Doctoral Dissertation (Shinshu University)

Self-assembling nanostructures created from protein nanobuilding blocks using the intermolecular folding structure of a dimeric *de novo* protein.

March 2017

Naoya Kobayashi

Table of contents

Table of contents 1

Abbreviations 2

Summary 3

Preface 5

Chapter 1 Domain-Swapped Dimeric Structure of a Stable and Functional *De Novo* Four-Helix Bundle Protein, WA20 11

1.1 Introduction 11

1.2 Materials and methods 15

1.3 Results and Discussion 23

Chapter 2 Self-Assembling Nano-Architectures Created from a Protein Nano-Building Block Using an Intermolecularly Folded Dimeric *De Novo* Protein 41

2.1 Introduction 41

2.2 Materials and methods 45

2.3 Results and Discussion 55

Chapter 3 Self-Assembling Supramolecular Nanostructures Created by *de Novo* Extender Protein Nanobuilding Blocks 76

3.1 Introduction 76

3.2 Materials and methods 80

3.3 Results and Discussion 86

Chapter 4 Conclusions and Future Directions 114

References 118

List of Publications 131

Acknowledgements 132

Abbreviations

AUC (analytical ultracentrifugation)

CD (circular dichroism)

ePN-Block (extender Protein Nano-building Block)

esPN-Block (extender-stopper Protein Nano-building Block)

DSF (Differential Scanning Fluorimetry)

DTT (dithiothreitol)

FL (Flexible Linker)

FM-AFM (Frequency Modulation Atomic Force Microscopy)

GdnHCl (guanidine hydrochloride)

HL (Helical Linker)

IFT (indirect Fourier transformation)

IMAC (Immobilized Metal ion Affinity Chromatography)

MAD (Multiwavelength Anomalous Dispersion)

MALDI-TOF (Matrix Assisted Laser Desorption/Ionization-Time of Flight)

MALS (multiangle light scattering)

NMR (Nuclear Magnetic Resonance)

PAGE (Polyacrylamide Gel Electrophoresis)

PCR (Polymerase Chain Reaction)

PDB (Protein Data Bank)

PN-Block or PNB (Protein Nano-building Block)

SAXS (Small-Angle X-ray Scattering)

SDS (Sodium Dodecyl Sulfate)

SEC (Size Exclusion Chromatography)

SeMet (Selenomethionine)

sPN-Block (stopper Protein Nano-building Block)

2D (two-dimensional)

3D (three-dimensional)

Summary

Living organisms are maintained by various self-assembling biomolecules including proteins, nucleic acids, sugars, and lipids. The chemical reconstitution of living matter is one of the ultimate goals of chemistry and synthetic biology. Rational design of artificial biomacromolecules that self-assemble into supramolecular complexes is an important step toward achieving the goal. Proteins are the most versatile biomacromolecules performing the complex and functional tasks in the living organisms. Therefore, design of novel proteins that self-assemble into supramolecular nanostructures is an important step in the development of synthetic biology and nanotechnology. However, *de novo* design of proteins and protein complexes is very difficult because structures of protein complexes are cooperatively stabilized by the contribution of many interactions throughout an amino-acid chain making a protein.

The purpose of this thesis is to create various self-assembling supramolecular nanostructures from protein nanobuilding blocks (PN-Blocks) using a characteristic dimeric *de novo* protein as a component of PN-Block.

In chapter 1, I describe the crystal structure of a stable and functional *de nov*o protein, WA20, used in this study. The WA20 crystal structure is not a monomeric four-helix bundle originally designed by binary patterning of polar and nonpolar residues, but a dimeric four-helix bundle. Each monomer comprises two long α-helices that intertwist with the helices of the other monomer. The two monomers together form a 3D domain-swapped four-helix bundle dimer. Small-angle X-ray scattering shows that the molecular weight of WA20 is ~25 kDa and the shape is rod-like, indicating that WA20 forms a dimeric four-helix bundle in solution.

In chapter 2, to harness the unusual intermolecular folding structure of the WA20 for the self-assembly of supramolecular nanostructures, I created a protein nanobuilding block (PN-Block), called WA20-foldon, by fusing the dimeric structure of WA20 to the trimeric foldon domain of fibritin from bacteriophage T4 as the first series of PN-Blocks. The WA20-foldon fusion protein was expressed in the soluble fraction in *Escherichia coli*, purified, and shown to form several homooligomeric forms. The stable oligomeric forms were further purified and characterized by a range of biophysical techniques. Size exclusion chromatography, multiangle light scattering, analytical ultracentrifugation, and small-angle X-ray scattering (SAXS) analyses indicate that the small (S form), middle (M form), and large (L form) forms of the WA20-foldon

oligomers exist as hexamer (6-mer), dodecamer (12-mer), and octadecamer (18-mer), respectively. These findings suggest that the oligomers in multiples of 6-mer are stably formed by fusing the interdigitated dimer of WA20 with the trimer of foldon domain. Pair-distance distribution functions obtained from the Fourier inversion of the SAXS data suggest that the S and M forms have barrel- and tetrahedron-like shapes, respectively. These results demonstrate that the *de novo* WA20-foldon is an effective building block for the creation of self-assembling artificial nanoarchitectures.

In chapter 3, to construct self-assembling extended chain-like nanostructures, I designed *de novo* extender protein nanobuilding blocks (ePN-Blocks) as the second series of PN-Blocks. The ePN-Blocks were constructed by fusing tandemly two *de novo* binary-patterned WA20 proteins with various linkers. The ePN-Block proteins with the long helical linkers or flexible linkers were expressed well in soluble fractions in *Escherichia coli*. The purified ePN-Blocks migrated as ladder bands in native PAGE, suggesting that the ePN-Blocks form several homooligo-meric states in the soluble fraction. Then, I reconstructed heteromeric complexes from extender and stopper PN-Blocks by denaturation and refolding. Size exclusion chromatography-multiangle light scattering and small-angle X-ray scattering analyses suggest that extender and stopper PN-Block (esPN-Block) complexes formed different types of extended chain-like nanostructures depending on their linker types. Moreover, atomic force microscopy imaging in liquid revealed that the esPN-Block complexes with metal ion further self-assembled into supramolecular nanostructures on mica surface.

In this thesis, I describe successful design and development of various self-assembling supramolecular nanostructures constructed from a few types of simple and basic PN-Blocks using the intermolecular folding structure of the *de novo* protein as their component. These results demonstrate that the PN-Block strategy is a powerful strategy to create various self-assembling supramolecular nanostructures.

Preface

Various self-assembling biomolecules including proteins, nucleic acids, sugars, and lipids organize living systems. The chemical reconstitution of living systems is an ultimate goal of chemistry and synthetic biology. Rational design of self-assembling artificial biomacromolecules is a key challenge for developments in synthetic biology and nanobiotechnology.

In recent years, DNA origami has been developed for the design and synthesis of various supramolecular nanostructures. In these applications, DNA base complementarity can be exploited in the rational design of artificial nanostructures with versatile two-dimensional (2D) and three-dimensional (3D) shapes, such as polyhedra (Ke, 2014). However, nucleic acids generally comprise the bases A, T, G, and C, and the ensuing limitations on numbers of combinations and chemical features may confine the potential to produce molecules with advanced functions. In addition, DNA nanostructures require thermal annealing in vitro and stabilization of electrostatic repulsion by addition of metal ions. Because such specific conditions are necessary, DNA nanostructures are not suitable for reconstruction of living systems under physiological conditions.

In contrast with DNA, proteins comprise 20 types of amino acids, allowing greater diversity of chemical properties. Accordingly, the enormous numbers of possible sequence combinations expand the probabilities to create diverse and advanced functions. Natural proteins perform various complex tasks in vivo, reflecting the spontaneous formation of intricate and refined tertiary and quaternary structures with versatile chemical properties and functionalities. There are four hierarchical levels of protein structures. The amino acid sequence of a protein's polypeptide chain is called its primary structure. The secondary structure can take the local regular form either of α-helices or of β-strands. In globular form of proteins, elements of α-helices and/or β-sheets as well as loops are folded into a tertiary structure. Many proteins are formed by self-assembling the folded chains of more than one polypeptide; this constitutes the quaternary structure of a protein. The complex and refined quaternary structures create versatile functionalities of proteins.

The design of *de novo* proteins is substantially very complicated to explore enormous amino-acid sequence space because the contribution of many cooperative and long-range interactions causes a significant gap between the primary structure and the tertiary and quaternary structure. *De novo* protein design and engineering have been performed with mainly two motivations: (1) recapitulation of natural systems to ultimately test our understanding of the principles of protein structure and function and (2) construction of tailor-made proteins as an essential step toward applied biotechnology. Research on *de novo* protein design has progressed toward the construction of novel proteins emanated mainly from three approaches: (1) rational and computational design (Bradley et al., 2005; Dahiyat and Mayo, 1997; Hecht et al., 2004; Huang et al., 2016; Kamtekar et al., 1993; Koga et al., 2012; Kuhlman et al., 2003), (2) combinatorial methods (Keefe and Szostak, 2001), and (3) semirational approaches, including elements of both rational design and combinatorial methods (Hecht et al., 2004; Kamtekar et al., 1993).

