 very bright colonies emitting similar fluorescence of GFP were isolated from 3300 colonies (Table I). An additional 916 bright colonies showing weak fluorescence were also selected. The remaining colonies did not emit fluorescence. By analyzing the 48 colonies that did not emit fluorescence by DNA sequencing, eight clones contained the $\alpha 3 \beta 3$ *de novo* protein gene with *gfp* gene. The rest (40/48) did not contain the *gfp* gene due to a mutation at the region of $\alpha 3 \beta 3$ *de novo* protein gene. Because this kind of designed artificial gene tends not to be expressed in *E. coli*, clones with the correct gene were more difficult to grow than those with a mutated gene. Thus, numerous negative colonies appeared on the plate. Out of green colonies analyzed, 13 and 12 clones containing $\alpha 3 \beta 3$ sequences were evaluated from the very bright and bright colonies, respectively.

Proteins from the very bright colonies were named vTAJ01, vTAJ03, vTAJ09, vTAJ12, vTAJ13, vTAJ22, vTAJ23, vTAJ27, vTAJ30, vTAJ32, vTAJ34, vTAJ36, and vTAJ38, whereas proteins from the bright colonies were named wTAJ111, wTAJ119, wTAJ213, wTAJ308, wTAJ419, wTAJ515, wTAJ603,

wTAJ609, wTAJ613, wTAJ614, wTAJ615, and wTAJ620. The amino acid sequences and schematic diagrams of these proteins were shown in Fig. 3, S1, and S2. Although a set of only seven amino acids was used in the core sequences on α -helices and β -strands in design of the genetic library, other kinds of amino acids were, as a result, incorporated in five of the 13 clones from the very bright colonies: Arg in vTAJ01, His and Asp in vTAJ13, Asn in vTAJ27, Arg in vTAJ30, Phe and Gln in vTAJ38. In contrast, two of the 12 clones from the bright colonies demonstrated the incorporation of unexpected amino acids (Asp in wTAJ613, Gln in wTAJ620). It seems that the amino acids in addition to Lys, Glu, Thr, Ala, Leu, Ile, and Val of vTAJ proteins could function to form the specific structures. Especially, His and Arg in vTAJ13 and vTAJ30, respectively, by which amino acids at the hydrophobic positions were replaced, might be used to make specific contacts in the hydrophobic core. On the other hand, only one Phe residue was used in both vTAJ and wTAJ sequences, indicating that the folded proteins might be generated without any aromatic amino acid by development of the strategy combining

Table I. Summary of the selection of the $\alpha 3 \beta 3$ *de novo* proteins from the library

Number	Very bright	Bright	Negative
Colonies ^a	201	916	2170
Analyzed colonies ^b	29	59	
Colonies $\alpha 3 \beta 3$ ^c	13	12	
Proteins expressed ^d	9	6	
Proteins analyzed ^e	8	4	

^a Colonies were observed on LB-agar plates.

^b Colonies analyzed by DNA sequencing.

^c Confirmed to contain the $\alpha 3 \beta 3$ sequences as designed.

^d Proteins isolated from the colonies could be solely expressed in *E. coli*.

^e Proteins could be refolded and analyzed.

	$\alpha 1$	$\alpha 2$	$\alpha 3$	$\beta 1$	$\beta 2$	$\beta 3$
vTAJ01	MYGGHGEIKKIKKIKKIKKGGHGGK	<u>RE</u> KILAKVKKIVIEGGHGGKIEEV	IKELKEVIKGGHGAIEVAVTGGHGGELKIKVKGHGGVEIAITLGGHGGSP			
vTAJ03	-----KIKKIEIEKIEKVIA-----	EVKKLLAELKEVLE-----	KLEELVEELBEEIK-----	KLEIEVE-----	TLTLEVD-----	ALEVEVT-----
vTAJ09	-----KVEEIIIEIEEVIE-----	ELAALIEVEVEIE-----	AIEKILEKIEEIVE-----	ALAIEVT-----	KVEVAVE-----	EIAVALT-----
vTAJ12	-----AIEELLEELKAVLE-----	EIKKILTEIKKIA-----	EIEAIVKEVAKILE-----	EVEVTVK-----	ELALELK-----	AVEVAIA-----
vTAJ13	-----KIKKIEIEKIKKVVE-----	EIEELHAELEAIVE-----	EIKKIEIKIEKILE-----	EVVVKLA-----	AIKIKVT-----	AVTVEIE-----
vTAJ22	-----KIEKIIETIKKVL-----	KIKALLKAIKAVLA-----	EIEELIKEIKKIVK-----	AVAIAIA-----	TIALKIE-----	TLAIEVK-----
vTAJ23	-----KLAELVEKLKEVLE-----	ALKAIKKIKELVE-----	ELEKIEEVEKIE-----	ELKLALK-----	EITIAVA-----	AIELKLE-----
vTAJ27	-----KVEEIIIEVEKIK-----	EVKTILEAIEELIK-----	EVKELVEKIEEILK-----	NITVAIE-----	TVEIALT-----	EIEVEVT-----
vTAJ30	-----EIEETIKKVAEVVE-----	ALEAILEEIAAILE-----	AIEKIIKKLKEILK-----	AVEVELA-----	AIKRLK-----	TVEVAVE-----
vTAJ32	-----EIEKIVAEIEAIVE-----	ELTELIETLAKIE-----	EIKKIEIKIKKVIE-----	ALTVEVE-----	KIKLKIT-----	AVKLKVK-----
vTAJ34	-----KLKELVAEIAITLLE-----	KLKAIKTEIEELLE-----	ELEKLIKIEKIVE-----	ELALELA-----	KLKIKLE-----	TVALAIA-----
vTAJ36	-----EIEKIEAIEETIK-----	ELEKVEEVEKIIA-----	KIKELIEEVEKLVE-----	EIEVEVE-----	KIKIKVK-----	TVAVALT-----
vTAJ38	-----KIEELITKIEEIK-----	KVEALLEELTKIIT-----	KIEKIITEVEEVVK-----	ALALEFE-----	AITIALQ-----	AVTVEVE-----
wTAJ111	-----EIKEIVEEIEKIK-----	EIEAILKELATILK-----	KLEKIIKKLEKIE-----	AITLAIA-----	TIEIEIT-----	EIAIALK-----
wTAJ119	-----EVKEVIEAVTKVLK-----	TLTAIVAELEEVLE-----	EIKKLEETIEKIVA-----	EVAIAVA-----	ELELAIE-----	AITLAIE-----
wTAJ213	-----KVKKVLKEIPEVVK-----	EIKKITEIKELIE-----	KLEKLIKIEKIVE-----	KVAIAIE-----	KVKLELE-----	ALAVTLA-----
wTAJ308	-----KIEELLKKIKIEIVA-----	EIEEVLKKIKELIA-----	KVPEIEEELKKVIE-----	AVELTVE-----	EVLTVA-----	TIALEVA-----
wTAJ419	-----EVKELIETIEELIK-----	TIAKIIIEAIEEILE-----	KLAKLLAEIKELIE-----	EVAVEIA-----	KLEITLA-----	EIKLEVA-----
wTAJ515	-----ALEKVIETIEKLK-----	ELTAVLEIEKVLIT-----	EIEKIIKKIETIIE-----	AVALKVE-----	KLEIKLE-----	KLTLELA-----
wTAJ603	-----EVKELIAEIAIVIK-----	KIKEILKELEALT-----	KIKEIKKAEIKLLE-----	ALTLKVT-----	KLTITLA-----	TVTLALA-----
wTAJ609	-----AVKKIKKAIKELIE-----	ELEKLIKIEVAALLE-----	KIKELIEEELKIE-----	TIEIELT-----	EVALAIE-----	ELAVEIA-----
wTAJ613	-----KIEEIVAEIEELVK-----	N--ALKDLIEKLEAIIK-----	KIKKIVKEVEEIK-----	ALTIALA-----	KIEIKIT-----	KVKIEVE-----
wTAJ614	-----EIAKILKTVKEILK-----	EVTTLVTELAEVL-----	TIEEVLAEIKIEVE-----	KLAIKVK-----	AIEVKIK-----	ELEIALA-----
wTAJ615	-----KLEKIIETIEKIEIL-----	EIKEILAEVEEILK-----	KVKLIKIEEIVK-----	KVAVKLE-----	AIKIELT-----	KVALAVE-----
wTAJ620	-----KLKEIKKLEKIK-----	KIQALLATIEAIEEC-----	KVEKILTEIEELIA-----	TLAVEIE-----	KITLAVK-----	TIALTLE-----

Figure 3. Amino acid sequences of vTAJ and wTAJ proteins. Unexpected amino acids presented in the core regions are underlined.

the combinatorial method and rational design of *de novo* proteins.

Circular dichroism study of selected proteins

To evaluate properties of the selected proteins, each protein was expressed solely and purified by reversed-phase HPLC with high purity. Although all vTAJ proteins were appeared in both the soluble and insoluble fractions, the proteins from insoluble fractions were used for the purification, because the purification was easier than that from the soluble fraction. In contrast, wTAJ proteins were tend to be appeared in insoluble fractions, indicating that the selection using the GFP fusion system may function to obtain the soluble and folded proteins. Some proteins, however, could not be expressed with *E. coli*. After refolding, eight and four proteins were obtained from very bright and bright colonies, respectively. Circular dichroism (CD) measurements of vTAJ proteins were performed in a buffer (50 mM phosphate, pH 7.5) with or without 100 mM NaCl (see Fig. 4). The shapes of the CD spectra of the proteins were similar and the spectra have negative maxima at 222 nm and 208 nm, which is characteristic of an α -helix structure. This observation suggests that the vTAJ proteins contain the α -helix structure. The intensities of the CD signals of the vTAJ proteins differed, suggesting that the vTAJ proteins represent a variety of the structural properties. For vTAJ27, vTAJ30, vTAJ34, and vTAJ38, CD spectra of each protein in the buffer were similar regardless of the presence or absence of NaCl, indicating that the structures of these vTAJ proteins were not salt sensitive and rather stable. On the other hand, the CD spectrum of the vTAJ32 protein could only be measured in the presence of 100 mM NaCl because the protein precipi-

tated within the buffer in the absence of NaCl. This result indicated that the salt stabilizes the structure of the vTAJ32 protein. Conversely, the CD intensities of vTAJ01, vTAJ13, and vTAJ36 decreased in the presence of 100 mM NaCl. This result suggests that the salt might destabilize the structures of vTAJ01, vTAJ13, and vTAJ36. Of note, the shapes of the CD spectra of vTAJ01 differed in the presence and absence of the salt, suggesting that the vTAJ01 protein has two different structures. Glu–Lys pairs corresponding to the i, i+4 positions on $\alpha 1$, $\alpha 2$, and $\alpha 3$ capable of forming salt bridges are depicted (Figs. S1 and S2). All vTAJ proteins have amino acid sequences that can be consistent with the formation of more than three ion pairs, indicating that the helical structures could be stabilized by the electrostatic interaction. However, the numbers of these ion pairs did not correlate with the structural changes of those in the presence and absence of the salt. Probably, the electrostatic interactions of helix–helix, helix–strand, and strand–strand could stabilize the secondary structures of the vTAJ proteins.