The design of self-assembling nanostructures using proteins as building blocks is an important step to achieve the goal of reconstituted living systems. In recent years, to design and create artificial self-assembling protein complexes, several biotechnological strategies using artificial and fusion proteins as nanoscale building blocks have been developed as described below.

Fusion proteins designed for symmetrical self-assembly. In natural protein complexes (Ahnert et al., 2015; Pieters et al., 2016), nanostructures self-assemble into symmetric polyhedral shapes, such as tetrahedrons, hexahedrons, octahedrons, dodecahedrons, and icosahedrons. Hence, symmetry provides a powerful tool for building large regular objects, and by Padilla *et al* described a foresighted general strategy for using symmetry to construct protein nanomaterials as “nanohedra” (Padilla et al., 2001). In this strategy, a protein that naturally forms a self-assembling oligomer is fused rigidly to another protein that forms another self-assembling oligomer, and the fusion protein self-assembles with other identical copies of itself into a designed nanohedral particle or material. The nanohedra strategy allows for construction of a wide variety of potentially useful protein-based materials. Accordingly, crystal structures of designed nanoscale protein cages have been solved (Lai et al., 2012a; Lai et al., 2013), and structures of a designed protein cage that self-assembles into a highly porous cube was reported (Lai et al., 2014) by Yeates and colleagues. In addition, Sinclair *et al.* previously generated protein lattices by fusing proteins with matching rotational symmetry (Sinclair et al., 2011). These strategies are described in detail in previous reviews (King and Lai, 2013; Lai et al., 2012b; Yeates et al., 2016).

Three-dimensional domain-swapped oligomers. Three-dimensional (3D) domain swapping involves exchanging one structural domain of a protein monomer with that of the identical domain from a second monomer, resulting in an intertwined oligomer (Bennett et al., 1995). As a pioneering work, Ogihara *et al.* designed and constructed artificial domain-swapped proteins that formed dimers and fibrous oligomers, and they proposed design principles of 3D domain-swapped protein oligomers for biomaterials (Ogihara et al., 2001). Moreover, Hirota and colleagues described cytochrome *c* polymerization following successive domain swapping (Hirota et al., 2010) and recently reported rational design of heterodimeric proteins using domain swapping for myoglobin (Lin et al., 2015b) and a nanocage encapsulating a Zn-SO_4 cluster in the internal cavity of a domain-swapped cytochrome cb_{562} dimer (Miyamoto et al., 2015).

Self-assembling designed coiled-coil peptide modules. Alpha-helical coiled coils are ubiquitous protein–protein interaction domains wherein folding and assembly of amphipathic α-helices directs of the production of multiple protein assemblies (Woolfson et al., 2012). Various self-assembling nanostructures, including fibers (Sharp et al., 2012), cyclized assemblies (Boyle et al., 2012), cages (Fletcher et al., 2013), and nanotubes (Burgess et al., 2015; Thomas et al., 2016) were constructed using designed coiled-coil peptide modules as building blocks. In addition, Sciore *et al.* recently reported a flexible, symmetry-directed approach to assembling protein cages using short, *de novo*-designed coiled-coil domains to mediate assembly (Sciore et al., 2016).

Metal-directed self-assembling engineered proteins. Metal ions are frequently found in natural protein–protein interfaces, where they stabilize quaternary or supramolecular protein structures, mediate transient protein–protein interactions, and serve as catalytic centers. Moreover, coordination chemistry of metal ions has been increasingly used to engineer and control the assembly of functional supramolecular peptide and protein architectures. In particular, Tezcan and colleagues developed design strategies of metal-directed protein self-assembly and metal-templated interface redesign (MeTIR) (Bailey et al., 2016; Salgado et al., 2010; Sontz et al., 2014). Using these strategies, a building block protein was designed and engineered to form homodimers bearing interfacial Zn-coordination sites, which enabled Zn-mediated self-assembly into 1D helical nanotubes and 2D and 3D crystalline arrays (Brodin et al., 2012). In addition, previous studies describe self-assembly of a Zn-binding protein cryptand via templated disulfide bonds (Medina-Morales et al., 2013) and a copper-inducible ferritin cage assembly by re-engineering of protein interfaces (Huard et al., 2013). Bai *et al.* also reported highly ordered self-assembling protein nanorings from engineered glutathione *S*-transferase with properly-placed metal-coordination motifs (Bai et al., 2013). Recently, Tezcan and colleagues expanded strategies that use metal-coordination by developing designed functional assembly of a metalloprotein with in vivo β-lactamase activity (Song and Tezcan, 2014), designed helical protein nanotubes with variable diameters from a single building block (Brodin et al., 2015) [32], a metal organic framework with spherical protein nodes (rational chemical design of 3D protein crystals) (Sontz et al., 2015), an allosteric metalloprotein assembly with strained disulfide bonds (Churchfield et al., 2016), and self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals (Suzuki et al., 2016).

Computationally designed self-assembling *de novo* proteins. Baker and colleagues recently advanced the Rosetta structure prediction and *de novo* design principles of proteins (Koga et al., 2012; Leaver-Fay et al., 2013; Lin et al., 2015a) and developed a general approach for designing computationally self-assembling protein nanomaterials with atomic-level accuracy (King et al., 2014; King et al., 2012) . Their approach involves docking of protein building blocks in a target symmetric architecture followed by the design of a low-energy protein–protein interface that drives the symmetry of self-assembly using RosettaDesign calculations (DiMaio et al., 2011). This method can be applied to various symmetric architectures, including protein arrays and complexes that extend in 1, 2, or 3 dimensions. These investigators also achieved computational design of a self-assembling icosahedral nanocage from 60-subunit trimeric protein building bloc (Hsia et al., 2016) and megadalton-scale icosahedral protein complexes from two-component 120-subunit proteins (Bale et al., 2016) with atomic-level accuracy. Moreover, there have been reports on parametric design of helical bundles with high thermodynamic stability (Huang et al., 2014), ordered two-dimensional arrays that are mediated by noncovalent protein–protein interfaces (Gonen et al., 2015), and *de novo* protein homo-oligomers with modular hydrogen-bond network-mediated specificity (Boyken et al., 2016). Current efforts to construct *de novo* proteins with ideal backbone arrangements have also led to the design of repeated proteins with idealized units and internal symmetry (Brunette et al., 2015; Doyle et al., 2015; Huang et al., 2015; Park et al., 2015; Rämisch et al., 2014). In addition, DeGrado and colleagues described computational design of virus-like protein assemblies on carbon nanotube surfaces (Grigoryan et al., 2011), computational design of a protein crystal (Lanci et al., 2012), and *de novo* design of a functional transmembrane protein with a Zn^{2+}-transporting four-helix bundle (Joh et al., 2014). Voet *et al.* also reported computationally designed symmetrical β-propeller proteins (Voet et al., 2014) and biomineralization of a cadmium chloride nanocrystal using a designed symmetrical protein (Voet et al., 2015). Furthermore, Woolfson and colleagues described computational design of water-soluble α-helical barrels (Thomson et al., 2014) and recently installed hydrolytic activity into a *de novo* α-helical barrel comprising seven helices with cysteine–histidine–glutamate catalytic triads (Burton et al., 2016). More studies of *de novo* protein design are included in recent reviews (Huang et al., 2016; Woolfson et al., 2015).

Although the above-mentioned nanostructure construction strategies are useful, it is still a long way to build living systems. In order to reconstruct living systems with diverse proteins, we need a strategy to generate a lot of protein nanostructures more efficiently. To strategically design various protein nanostructures, I adopted the idea derived from mathematics and geometry as a design philosophy. A typical geometric symmetrical structure forms a regular polyhedron. There are five finite convex regular polyhedra known as the Platonic solids, i.e., tetrahedron, cube, octahedron, dodecahedron, and icosahedron. In these Platonic solids, tetrahedron, hexahedron, and dodecahedron have three frames for one vertex as a common structural feature. Focusing on this geometrical feature, design of a basic structure with three frames for one vertex enable to efficiently produce geometric nanostructures. Therefore, I designed and constructed this common basic structure using a simple and symmetrical rod-like structure of a dimeric *de novo* protein as an edge of the geometric structure.

In this thesis, the purpose is to create various self-assembling supramolecular nanostructures from protein-nanobuilding blocks (PN-Blocks) using a characteristic dimeric *de novo* protein as a component of PN-Block. In chapter 1, I describe the crystal structure of a stable and functional *de nov*o protein, WA20, used in this study. I present the novel structure and discuss its unusual feature. In chapter 2, I describe the design and construction of the WA20-foldon fusion protein as a novel protein nanobuilding block (PN-Block) and demonstrate its characteristic self-assembling nano-architectures. In chapter 3, I report the *de novo* extender protein nanobuilding blocks (ePN-Blocks), constructed by fusing tandemly two WA20 proteins with various linkers, as a new series of PN-Blocks for self-assembling extended and cyclized chain-like nanostructures. Moreover, I demonstrate reconstruction of quaternary structural heteromeric complexes from extender and stopper PN-Blocks by denaturation and refolding. These studies demonstrate that the PN-Block strategy is a powerful strategy to create various self-assembling supramolecular nanostructures.