To evaluate the formation of secondary structures, the CD spectra were analyzed using the SELCON3 program²³ (Table S1). The theoretical α -helix and β -strand content of $\alpha 3\beta 3$ *de novo* proteins was 41 and 20%, respectively, according to our original design of the genetic library. Calculated values of α -helix and β -strand contents of vTAJ proteins, except for vTAJ01, were around 40 and 12%, respectively, indicating the presence of both α -helix and β -strand structures in the designed $\alpha 3\beta 3$ proteins. The calculated α -helix content matched the theoretical value, whereas the calculated β -strand content of the vTAJ proteins was smaller. Although the β -strand value is difficult to evaluate

Figure 4. CD spectra of (A) vTAJ01, (B) vTAJ13, (C) vTAJ27, (D) vTAJ30, (E) vTAJ32, (F) vTAJ34, (G) vTAJ36, and (H) vTAJ38 proteins in 50 mM phosphate buffer (pH 7.5) with 100 mM NaCl (solid line) and without NaCl (dashed line) at 4°C. [vTAJ01] = 2.0 μ M, [vTAJ13] = 9.0 μ M, [vTAJ27] = 7.3 μ M, [vTAJ30] = 3.7 μ M, [vTAJ32] = 5.4 μ M, [vTAJ34] = 5.0 μ M, [vTAJ36] = 5.1 μ M, [vTAJ38] = 9.3 μ M.

Figure 5. CD spectra of (A) wTAJ515, (B) wTAJ603, (C) wTAJ613, and (D) wTAJ615 proteins in 50 mM phosphate buffer (pH 7.5) with 100 mM NaCl at 4°C. [wTAJ515] = 8.7 μ M, [wTAJ603] = 6.7 μ M, [wTAJ613] = 7.3 μ M, [wTAJ615] = 7.0 μ M.

exactly by this kind of calculation from the CD data, the β -strand regions might not be fully organized as designed in vTAJ proteins. The α -helix and β -strand content of wTAJ proteins, specifically wTAJ515, wTAJ603, and wTAJ615 was similar to the vTAJ proteins, indicating that wTAJ515, wTAJ603, and wTAJ615 proteins have similar properties of the secondary structures as vTAJ proteins (see Fig. 5).

Thermal denaturation of vTAJ13, vTAJ27, vTAJ30, and vTAJ38 was also examined by monitoring of the CD signals at 220 nm (see Fig. 6). As the temperature increased, a gradual reduction in the signal intensity of each protein was observed, which coincided with protein unfolding. Intense signals were observed at high temperatures and the denaturation curves indicated that the proteins were stable at high temperatures. Even at temperatures approaching 100°C, vTAJ13 was still capable of folding up to 50%. Moreover, almost all proteins evaluated in this study, including wTAJ proteins were difficult to be denatured in high temperatures. This finding indicated that the approach combining hydrophobic and hydrophilic amino acids using the binary code has a potential to generate thermally stable proteins including folded and molten-globule like proteins.

8-Anilino-1-Naphthalenesulfonate binding study for selected proteins

The packing states of the $\alpha\beta\beta$ *de novo* proteins were evaluated using 8-anilino-1-naphthalenesulfonate (ANS).^{24,25} ANS itself is weakly fluorescent in aqueous

buffer. In contrast, a dramatic increase in its fluorescence intensity is normally observed when the ANS molecule binds to solvent-exposed hydrophobic regions. Hence, ANS can be used as a probe to investigate the molten-globule state of proteins. Binding of ANS to vTAJ and wTAJ proteins as well as lysozyme and apomyoglobin is shown in Figure 7. Although lysozyme, which folds in well-packed structure, could not bind to ANS, apomyoglobin binds to ANS very well, and the strong fluorescence was observed with

Figure 6. Thermal stabilities of vTAJ13 (○), vTAJ27 (□), vTAJ30 (◇), and vTAJ38 (▲) proteins in 50 mM phosphate buffer (pH 7.5) with 100 mM NaCl monitored by CD signal at 220 nm. [protein] = 5.0 μ M.

Figure 7. ANS binding assay of vTAJ and wTAJ proteins. (A) Fluorescence spectra of ANS incubated with vTAJ13, vTAJ36, vTAJ38, wTAJ515, lysozyme, and apomyoglobin. Although ANS cannot bind to lysozyme that folds in a well-packed structure, the strong fluorescence was observed in binding of ANS to apomyoglobin. [ANS] = 10 μ M and [protein] = 1.5 μ M in 50 mM phosphate buffer (pH 7.5) with 100 mM NaCl at 25°C. λ_{exc} = 390 nm. (B) Comparison of ANS binding activities of $\alpha 3\beta 3$ *de novo* proteins. $I - I_0$ represents difference from the fluorescence intensity of ANS at 500 nm in the presence of protein (I) to that in the absence of protein (I_0).

emission maxima at 470 nm [Fig. 7(A)]. In vTAJ and wTAJ proteins, comparing the fluorescence intensities of ANS at 500 nm revealed that approximately 75% of the vTAJ proteins (vTAJ13, vTAJ27, vTAJ30, vTAJ32, vTAJ34, and vTAJ38) exhibited a relatively low increase in fluorescence corresponding to ANS binding ($(I - I_0)/I_0 \leq 3$; I and I_0 represent the fluorescence intensities of ANS with and without proteins, respectively). In particular, vTAJ13, vTAJ27, and vTAJ38 proteins demonstrated minimal fluorescence ($(I - I_0)/I_0 \leq 1$), indicating that these proteins may have packed structures similar to native proteins. In contrast to vTAJ proteins, 75% of wTAJ proteins (wTAJ515, wTAJ603, and wTAJ613) exhibited the strong fluorescence of ANS suggesting molten globular protein formation. These results indicated that adequate amounts of folded $\alpha 3\beta 3$ *de novo* proteins can be obtained from very bright colonies (vTAJ) not from bright colonies (wTAJ).

Association states of selected proteins

To determine the association states of the selected $\alpha 3\beta 3$ *de novo* proteins, size-exclusion chromatography (SEC) was performed using a Superdex 75 column²⁶ (see Fig. 8). vTAJ and wTAJ proteins were analyzed by SEC, but only five vTAJ proteins (vTAJ01, vTAJ13, vTAJ27, vTAJ36, and vTAJ38) demonstrated peaks. None of the wTAJ proteins selected in this study could be analyzed using the Superdex column. The ANS binding assay indicated that some proteins have molten globule-like structures, and therefore might expose their hydrophobic amino acids on their surface. As a result, these proteins may have interacted non-specifically with the gel matrix of the dextran-based col-

umn.²⁷ Thus, peaks corresponding to the molecular weight of these proteins could not be observed. In this regard, GFP selection is a preferred method to identify the folded proteins. In contrast, the vTAJ13 protein showed a single peak corresponding to the monomer state ($M_{\text{obsd}} = 12.1$ kDa, $M_{\text{calcd}} = 10.4$ kDa). Although the observed molecular weight of vTAJ13 was determined using globular well-packed proteins as a standard, the difference between theoretical and observed values was small. This finding indicated that the vTAJ13 protein may form globular structure. The vTAJ36 protein also produced a peak thought to correspond to a monomer ($M_{\text{obsd}} = 16.1$ kDa). Because the vTAJ36 protein was thought to have a molten globule-like structure based on the ANS binding assay, its molecular volume was probably greater than that of folded globular protein such as vTAJ13. The vTAJ27 protein demonstrated a peak corresponding to the dimer state of the protein ($M_{\text{obsd}} = 23.6$ kDa). vTAJ38 also demonstrated a peak thought to correspond to the dimer state of the protein ($M_{\text{obsd}} = 21.2$ kDa); however, an ultracentrifugation study of vTAJ38 indicated that it might exist in an equilibrium between the monomer and dimer formations (Figs. S3 and S4). The association constant of the vTAJ38 protein is calculated to be $6.2 \pm 2.9 \times 10^4$ M⁻¹. The vTAJ27 and vTAJ38 proteins may have packed structures, as suggested by the ANS binding assay, and these well-folded conformations might be stabilized by dimerization. The vTAJ27 and vTAJ38 proteins have five and six Thr residues, respectively, the numbers being more than those of the other vTAJ proteins. It appears that the Thr residues positioned at the hydrophilic sites might induce intermolecular interactions to form association structures. The vTAJ01

Figure 8. Size exclusion chromatography of (A) vTAJ01, (B) vTAJ13, (C) vTAJ27, (D) vTAJ36, and (E) vTAJ38 proteins on Superdex 75 column monitored by the absorbance at 220 nm. [vTAJ01] = 2.0 μ M, [vTAJ13] = 9.0 μ M, [vTAJ27] = 7.3 μ M, [vTAJ36] = 5.1 μ M, [vTAJ38] = 24 μ M. Molecular weights of vTAJ proteins are calculated using standard proteins.

protein demonstrated two main peaks corresponding to oligomers and monomer formations. One of the peaks appeared in the void volume of the column. When the vTAJ01 protein has a similar volume of vTAJ36 showing molten-globule characteristics, the oligomers might be composed of more than four molecules. The structure of vTAJ01 protein was different from the other vTAJ proteins judged by the shapes of the CD spectra. An Arg residue at the hydrophobic position on α 2 might affect the folding and aggregation of vTAJ01. These findings suggest that designing the α 3 β 3 genetic library with a binary combination of hydrophobic and hydrophilic amino acids can generate a variety of *de novo* proteins not only with folded and molten globule-like structures but also of various association states using a selection system with GFP as a folding reporter.

NMR analysis of vTAJ13

Various proteins were obtained with the design of an α 3 β 3 genetic library. Among them, the structure of

vTAJ13 was examined by NMR spectroscopy because the above results suggest that vTAJ13 has a folded globular structure and higher thermal stability in the monomeric state than the other selected proteins. Sequence-specific assignments of the backbone and side-chain chemical shifts were identified by examining the heteronuclear multidimensional NMR spectra^{28–31} measured in 50 mM sodium acetate buffer (pH 5.0) using uniformly $^{13}\text{C}/^{15}\text{N}$ -labeled vTAJ13. As shown in Figure 9, the ^1H - ^{15}N HSQC spectrum of vTAJ13 displays well-dispersed signals, similar to native proteins. This result indicates that the vTAJ13 protein has a well-folded unique structure. Although 78 amino acids (out of a total of 103 amino acids) corresponding to backbone signals (HN, CO, C α , and/or H α) were successfully assigned (see Supp. Info.), it was difficult to identify the remaining amino acids. The linker sequence (GGHGG) exists at several positions including both ends, and each linker seems to have some flexibility. Therefore, the overlapping of signals for the Gly residues made the assignment difficult. Moreover, signals from the first α -helix were not

Figure 9. Two-dimensional ^1H - ^{15}N HSQC spectrum of ^{15}N - or $^{13}\text{C}/^{15}\text{N}$ -labeled vTAJ13 in 50 mM sodium acetate buffer (pH 5.0), showing sequential assignments of the backbone amides. [vTAJ13] = 0.9 mM.

completely assigned. It appears that the first α -helix of the design might occur in the undetectable interaction with other helices and strands or not fold properly and exist as a molten globule-like structure.