Chapter 1 Domain-Swapped Dimeric Structure of a Stable and Functional *De Novo* Four-Helix Bundle Protein, WA20

1.1 Introduction

The field of *de novo* protein design and engineering is motivated by two considerations: (i) Recapitulation of natural systems to ultimately test our understanding of biological systems including protein structure and function, and (ii) construction of novel, "tailor-made" proteins as an essential step toward future advances in biotechnology and synthetic biotechnology. Progress toward the construction of novel proteins emanates mainely from two approaches: rational design (Dahiyat and Mayo, 1997; Kuhlman et al., 2003) and combinatorial methods (Keefe and Szostak, 2001). Hecht and colleagues have developed a semirational approach that incorporates elements of both rational design and combinatorial methods to produce focused libraries of novel proteins (Hecht et al., 2004; Kamtekar et al., 1993). In these libraries, the hydrophobic or hydrophilic nature of each amino acid side chain is rationally designed on the basis of the template structure of a globular protein, but the exact identities of individual polar and nonpolar residues are varied combinatrially. Using this binary code strategy, several libraries of *de novo* α-helical or β-sheet proteins have been constructed. (Bradley et al., 2005; Hecht et al., 2004; Kamtekar et al., 1993; Wei et al., 2003b; West et al., 1999) For example, the binary patterned design of amphipathic α-helical sequences places a hydrophobic amino acid every three or four residues in accordance with its secondary structure periodicity of 3.6 residues/turn, thereby generating the following pattern: ○●○○●●○○●○○●●○, where ○ and ● represent polar and nonpolar residues, respectively. (Hecht et al., 2004; Kamtekar et al., 1993) When four such helices are linked, hydrophobic effect drives them to form a four-helix bundle with nonpolar residues forming a hydrophobic protein core and polar residues oriented the aqueous solvent.

Using this approach, three libraries of binary patterned four-helix bundles have been constructed. The first library encoded 74-residue sequences and produced structures that were moderately stable but mostly dynamic. (Kamtekar et al., 1993; Roy and Hecht, 2000) A second generation library 102-amino acid residue library was designed to be well-folded, by increasing the number of hydrophobic core residues, similar to that observed in naturally occurring four-helix bundles. (Wei et al., 2003b) Of the five proteins characterized from this second generation library, four were shown to form stable well-ordered structures, and two of these structures, S-824 and S-836, were shown by nuclear magnetic resonance (NMR) spectroscopy to form monomeric four helix bundles with the hydrophobic residues sequenced in the core and the hydrophilic residues sequenced exposed on the exterior, as designed (Figure 1-1) (Go et al., 2008; Wei et al., 2003a).

On the basis of this 102 amino acid template, a high-quality third generation combinatorial library comprising ~106 *de novo* four-helix bundles was constructed to probe functional activity in unevolved amino acid sequence space (Bradley et al., 2005; Patel et al., 2009) (Figure 1-1 A and Figure 1-2). Recently, it was demonstrated that proteins from this third generation library of binary patterned four-helix bundles are rich in rudimentary enzymatic activities *in vitro* (Patel et al., 2009) and have the ability to restore cell growth in auxotrophic *E. coli in vivo*, (Fisher et al., 2011) thereby establishing the functional potential of an unevolved artificial superfamily of proteins. During our initial characterization of highly expressed sequences arbitrarily chosen from this collection of novel functional proteins, WA20 was identified as one of the most stable structures with a cooperative guanidine hydrochloride (GdnHCl) denaturation midpoint value (3.8 M) significantly higher than those determined for second generation proteins S-824 (3.2M) and S-836 (3.0 M), respectively. (Platt, 2007; Wang, 2006; Wei et al., 2003b) (Bradley et al., unpublished). Moreover, WA20 has rudimentary activities as a peroxidase (with bound heme cofactor), esterase, and lipase. (Patel et al., 2009) To analyze the structural details of *de novo* protein WA20, its crystal structure was solved using multiwavelength anomalous dispersion (MAD). (Hendrickson, 1991) Here I describe the novel structure and discuss its feature.

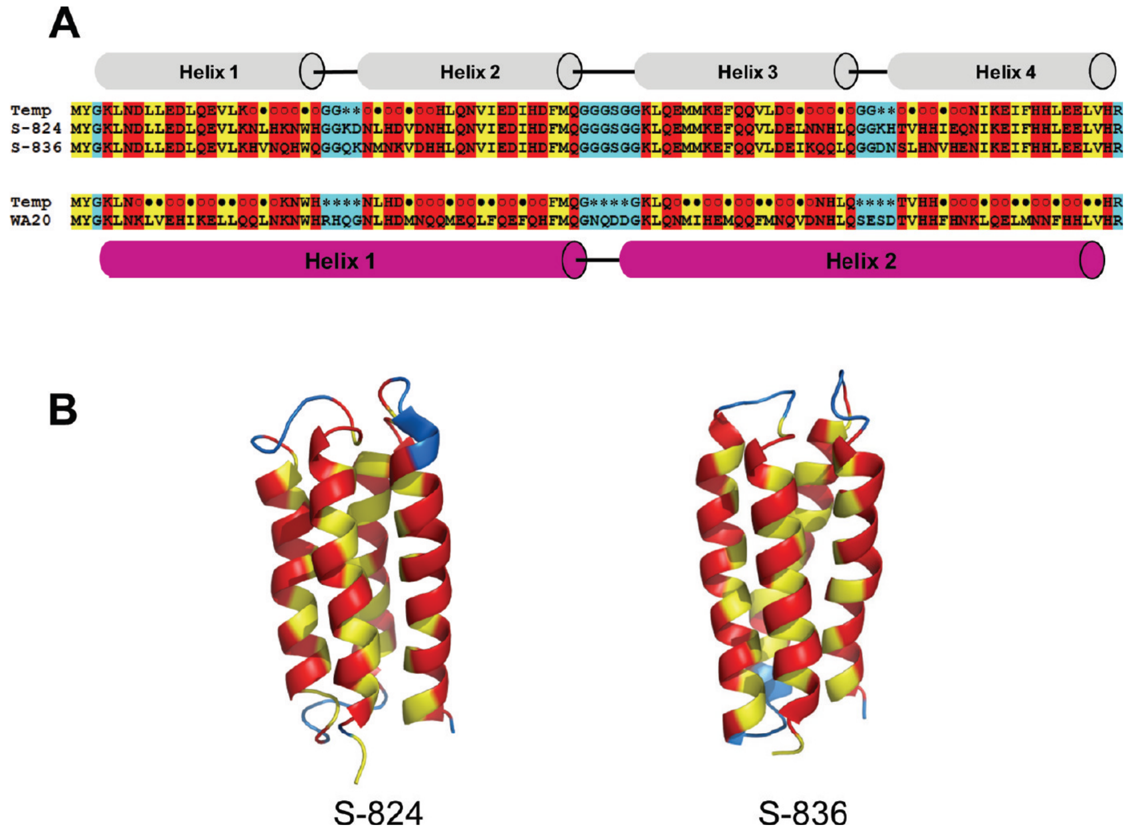
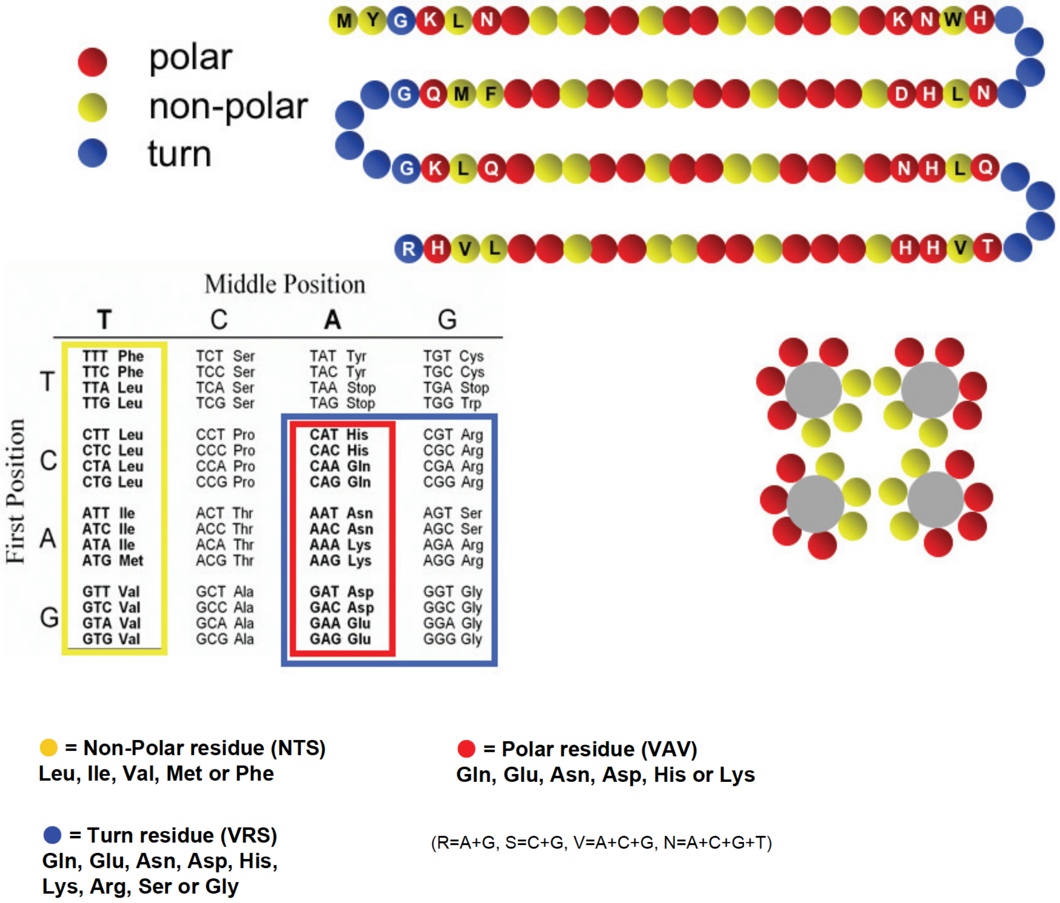