To identify regular secondary structure regions of vTAJ13, interproton-distance constraints were derived from short- and medium-range NOE connectivities, and the secondary chemical shifts ($\Delta\delta$), the observed displacements of C_α and C_β chemical shifts from their random coil values,³² were calculated [Fig. 10(A)]. In the designed α_2 and α_3 regions, $\text{H}\alpha$ - $\text{H}\beta(i, i+3)$ NOEs were clearly observed in a consecutive manner, indicating that these regions actually adopted α -helix

structures as designed. Secondary-shift analysis also supports that the α -helical conformation appears in the designed α_2 and α_3 regions as demonstrated by the positive and negative values of ΔC_α and ΔC_β , respectively. Moreover, β -strands were present at the region of β_1 , β_2 , and β_3 as designed as shown in the negative and positive values of ΔC_α and ΔC_β , respectively. Dihedral angles, ϕ and ψ , of vTAJ13 were evaluated from a database analysis of backbone ($^{13}\text{C}\alpha$, $^{13}\text{C}\beta$, ^{13}CO , $^1\text{H}\alpha$, ^{15}N) chemical shifts using the program TALOS³³ (Fig. S5). The predicted values of ϕ and ψ at the regions of α_1 , α_2 , and α_3 were around -60° and 40° , respectively, indicating that the regions

Figure 10. (A) Diagram of the NMR data used to establish the secondary structure of vTAJ13. NOE data were obtained from three- and four-dimensional spectra measured at 15°C, and the height of the bar indicates the strength of the NOE correlation. Secondary chemical shifts of C_α and C_β are displacements from their random-coil values. (B) and (C) The β-strand topologies of vTAJ13 showing long range NOEs: H_α-H_α NOE pairs (thin arrow), H_γ-H_γ NOE pairs (bold arrow), and weak H_β-H_β NOE pairs (dashed arrow).

designed as α -helix actually formed the α -helix structure. It appears the α_1 region folded into an α -helical structure, judging from the limited data obtained. The values of ϕ and ψ at designed β -strand regions (β_1 , β_2 , and β_3) calculated around -110° and $+140^\circ$, respectively, proved that these regions formed β -structures. Long-range NOEs were observed between α_2 and α_3 (H32-I54, E36-K49), indicating that the α_2 and α_3 regions could associate together in an anti-parallel orientation. The H32 residue might be buried inside the hydrophobic core and the E36 residue could interact electrostatically with K49. The inter-strand NOEs were observed between β_1 and β_2 (E65-K81, V66-I80, D67-K79, V68-I78, K69-A77), indicating that the β_1 and β_2 strands could pack into an anti-parallel β -sheet structure. Of note, two ion pairs (E65-K81, D67-K79) could be employed to stabilize the β -sheet structure between β_1 and β_2 . Although inter-strand NOEs were also observed between β_2 and β_3 , two kinds of β -sheet ori-

entations might produce this result, as shown in Figure 10(B,C). This result suggests that the β -sheets of β_2 and β_3 existed in an equilibrium between structures with different packing conditions. It seems that a proline residue at the C-terminus, which can assume both *cis* and *trans* isomers, affects the β -sheet structures at the end of the vTAJ13 molecule. Because the linker sequences (GGHGG) between β -strands might not be appropriate to form the well-packed β -sheet structure, redesign of vTAJ13 with regard to the linker and C-terminal sequences may enhance packing of the β -sheet structure. Of note, long-range NOEs were also observed between α_2 and β_1 (L35-A71, V39-L70). This finding indicating that the orientation between α_2 and β_1 was anti-parallel. Although we attempted to calculate the overall structure of vTAJ13 based on the NMR data, it was difficult to determine its 3D structure due to the insufficient amounts of NOE signals for the calculation.

Methods

All chemicals and solvents were of reagent or HPLC grade. The designed DNA fragments and primers were purchased from JBioS and Sigma Genosys. Klenow fragment, T4 DNA ligase, and KOD-Plus DNA polymerase were purchased from Toyobo and GFPuv was purchased from Clontech Laboratories, and F64L/S65T mutation was incorporated using PCR technique.

Construction of $\alpha 3\beta 3$ genetic library

Single-stranded DNAs (sense-NDE, 5'-GAGATTCAT ATGTACGGTGGACACGGAGGC-3'; sense-BSA-alpha, 5'-GAGATTCGGTCTCACACGGAGGCRMRVTHRMRR MRVTHVTHRMRRMRVTHRMRRMRVTHVTHRMRG GTGGACACGGAGGC-3'; antisense-alpha-BBS, 5'-GAGATTCGAAGACTGCGTGGCCACCYKYDABDABYK YYKYDABYKYYKYDABDABYKYYKYDABYKYGCCTCC GTGTCC ACC-3'; sense-BSA-beta, 5'-GAGATTCGGTC TCACACGGAGGCRMRVTHRMVTHRMVTHRMVTHRMRG GTGGACACGGAGGC-3'; antisense-beta-BBS, 5'-GAG ATTCGAAGACTGCGTGGCCACCYKYDABYKYDAMYK YDABYKYGCCTCCGTGTCCACC-3'; antisense-KPN, 5'-GAGATTCGGTACCGAGCCTCCGTGTCCACC-3') were elongated to double stranded DNAs by combinations of sense-NDE and antisense-alphaBBS, sense-BSA-alpha and antisense-alpha-BBS, sense-BSA-beta and antisense-beta-BBS, and sense-BSA-beta and antisense-KPN using Klenow enzyme. DNAs were digested by restriction enzymes (*Bsa*I and *Bbs*I), and then the digested DNAs were ligated using T4 DNA ligase. Iterative reactions involving restriction and ligation generated the DNA library coding $\alpha 3\beta 3$ gene. The DNA library was amplified by PCR (nine cycle) using KOD plus enzyme and primers (primer-N, 5'-GAGATTCCTA TATGTACGG-3'; primer-C, 5'-GAGATTCGGTACC GAG-3'), and then the product was digested by *Nde*I and *Kpn*I. A vector (pET-GFP) was constructed according to the reported method with some modifications.²⁰ The DNA library was inserted into the pET-GFP vector at the *Nde*I and *Kpn*I restriction site, and XL1-Blue *E. coli* cells (Stratagene) were transformed with the library plasmids on LB agar plates containing 34 μ g/mL kanamycin. Colonies were pooled and plasmids were extracted using a plasmid purification kit (Qiagen Plasmid Midi Kit).