A
Helix 1
Helix 2
Helix 3
Helix 4
Temp MYGKLNDLLEDLQEVLK○●○○○●○GG**○●○○●○○HLQNVIEDIHDFMQGGGSGGKLQEMMKEFQQVLD○●○○○●○GG**○●○○●○○NIKEIFHHLEELVHR
S-824 MYGKLNDLLEDLQEVLKNLHKNWHGGKDNLHDVDNHLQNVIEDIHDFMQGGGSGGKLQEMMKEFQQVLDELNNHLQGGKHTVHHIEQNIKEIFHHLEELVHR
S-836 MYGKLNDLLEDLQEVLKHVNQHWQGGQKNMNKVDHHLQNVIEDIHDFMQGGGSGGKLQEMMKEFQQVLDEIKQQLQGGDNSLHNVHENIKEIFHHLEELVHR
Temp MYGKLN○●●○○●○○●●○○●○KNWH****NLHD●○○○●○○●●○○●○○FMQG****GKLQ○●●○○●○○●●○○●○NHLQ****TVHH●○○○●○○●●○○●○○●●HR
WA20 MYGKLNKLVEHIKELLQQLNKNWHRHQGNLHDMNQQMEQLFQEFQHFMQGNQDDGKLQNMIHEMQQFMNQVDNHLQSESDTVHHFHNKLQELMNNFHHLVHR
Helix 1
Helix 2
B
S-824
S-836

Figure 1-1. ***De novo*** **four-helix bundle proteins.** (A) Amino acid sequences of binary patterned four-helix bundle *de novo* proteins. Top: Design template for second generation library and the amino acid sequences of S-824 and S-836.7 The cylindrical shapes above the top represent designed regions of four helices. Bottom: Design template for third generation library and the amino acid sequences of WA20 (Patel et al., 2009). The magenta cylindrical shapes below the bottom represent two helices in the crystal structure of WA20 (chain A). The sequences follow the binary pattern design with red indicating polar residues and yellow indicating nonpolar residues. Turn residues are highlighted in blue. In the template (Temp) sequences, ○, ●, and ∗ represent polar, nonpolar, and turn residues, respectively (Figure 1-2). (B) Ribbon representation of S-824 (PDB: 1P68) (Wei et al., 2003a) and S-836 (PDB: 2JUA) (Go et al., 2008) structures determined by NMR. The color coding is the same as part A.



Middle Position

First Position	T	C	A	G
T	TTT Phe TTC Phe TTA Leu TTG Leu	TCT Ser TCC Ser TCA Ser TCG Ser	TAT Tyr TAC Tyr TAA Stop TAG Stop	TGT Cys TGC Cys TGA Stop TGG Trp
C	CTT Leu CTC Leu CTA Leu CTG Leu	CCT Pro CCC Pro CCA Pro CCG Pro	CAT His CAC His CAA Gln CAG Gln	CGT Arg CGC Arg CGA Arg CGG Arg
A	ATT Ile ATC Ile ATA Ile ATG Met	ACT Thr ACC Thr ACA Thr ACG Thr	AAT Asn AAC Asn AAA Lys AAG Lys	AGT Ser AGC Ser AGA Arg AGG Arg
G	GTT Val GTC Val GTA Val GTG Val	GCT Ala GCC Ala GCA Ala GCG Ala	GAT Asp GAC Asp GAA Glu GAG Glu	GGT Gly GGC Gly GGA Gly GGG Gly

● = Non-Polar residue (NTS)
Leu, Ile, Val, Met or Phe

● = Polar residue (VAV)
Gln, Glu, Asn, Asp, His or Lys

● = Turn residue (VRS)
Gln, Glu, Asn, Asp, His,
Lys, Arg, Ser or Gly

(R=A+G, S=C+G, V=A+C+G, N=A+C+G+T)

Figure 1-2. Design template of the third generation library for binary-patterned four helix bundle proteins. This figure is derived from Figure S1 of the reference (Patel et al., 2009).

1.2 Materials and methods

Protein Expression and Purification.

WA20 was indentified from a library of genes encoding the *de novo* proteins, as described previously (Figure 1-1A) (Bradley et al., 2005; Patel et al., 2009). The WA20 gene was coloned into an isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible protein expression vector, pET-3a (Novagen, MERCK, Darmstadt, Germany), with the T7 promoter and ampicillin resistance. WA20 protein was expressed in *E. coli* BL21 Star(DE3) (Invitrogen, Carlsbad, CA) using 2 L of LB medium at 30˚C. The selenomethionine (SeMet)-labeled WA20 protein was expressed in the methionine auxotroph *E. coli* B834(DE3) strain (Novagen, MERCK) using 2 L of LeMaster medium (LeMaster and Richards, 1985) with 100 mg of L-SeMet at 30˚C. Expression was induced with 0.2 mM IPTG (at OD600 = ~0.5), and cells were further cultured for 16 h at 30˚C. The protein was extracted from harvested cells using the freeze–thaw method (Johnson and Hecht, 1994) in lysis buffer (50 mM sodium phosphate (pH 7.0), 300 mM NaCl, 10% glycerol). The protein was purified by immobilized metal ion affinity chromatography with TALON metal affinity resin (Clontech, Takara Bio, Mountain View, CA) according to the manufacturer's protocols (equilibration/wash buffer: 50 mM sodium phosphate buffer (pH 7.0) containing 300 mM NaCl, 10% glycerol, and 250 mM imidazole). Even without a His-tag, the WA20 protein binds to a TALON metal affinity resin, presumably because of the relatively high percentage (12.7%) of histidine residues in its sequence (Table 1-1). WA20 was further purified by cation exchange chromatography (25 mM MES buffer (pH 6.0) containing 10% glycerol, with a linear gradient of NaCl from 0 to 1.5 M) with a Poros HS/M column (Perspective Biosystems) and gel filtration chromatography (25 mM HEPES buffer (pH 7.0) containing 100 mM NaCl, 10% glycerol, 1 mM dithiothreitol (DTT)) with a Superdex 75 10/300 GL column (GE healthcare, Little Chalfont, U.K.). the SeMet-labeled WA20 protein was used for the other experiments, the native WA20 protein was used for the other experiments.

Table 1-1. Amino acid composition of the *de novo* protein WA20.

Amino acids	Number	Percentage (%)
Nonpolar residues in the binary-patterned library		
Ile	2	1.96
Leu	12	11.76
Met	8	7.84
Phe	6	5.88
Val	4	3.92
Polar residues in the binary-patterned library		
Asn	12	11.76
Asp	5	4.90
Gln	16	15.69
Glu	7	6.86
His	13	12.75
Lys	6	5.88
Additional residues in the turn part		
Arg	2	1.96
Gly	4	3.92
Ser	2	1.96
Others		
Ala	0	0.00
Cys	0	0.00
Pro	0	0.00
Thr	1	0.98
Trp	1	0.98
Tyr	1	0.98
(Total)	102	100.0

Crystallization.

The crystals of the SeMet-labeled WA20 protein were obtained in a drop composed of 0.5 μL of the protein solution and 0.5 μL of the reservoir solution (0.056 M sodium phosphate monobasic monohydrate, 1.344 M potasodium phosphate dibasic, pH 8.2) by the sitting drop vapor diffusion method against 50 μL of the reservoir solution at 4˚C. Rod-like crystals were obtained several weeks (Figure 1-3).