Screening of $\alpha 3\beta 3$ de novo proteins

To select the folded $\alpha 3\beta 3$ de novo proteins, Rosetta (DE3)pLysS (Novagen) *E. coli* cells for protein expression were transformed with the purified plasmids coding $\alpha 3\beta 3$ de novo protein gene fused with *gfp*. Obtained colonies on the LB plates containing 34 μ g/mL kanamycin were covered with nitrocellulose membranes (Millipore) for 1 h, and then the membranes were transferred into the LB plates containing 34 μ g/mL kanamycin and 1 mM isopropyl-1-thio- β -D-galactopyranoside (IPTG) for the induction of protein synthesis. The plates were

incubated at 37°C for 3 h. Green fluorescence colonies were picked by UV illumination around 366 nm using a UV lamp (MODEL UVGL-25; UVP, Inc.) as comparing to the colonies having GFP without de novo protein gene. The colonies were categorized: (1) very bright colonies which emit the fluorescence as strong as wild-type GFP; (2) bright colonies which emit weaker fluorescence than wild-type GFP; (3) colorless colonies. From each colony, plasmid DNA was extracted using a plasmid purification kit (Qiagen Miniprep Kit). The plasmids were digested by *Nde*I and *Kpn*I, and then the fragments coding $\alpha 3\beta 3$ de novo proteins were ligated into a vector without the *gfp* gene.

Expression and purification of $\alpha 3\beta 3$ de novo proteins

The plasmids coding $\alpha 3\beta 3$ de novo proteins without the *gfp* gene from very bright and bright colonies were transformed into Rosetta(DE3)pLysS. LB medium (5 mL supplemented with 34 μ g/mL kanamycin was inoculated with a single colony of each Rosetta(DE3)-pLysS cell and incubated overnight at 37°C. LB medium (100 mL) supplemented with 34 μ g/mL kanamycin was inoculated using 0.2 mL of the overnight culture. The cells were grown at 37°C to an OD 600 of 0.6 and then induced by addition of IPTG to a final concentration of 1.0 mM. The culture solutions were then incubated at 30°C for 5 h. The cells were collected by centrifugation, and the pellets were stored in -80°C. The cell pellets were resuspended in a buffer (20 mM Tris-HCl, 0.5M NaCl, pH 8.0), and then the cells were broken by sonication. The resulting solids, which are inclusion bodies, were collected by centrifugation and washed with PBS containing 4% Triton X-100 and water.

The purified inclusion bodies were dissolved with 8M urea aqueous solution. After centrifugation, the solution was purified by RP-HPLC as a linear gradient of acetonitrile/0.1% TFA. Although the soluble fractions included the de novo proteins, it was easy to purify the proteins from the insoluble fractions, because the proteins did not have any tag sequence such as His6 and Flag sequences. Each protein was identified by MALDI TOF-MS: m/z found (calcd.); vTAJ13, 10401.0 [M+H]⁺ (10396.8); vTAJ27, 10418.1 [M+Na]⁺ (10415.6); vTAJ30, 10167.0 [M⁺Na]⁺ (10163.6); vTAJ32, 10347.2 [M+H]⁺ (10343.9); vTAJ34, 10257.3 [M+H]⁺ (10253.9); vTAJ36, 10407.1 [M+H]⁺ (10403.8); vTAJ38, 10281.1 [M+H]⁺ (10281.6); wTAJ603, 10159.9 [M+H]⁺ (10159.8); wTAJ613, 10316.2 [M+H]⁺ (10317.9); wTAJ615, 10394.6 [M+H]⁺ (10398.1).

Refolding of $\alpha 3\beta 3$ de novo proteins

The lyophilized powder of each protein was dissolved in a denaturing buffer (6M guanidine hydrochloride, 50 mM sodium phosphate, pH 7.5, with or without 100 mM NaCl). The step-wise dialysis using a dialysis

membrane (3500 MWCO; Spectrum) with decreasing the guanidine concentration (6, 2, 1, 0.5, and 0 mM) was performed to refold the protein. The concentrations of the *de novo* proteins were determined by amino acid analysis using phenyl isothiocyanate (PTC) method on a Wakopak WS-PTC column (Wako Chemical).

CD measurement of $\alpha\beta\beta$ *de novo* proteins

CD spectra of the $\alpha\beta\beta$ *de novo* proteins were measured on a Jasco J-720WI spectropolarimeter using a quartz cell with 1.0 or 2.0 mm pathlength in the region of 195–250 nm. Thermal stability of the secondary structures of the proteins was monitored by the CD signals recorded at 220 nm from 4 to 100°C. Temperature increased 1°C/min. Estimation of the secondary structures of the proteins was performed using a computer program SELCON3.

ANS binding assay of $\alpha\beta\beta$ *de novo* proteins

Fluorescence spectra were collected on a Hitachi F2500 fluorescence spectrophotometer using a 5 mm $\times$ 5 mm quartz cell. 8-anilino-1-naphthalene sulphonate (ANS; 10 μ M) and each protein (1.5 μ M) were mixed and incubated in 50 mM phosphate buffer (pH 7.5) containing 100 mM NaCl at room temperature for 30 min. Fluorescence spectra were measured (λ_{ex} = 390 nm), and the fluorescence intensities at 500 nm were collected.

SEC analysis of $\alpha\beta\beta$ *de novo* proteins

SEC was performed using Superdex 75 10/300 GL column (Amersham Bioscience) on a Shimadzu LC-2010C HPLC system with 50 mM phosphate buffer (pH 7.5) containing 100 mM NaCl. Following proteins were used as molecular weight standards; BSA (66 kDa), carbonic anhydrase (29 kDa), cytochrome c (12.4 kDa), aprotinin (6.5 kDa), and insulin B chain (3.5 kDa).

NMR analysis of vTAJ13 protein

Uniformly ^{15}N - and $^{13}\text{C}/^{15}\text{N}$ -labeled vTAJ13 were prepared by growing cells in M9 minimum media supplemented with 1 g/L [^{15}N]NH₄Cl and 2 g/L glucose or [^{13}C]glucose. Protein was purified and refolded as described above. Refolded protein solution was concentrated using Amicon Ultra-15 centrifugal filter device (Millipore). NMR samples contained about 0.9 mM ^{15}N - or $^{13}\text{C}/^{15}\text{N}$ -labeled protein in 50 mM sodium acetate buffer, pH 5.0.

All NMR spectra were recorded at 15°C on a Bruker DMX750 spectrometer equipped with a 5 mm inverse triple-resonance probehead with three-axis gradient coils. ^1H , ^{13}C , and ^{15}N sequential resonance assignments were obtained using 2D double resonance, and 3D double and triple resonance through-bond correlation experiments: 2D ^1H - ^{15}N HSQC, 2D ^1H - ^{13}C HSQC, 3D HNCO, 3D CNCA(CO)NH, 3D

HNCA, 3D HCABGCO, and 3D HCCH-TOCSY.^{28–31} Interproton distance information were derived from multidimensional nuclear overhauser effect (NOE) spectra with mixing times of 100 ms: 3D ^{15}N -separated NOESY-HSQC, 3D $^{13}\text{C}/^{15}\text{N}$ -separated NOESY-HSQC, and 4D $^{13}\text{C}/^{13}\text{C}$ -separated NOESY-HSQC.^{29–31} All spectra were processed using NMRPipe software³⁴ and analyzed using Capp/Pipp/Stapp software.³⁵ ^1H , ^{13}C , and ^{15}N chemical shifts were referenced, respectively, to mono-deuterated water (4.87 ppm at 15°C), indirectly to sodium 3-(trimethylsilyl)-propionate (^{13}C),³⁶ and to liquid ammonia (^{15}N).³⁷

Conclusions

In conclusion, we have successfully constructed and screened *de novo* proteins from a genetic library designed to encode $\alpha\beta\beta$ *de novo* proteins. The GFP fusion system enabled not only the easy elimination of unexpected products from genes of single base insertion or deletion, but also the emergence of the folded proteins. In the design strategy of the genetic library, we used the binary combination. If a protein cannot fold well, a lot of hydrophobic amino acids exposed could make aggregation and then result in disturbing the GFP folding. Therefore, selected proteins can fold with burying the hydrophobic amino acids by three-dimensional interactions. The selected proteins contained both the α -helix and β -strand, and some of the proteins showed native-like properties. Furthermore, the selected proteins possess different properties such as structures and association states. We utilized Ala, positioned at the hydrophilic sites of helices and strands as designed, as a modulator in the hydrophobic interaction. The Ala residues in vTAJ and wTAJ proteins might modulate the interactions between helix–helix, helix–strand, and strand–strand to generate a variety of three-dimensional structures. Thr was also utilized to construct the proteins, and might bring the association abilities for vTAJ27 and vTAJ38 proteins. The vTAJ proteins isolated from the very bright colonies showing similar fluorescence to GFP tend to have packed structures to which ANS is unable to bind. Although these selected proteins have many hydrophobic amino acids, the proteins can form the tertiary structures by the hydrophobic packing. Thus, the designed combinatorial library coding both α -helices and β -strands has the potential to generate *de novo* proteins with folding properties with artificial scaffolds using a screening system. The possibility will need to be proved by detailed analyses using NMR studies. These selected small proteins possess stable structures even when the proteins fold using both α -helix and β -strand as the secondary structure components. Especially, vTAJ13 protein is monomeric and might have a globular structure. The NMR structural data indicated that vTAJ13 has the three-dimensional structure and predominantly forms the $\alpha\beta\beta$ structure as intended by our design, although determination of its complete

structure was hampered due to the limited separation of NOE signals and the loss of NMR signals for the majority of the $\alpha 1$ region that might not fold fully into a specific structure. Like vTAJ13, the other *de novo* proteins selected in this study might have folded in well-packed structures. Of note, vTAJ proteins can be engineered to function such as catalytic and binding properties for some targets because the proteins are not only folded but also small and stable without any aromatic amino acids except vTAJ38. It appears that the linker sequences and the surface residues positioned at the hydrophilic sites on helices and strands are available to be engineered. Using the selected proteins as scaffolds, evolution of these selected proteins will ultimately lead to the generation of novel functional proteins in the future.

Acknowledgments

The authors thank Prof. Fumio Arisaka, Tokyo Institute of Technology, Japan, for helping with ultracentrifugation analysis.