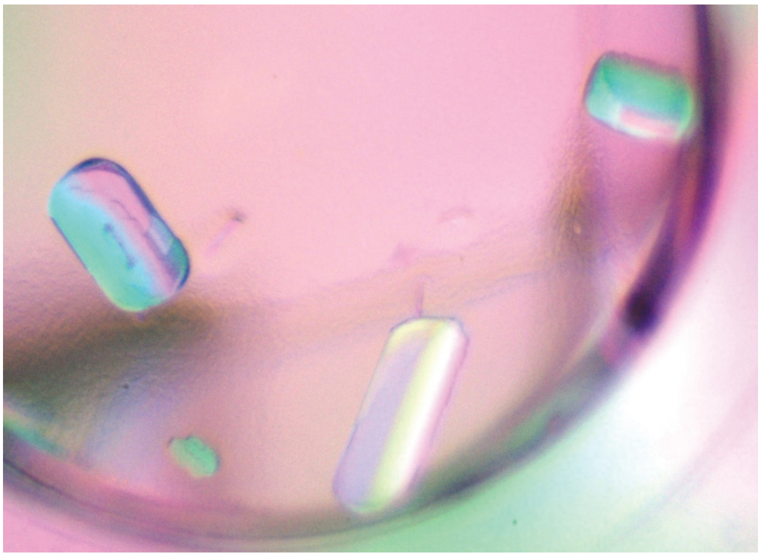


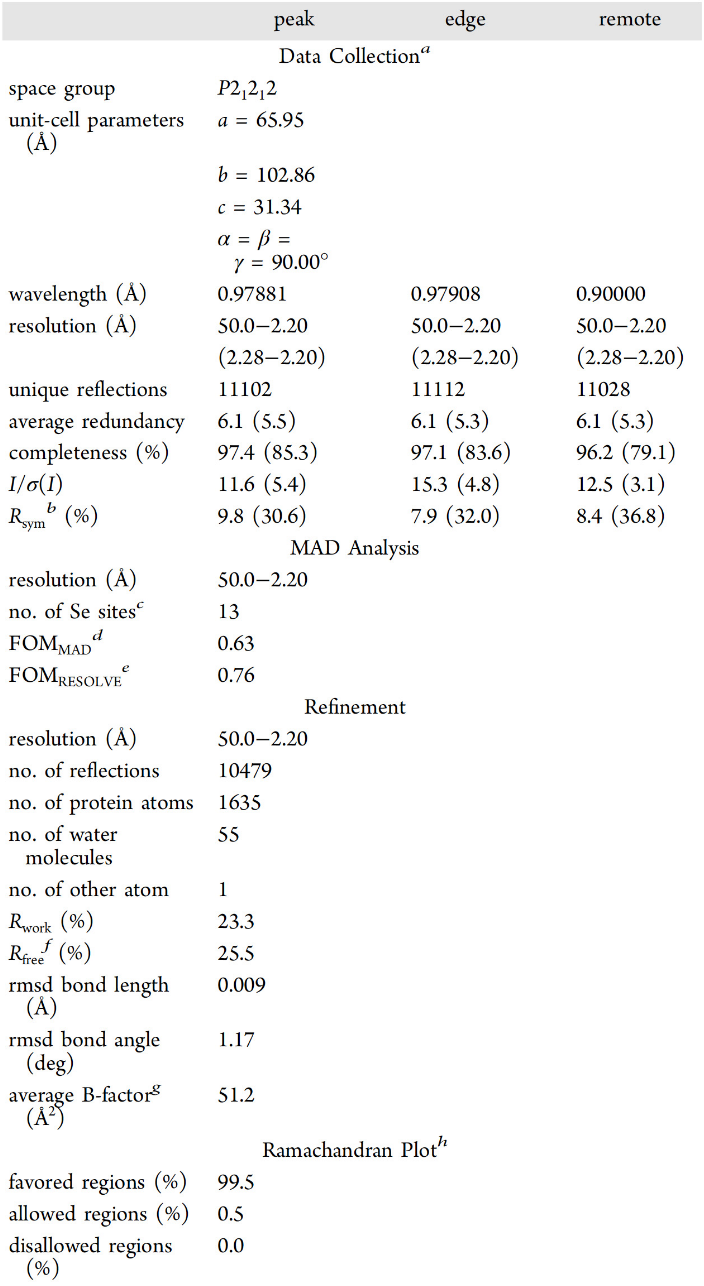
Figure 1-3. WA20 (SeMet) crystals. The crystals were obtained by the sitting drop vapor diffusion method against the reservoir solution (0.056 M sodium phosphate monobasic monohydrate, 1.344 M potassium phosphate dibasic, pH 8.2) at 4 ºC.

Data Collection, Structure Determination, and Refinement.

The X-ray diffraction data were collected at the Photon Factory, BL-5A (KEK, Tsukuba, Japan). The data collection was carried out at 95 K with a mixture of equal parts of Paratone-N (Hampton Research, Aliso Viejo, CA) and paraffin oil as a cryoprotectant. All diffraction data were processed with the program HKL2000 (Otwinowski and Minor, 1997) (Table 1-2).

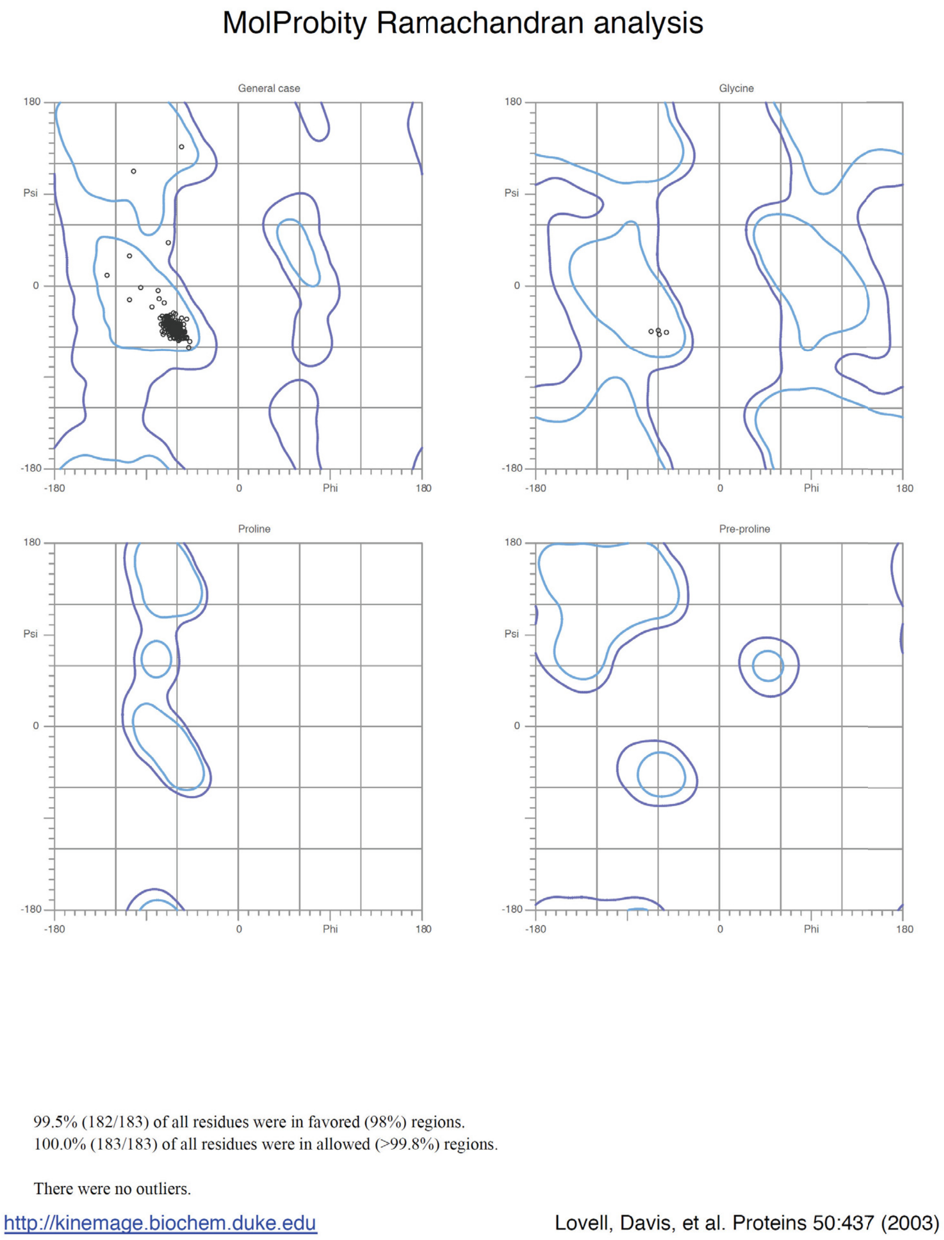
The program SOLVE (Terwilliger and Berendzen, 1999) was used to locate the selenium sited and to calculate the phased by the MAD method, and the program RESOLVE (Terwilliger, 2002) was used for the density modification and partial model building. The model was build and corrected with the program COOT (Emsley et al., 2010) and was refined with the program REFMAC5 (Murshudov et al., 2011; Murshudov et al., 1997) in the CCP4 suite (Collaborative Computational Project, 1994). All refinement statics are presented in Table 1. The quality of the model was inspected by the programs PROCHECK (Laskowski et al., 1993) and MolProbity (Chen et al., 2010; Davis et al., 2007; Lovell et al., 2003) (Figure 1-4). The atomic coordinates and the structure factors have been deposited in the Protein Data Bank, with the accession code 3VJF. The graphic figures were created using the program PyMOL (Delano Scientific LLC).