References

- Lerner RA (2006) Manufacturing immunity to disease in a test tube: the magic bullet realized. *Angew Chem Int Ed* 45:8106–8125.
- Kehoe JW, Kay BK (2005) Filamentous phage display in the new millennium. *Chem Rev* 105:4056–4072.
- Rothe A, Hosse RJ, Power BE (2006) *In vitro* display technologies reveal novel biopharmaceutics. *FASEB J* 20:1599–1610.
- Seelig B, Szostak JW (2007) Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* 448:828–831.
- Luginbühl B, Kanyo Z, Jones RM, Fletterick RJ, Prusiner SB, Cohen FE, Williamson RA, Burton DR, Plückthun A (2006) Directed evolution of an anti-prion protein scFv fragment to an affinity of 1 pM and its structural interpretation. *J Mol Biol* 363:75–97.
- Graziano JJ, Liu W, Perera R, Geierstanger BH, Lesley SA, Schultz PG (2008) Selecting folded proteins from a library of secondary structural elements. *J Am Chem Soc* 130:176–185.
- Baker D (2006) Prediction and design of macromolecular structures and interactions. *Phil Trans R Soc B* 361:459–463.
- Jin W, Kambara O, Sasakawa H, Tamura A, Takada S (2003) *De novo* design of foldable proteins with smooth folding funnel: automated negative design and experimental verification. *Structure* 11:581–590.
- Isogai Y, Ito Y, Ikeya T, Shiro Y, Ota M (2005) Design of λ Cro fold: solution structure of a monomeric variant of the *de novo* protein. *J Mol Biol* 354:801–814.
- Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* 302:1364–1368.
- Kaplan J, DeGrado WF (2004) *De novo* design of catalytic proteins. *Proc Natl Acad Sci USA* 101:11566–11570.
- Calhoun JR, Natri F, Maglio O, Pavone V, Lombardi A, DeGrado WF (2005) Artificial diiron proteins: from structure to function. *Biopolymers* 8:264–278.
- Hecht MH, Das A, Go A, Bradley LH, Wei Y (2004) *De novo* proteins from designed combinatorial libraries. *Protein Sci* 13:1711–1723.
- Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH (1993) Protein design by binary patterning of polar and nonpolar amino acids. *Science* 262:1680–1685.
- Wei Y, Liu T, Sazinsky SL, Moffet DA, Pelczer I, Hecht MH (2003a) Stably folded *de novo* proteins from a designed combinatorial library. *Protein Sci* 12:92–102.
- Wei Y, Hecht MH (2004) Enzyme-like proteins from an unselected library of designed amino acid sequences. *Protein Eng Des Sel* 17:67–75.
- Wei Y, Kim S, Fela D, Baum J, Hecht MH (2003b) Solution structure of a *de novo* protein from a designed combinatorial library. *Proc Natl Acad Sci* 100:13270–13273.
- Wang W, Hecht MH (2002) Rationally designed mutations convert *de novo* amyloid-like fibrils into soluble monomeric β -sheet proteins. *Proc Natl Acad Sci* 99:2760–2765.
- Walter K, Vamvaca K, Hilvert D (2005) An active enzyme constructed from a 9-amino acid alphabet *J Biol Chem* 280:37742–37746.
- Waldo GS, Standish BM, Berendzen J, Terwilliger TC (1999) Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol* 17:691–695.
- Cabantous S, Pédelacq J-D, Mark BL, Maranjo C, Terwilliger TC, Waldo GS (2005) Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. Application to improving the folding and solubility of recalcitrant proteins from *Mycobacterium tuberculosis*. *J Struct Funct Genomics* 6:113–119.
- Wurth, C., Guimard NK, Hecht MH (2002) Mutations that reduce aggregation of the Alzheimer's A β 42 peptide: an unbiased search for the sequence determinants of A β amyloidogenesis. *J Mol Biol* 319:1279–1290.
- Sreerama N, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem* 287:252–260.
- Matulis D, Baumann CG, Bloomfield VA, Lovrien RE (1999) 1-anilino-8-naphthalene sulfonate as a protein conformational tightening agent. *Biopolymers* 49:451–458.
- Betz SF, DeGrado WF (1996) Controlling topology and native-like behavior of *de novo*-designed peptides: design and characterization of antiparallel four-stranded coiled coils. *Biochemistry* 35:6955–6962.
- MacBeath G, Kast P, Hilvert D (1998) A small, thermostable, and monofunctional chorismate mutase from the Archeon *Methanococcus jannaschii*. *Biochemistry* 37:10062–10073.
- Hepler RW, Grimm KM, Nahas DD, Breese R, Dodson EC, Acton P, Keller PM, Yeager M, Wang H, Shughrue P, Kinney G, Joyce JG (2006) Solution state characterization of amyloid β -derived diffusible ligands. *Biochemistry* 45:15157–15167.
- Clore GM, Gronenborn AM (1991) Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* 252:1390–1399.
- Clore GM, Gronenborn AM (1998) Determining the structures of large proteins and protein complexes by NMR. *Trends Biotechnol* 16:22–34.
- Bax A, Grzesiek S (1993) Methodological advances in protein NMR. *Acc Chem Res* 26:131–138.
- Yamazaki T, Nicholson LK, Torchia DA, Wingfield P, Stahl SJ, Kaufman JD, Eyermann CJ, Hodge CN, Lam PYS, Ru Y, Jadhav PK, Chang C-H, Weber PC (1994)

- NMR and X-ray evidence that the HIV protease catalytic aspartyl groups are protonated in the complex formed by the protease and a non-peptide cyclic urea-based inhibitor. *J Am Chem Soc* 116:10791–10792.
32. Spera S, Bax A (1991) Empirical correlation between protein backbone conformation and C α and C β ^{13}C nuclear magnetic resonance chemical shifts. *J Am Chem Soc* 113: 5490–5492.
 33. Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13:289–302.
 34. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6: 277–293.
 35. Garrett DS, Powers R, Gronenborn AM, Clore GM (1991) A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J Magn Reson* 95:214–220.
 36. Bax A, Subramanian S (1986) Sensitivity-enhanced two-dimensional heteronuclear shift correlation NMR spectroscopy. *J Magn Reson* 67:565–569.
 37. Live DH, Davis DG, Agosta WC, Cowburn D (1984) Long range hydrogen bond mediated effects in peptides: ^{15}N NMR study of gramicidin S in water and organic solvents. *J Am Chem Soc* 106:1939–1941.