Table 1-2. X-ray Data Collection and Refinement Statistics



	peak	edge	remote
	Data Collection[a]		
space group	$P2_12_12$		
unit-cell parameters (Å)	a = 65.95		
	b = 102.86		
	c = 31.34		
	$\alpha = \beta = \gamma$ = 90.00°		
wavelength (Å)	0.97881	0.97908	0.90000
resolution (Å)	50.0−2.20 (2.28−2.20)	50.0−2.20 (2.28−2.20)	50.0−2.20 (2.28−2.20)
unique reflections	11102	11112	11028
average redundancy	6.1 (5.5)	6.1 (5.3)	6.1 (5.3)
completeness (%)	97.4 (85.3)	97.1 (83.6)	96.2 (79.1)
$I/\sigma(I)$	11.6 (5.4)	15.3 (4.8)	12.5 (3.1)
R_{sym}[b] (%)	9.8 (30.6)	7.9 (32.0)	8.4 (36.8)
	MAD Analysis		
resolution (Å)	50.0−2.20		
no. of Se sites[c]	13		
FOM_{MAD}[d]	0.63		
$FOM_{RESOLVE}$[e]	0.76		
	Refinement		
resolution (Å)	50.0−2.20		
no. of reflections	10479		
no. of protein atoms	1635		
no. of water molecules	55		
no. of other atom	1		
R_{work} (%)	23.3		
R_{free}[f] (%)	25.5		
rmsd bond length (Å)	0.009		
rmsd bond angle (deg)	1.17		
average B-factor[g] (Å^2)	51.2		
	Ramachandran Plot[h]		
favored regions (%)	99.5		
allowed regions (%)	0.5		
disallowed regions (%)	0.0		

[a] All numbers in parentheses represent last outer shell statistics. [b] $R_{sym} = \Sigma|I_i - I_{avg}|/\Sigma I_i$, where I_i is the observed intensity and I_{avg} is the average intensity. [c] Number of selenium sites located using SOLVE. [d] Figure of merit after SOLVE phasing. [e] Figure of merit after RESOLVE density modification. [f] Rfree is calculated for 5% of randomly selected reflections excluded from refinement. [g] Average B-factor is average of sum of TLS (Translation, Libration, and Screw-rotation) and residual B-factors. [h] The Ramachandran plot is shown in Figure 1-4



MolProbity Ramachandran analysis

99.5% (182/183) of all residues were in favored (98%) regions.
100.0% (183/183) of all residues were in allowed (>99.8%) regions.

There were no outliers.

http://kinemage.biochem.duke.edu

Lovell, Davis, et al. Proteins 50:437 (2003)

Figure 1-4. Ramachandran plot of the WA20 structure analyzed by MolProbity (Chen et al., 2010; Davis et al., 2007; Lovell et al., 2003).

Small-Angle X-ray Scattering (SAXS).

SAXS measurements on WA20 (5.4 mg/mL) were performed to examine the static structure in solution (25 mM HEPES buffer (pH 7.0) containing 100 mM NaCl, 10% glycerol, and 1 mM DTT). I used a SAXSess camera (Anton Paar, Graz, Austria) attached to a sealed tube anode X-ray generator was operated at 40 kV and 50 mA. A Göbel mirrer and a block collimator provide a focused monochromatic X-ray beam of Cu Kα radiation (λ = 0.1542 nm) with a well-defined shape. A thermostatted sample holder unit (TCS 120, Anton Paar) was used to control the sample temperature. The two-dimensionl scattering patterns recorded by an imaging-plate (IP) detector (Cyclone, Perkin-Elmer) were integrated into one-dimensional scattered intensities, $I(q)$, as a function of the magnitude of the scattering vector $q = (4\pi/\lambda)\sin(\theta/2)$ using SAXSQuant software (Anton Paar), where θ is the total scattering angle. For all experiments, the attenuated primary beam at $q = 0$ was monitored using a semitransparent beam stop. All the measured intensities were semiautomatically calibrated for transmission by normalizing a zero-q primary intensity to unity. The background scattering contributions from capillary and solvent were corrected. The absolute intensity calibration was made by using water intensity as a secondary standard (Orthaber et al., 2000).

Assuming the structure factor $S(q) = 1$ for dilute samples, $I(q)$ is given by Fourier transformation of the so-called pairdistance distribution function $p(r)$, i.e., the spatial autocorrelation function of the electron density fluctuations, $\Delta\rho(r)$, as

$$I(q) = 4\pi \int_0^\infty p(r) \frac{\sin qr}{qr}\, \mathrm{d}r$$

where r is the distance between two scattering centers chosen inside the molecule. The indirect Fourier transformation (IFT) technique (Glatter, 1980a; Glatter, 1980b; Glatter and Kratky, 1982) was used to calculate $p(r)$.

Refolding and Gel Filtration Chromatography.

The WA20 protein was denatured by 6 M guanidine hydrochloride (GdnHCl) for 2 h at 4 °C in 25 mM HEPES buffer (pH 7.0) containing 100 mM NaCl, 10% glycerol, and 1 mM DTT. For refolding, the denatured WA20 proteins at different concentrations (2.1, 0.21, and 21 μg/mL) were dialyzed three times for ~4 h (×3) against 200× volume of 25 mM HEPES buffer (pH 7.0) containing 100 mM NaCl, 10% glycerol, and 1 mM

DTT. The refolded WA20 proteins were analyzed by gel filtration chromatography (25 mM HEPES buffer (pH 7.0) containing 100 mM NaCl, 10% glycerol, and 1 mM DTT) on a Superdex 75 10/300 GL column (GE healthcare). Absorbance was monitored at 280 nm. The calibration curve for molecular weight (M_w) estimation was plotted with a Gel Filtration Calibration kit LMW (GE Healthcare) (Figure 1-5).

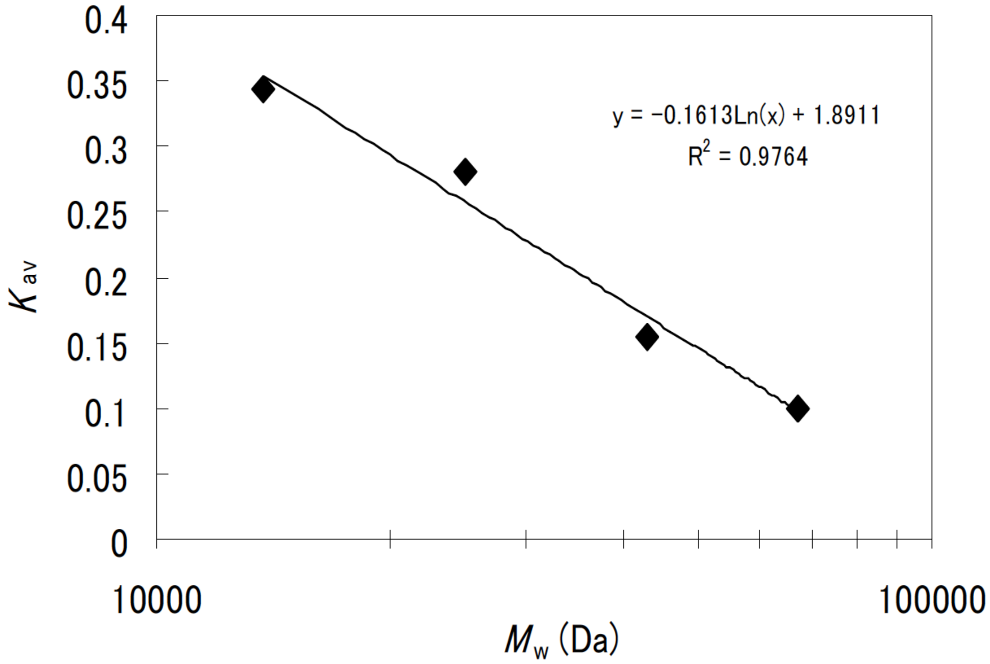


Figure 1-5. Calibration curve by gel filtration chromatography on Superdex 75 10/300 GL. The x-axis is molecular weight (M_w) in log scale. The calibration protein samples are ribonuclease A (13700 Da), chymotrypsinogen A (25000 Da), ovalbumin (43000 Da), and bovine serum albumin (67000 Da). K_av values for each protein were calculated using the equation, $K_\mathrm{av} = (V_\mathrm{e} - V_\mathrm{o})/(V_\mathrm{t} - V_\mathrm{o})$, where V_e is elution volume for the protein, and V_o is column void volume (elution volume of Blue Dextran 2000), and V_t is total bed volume (24 ml).

Differential Scanning Fluorimetry (DSF).

A real-time PCR device, MiniOpticon (Bio-Rad, Hercules, CA), was used to monitor protein unfolding by the increase in the fluorescence of the fluorophore SYPRO Orange (Invitrogen) with affinity for hydrophobic parts of the protein, which are exposed as the protein unfolds (Niesen et al., 2007; Vedadi et al., 2006). WA20 protein (1 mg/mL) samples (20 μL) with SYPRO Orange (5× concentration) in 25 mM HEPES buffer (pH 7.0) containing 100−1000 mM NaCl, 10% glycerol, and 1 mM DTT were analyzed in 48-well PCR microplates (Bio-Rad). The relative fluorescence intensity was plotted as a function of temperature; this generates a sigmoidal curve that can be described by a two-state transition (Niesen et al., 2007). The inflection point of the transition curve (T_m) was calculated using the curve fitting function in KaleidaGraph (Synergy Software, Reading, PA) with the following equation:

$$y = LL + \frac{UL - LL}{1 + \exp\left(\frac{T_m - x}{a}\right)}$$

where LL and UL are the values of minimum and maximum intensities, respectively, and a denotes the slope of the curve within T_m (Niesen et al., 2007).

1.3 Results and Discussion

Overall Structure.

The crystal structure of the *de novo* four helix bundle protein WA20 was solved by the MAD method, and refined to 2.2 Å. The crystallographic data are summarized in Table 1-2. The WA20 crystal contains two protein molecules per asymmetric unit. The final model includes 189 amino acid residues of two WA20 monomers, 55 water molecules, and one metal ion in the asymmetric unit. The metal ion is probably potassium, because of the metal−ligand geometry (Harding, 2002). The N-terminal and C-terminal residues and some loop residues (chain A, 1−2, 102; chain B, 1−4, 49−55, 102) are invisible due to disorder. Surprisingly, the WA20 crystal structure is not a monomeric four-helix bundle like the *de novo* proteins S-824 (Wei et al., 2003a) and S-836 (Go et al., 2008) (Figure 1-1B) but a dimeric four-helix bundle (Figure 1-6). Each monomer comprises two long α-helices, which span residues 4−50 (α1), 54−100 (α2) in chain A (Figure 1-1A) and residues 8−47 (α1), 58−100 (α2) in chain B. The helices intertwine with the helices of the other monomer, and the two monomers together form a 3D domain-swapped (Bennett et al., 1995; Liu and Eisenberg, 2002) dimer. The four α-helices wrap around into a left-handed coiled coil (Mason and Arndt, 2004). The overall shape of WA20 is cylindrical with a length of ~8 nm and a diameter of ~3 nm. Helices α1 (chain A) and α2 (chain B) are roughly parallel. Helices α1 (B) and α2 (A) are also roughly parallel. In contrast, the angles between helices α1 (A) and α1 (B) and helices α2 (A) and α2 (B) are about 20°, similar to the angle found in the "knobs-into-holes" packing of many natural α-helical proteins. (Crick, 1953) Four-helix bundles, in which some angles are ~20° and others are more parallel (or antiparallel), also occur in nature (e.g., cytochrome b_{562}) (Lederer et al., 1981) and in *de novo* proteins (e.g., S-824 and S-836). (Go et al., 2008; Wei et al., 2003a) In more detailed views, there are differences of architecture between the four-helix bundle *de novo* proteins, monomeric S-824 and dimeric WA20. In the monomeric S-824, helices 1 (residues 5−20) and 2 (residues 28−48), and helices 3 (residues 56−72) and 4 (residues 80−99) are roughly antiparallel, and the angle between helices 1 and 4 and between helices 2 and 3 is ~20°. In contrast, in the view of the upper half part of WA20, their corresponding regions 1 (residues 5−20 in chain B) and 3 (residues 56−72 in chain A), and regions 2 (residues 28−48 in chain A) and 4 (residues 80−99 in chain B) are roughly parallel, and the angle between regions 1 and 2 and between regions 3 and 4 is ~20°.

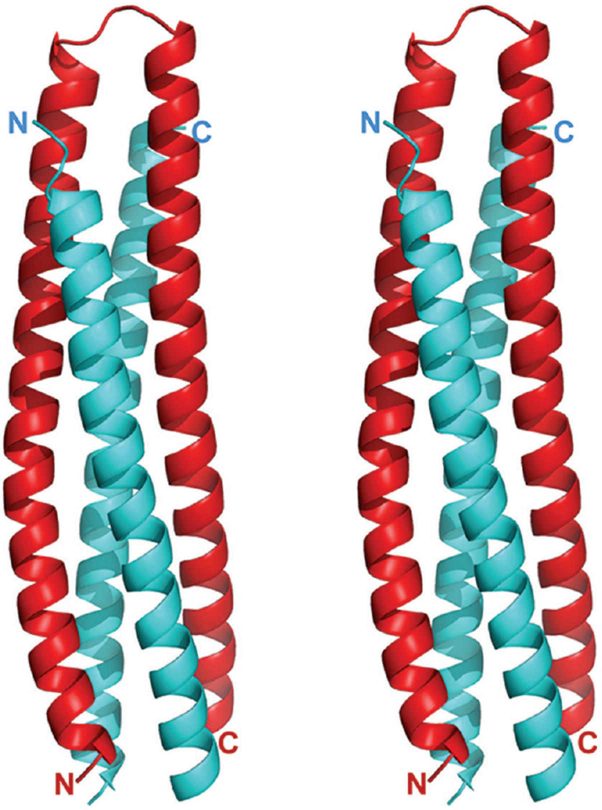


Figure 1-6. Ribbon representation of the crystal structure of WA20 (stereoview). Chains A and B are shown in red and cyan, respectively.

The Binary Patterned Structure and the Dimer Interface.

Figure 1-7A shows that the hydrophobic residues of WA20 form core regions in the four-helix bundle dimeric structure. The head-on views in Figure 1-7B and C indicate that the side chains in the crystal structure are clearly partitioned with nonpolar residues (yellow) in the interior and polar residues (red) on the surface, as specified by the binary code design strategy. In addition, Figure 1-7B shows that four helices are located on the diamond shape at the end part of the WA20 structure. In the diagonal distance of the diamond shape, the helices at the loop region are proximal (ca. 1.2−1.3 nm) and the helices at the terminal regions are distant (ca. 1.8−1.9 nm).

The dimer interface is predominantly hydrophobic with several hydrophobic clusters (Figure 1-8A and Figure 1-9). The major residues, involved in these hydrophobic interactions, include Val9, Ile12, Leu16, Leu19, Trp23, Leu30, Met33, Met37, Leu40, Phe41, Phe44, Met48, Met64, Phe67, Val71, Leu75, Phe85, Leu89, Leu92, Phe96, and Leu99 (Figure 1-8A). I suggest that one reason for the stable structure of WA20 is because there are roughly twice as many hydrophobic interactions in a dimeric structure of WA20 relative to a monomeric structure like S-824. The dimer interface is further stabilized by interchain salt bridges and/or hydrogen bonds between the atoms of Gln27(A)-Thr81(B), Glu63(A)-Lys13(B), Asp72(A,B)-His86-(B,A), His74(A)-Asn20(B), and Ser79(A)-His83(B) (Figure 1-8B and Figure 1-9). To estimate the key residues of dimerization, I carefully compared the structure and amino acid sequence of dimeric WA20 with those of monomeric S-824. Significant differences occur in the loop regions of residues 25−28 and 77−80. The regions (GGKD and GGKH) in S-824 are glycine-rich loops. In contrast, the regions (RHQG and SESD) in WA20 are located in the middle of α-helices, which are stabilized by intrachain salt bridges or hydrogen bonds between His26 and Glu78 (Figure 1-8C). These observations suggest that the His26 and Glu78 residues in the designed loop regions are potential key residues leading to the formation of the domain-swapped dimer, since the interactions at loops affect the free energy difference between the monomer and the 3D domain-swapped multimer (Liu and Eisenberg, 2002).

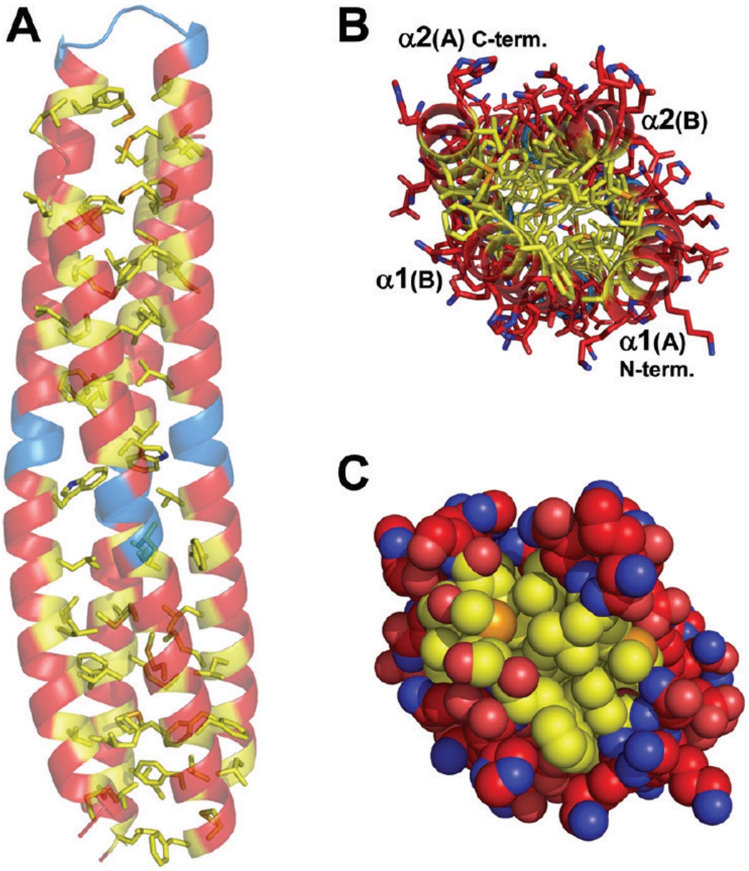


Figure 1-7. The binary patterned structure of WA20. (A) The hydrophobic core of WA20 in the four-helix bundle dimeric structure. Nonpolar side chains are shown as stick models. The color coding is the same as Figure 1-1A. Polar (red), nonpolar (yellow), and turn (cyan) residues as the design template. (B) Head-on view (from the side with disordered loop region in chain B) with the polar (red) and nonpolar (yellow) side chains shown as stick models. (C) Same as part B but in space-filling representation.

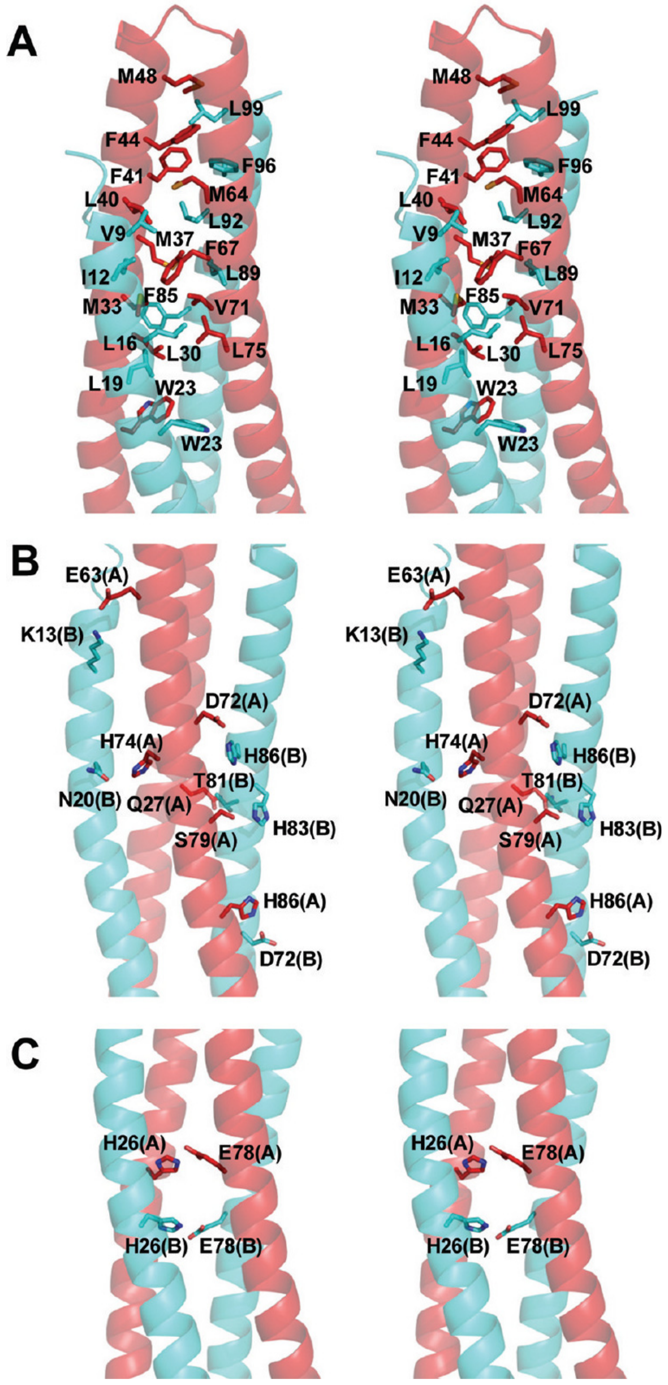


Figure 1-8. Close-up view of the dimer interface of WA20 (stereoview). Chains A and B are shown in red and cyan, respectively. (A) The major clusters of hydrophobic residues in the dimer interface are shown as stick models. (B) The residues of interchain salt bridges and hydrogen bonds, as determined by DIMPLOT in LIGPLOT, (Wallace et al., 1995) are shown as stick models. (C) Potential key residues of intrachain salt bridges or hydrogen bonds for dimer formation are shown as stick models.

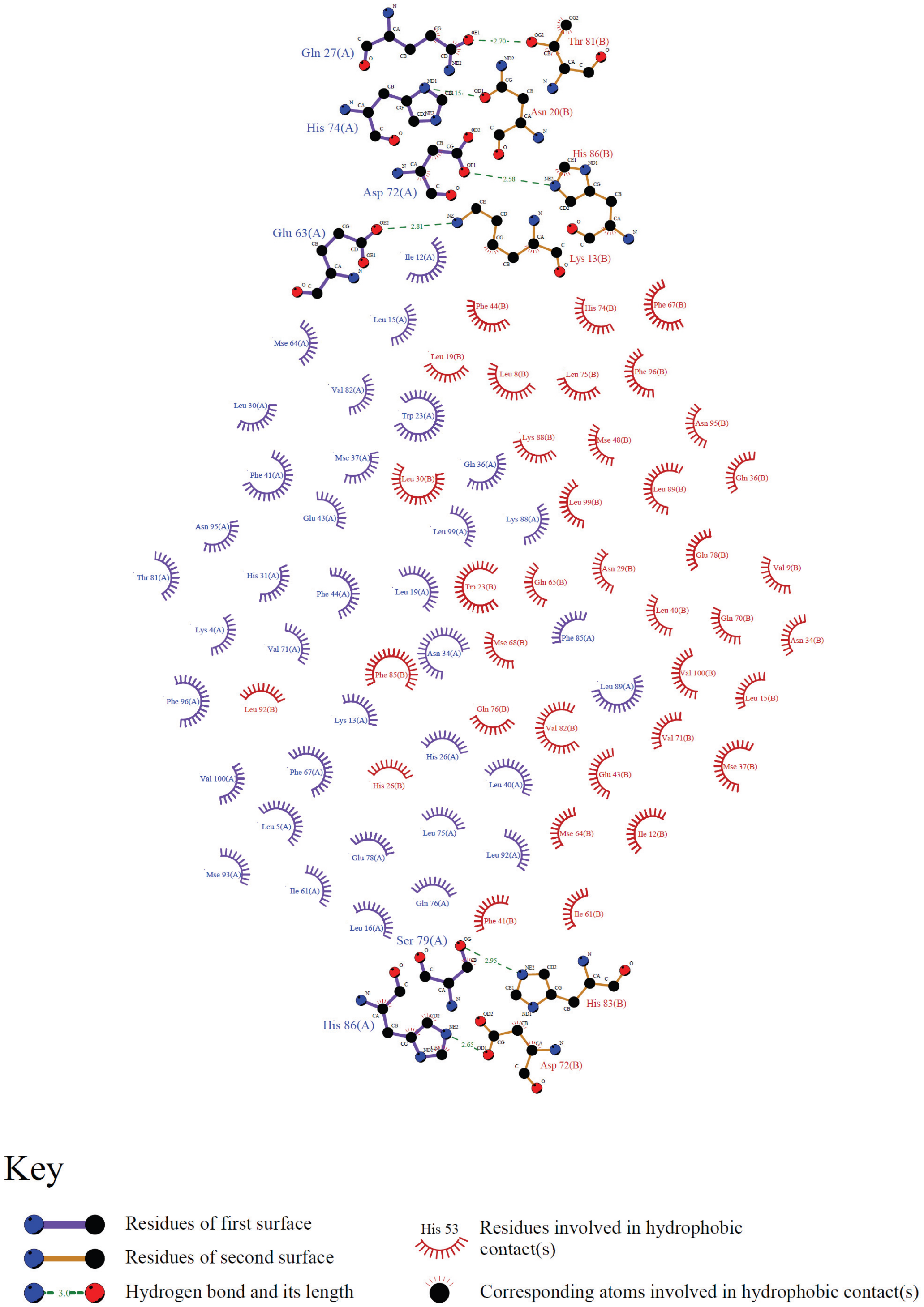


Figure 1-9. Interactions at dimer interface of WA20, analyzed and depicted by the program DIMPLOT in LIGPLOT (Wallace et al., 1995).

Small-Angle X-ray Scattering (SAXS).

To examine the solution structure of WA20, SAXS experiments were performed. Figure 1-10A and B show that X-ray scattering intensities of WA20 and chicken egg lysozyme as a molecular weight reference (lysozyme M_w = 14.3 kDa). Assuming that WA20 and lysozyme have identical scattering length densities, no aggregation in solution, and the structure factor $S(q) = 1$ for dilute samples, the forward scattering intensity $I(q\rightarrow0)$ is proportional to the molecular weight (M_w) at the same concentration (5.4 mg/mL). The $I(q\rightarrow0)$ of WA20 and lysozyme are 0.0812 and 0.0459 cm^{-1}, respectively. The molecular weight of WA20 was estimated to be 25.3 kDa. Since the molecular weight of WA20 monomer is 12.5 kDa, these data show that WA20 forms a dimer in solution. To extract intuitive real-space information via a virtually model-free routine, the pair-distance distribution function, $p(r)$, of WA20 was obtained using an indirect Fourier transformation (IFT) technique (Figure 1-10D). The $p(r)$ indicates that the maximum diameter, D_max, is ~8 nm, which is consistent with the crystal structure. The observed pronounced peak of $p(r)$ in the low-r regime and extended linear tail in the high-r regime are significant features of rod-like structure (Sato et al., 2010). The inflection point located on the higher-r side of the maximum in $p(r)$, highlighted by broken lines in Figure 1-10D, gives a measure of the cross section diameter, D_c max. The D_c max value of WA20 is roughly ~3 nm, which is also consistent with the crystal structure. Furthermore, $I(q)$ and $p(r)$ of WA20 resemble those simulated from the crystal structure of WA20 (Figure 1-10C and D). These SAXS results show that WA20 forms the dimeric four-helix bundle structure in solution.

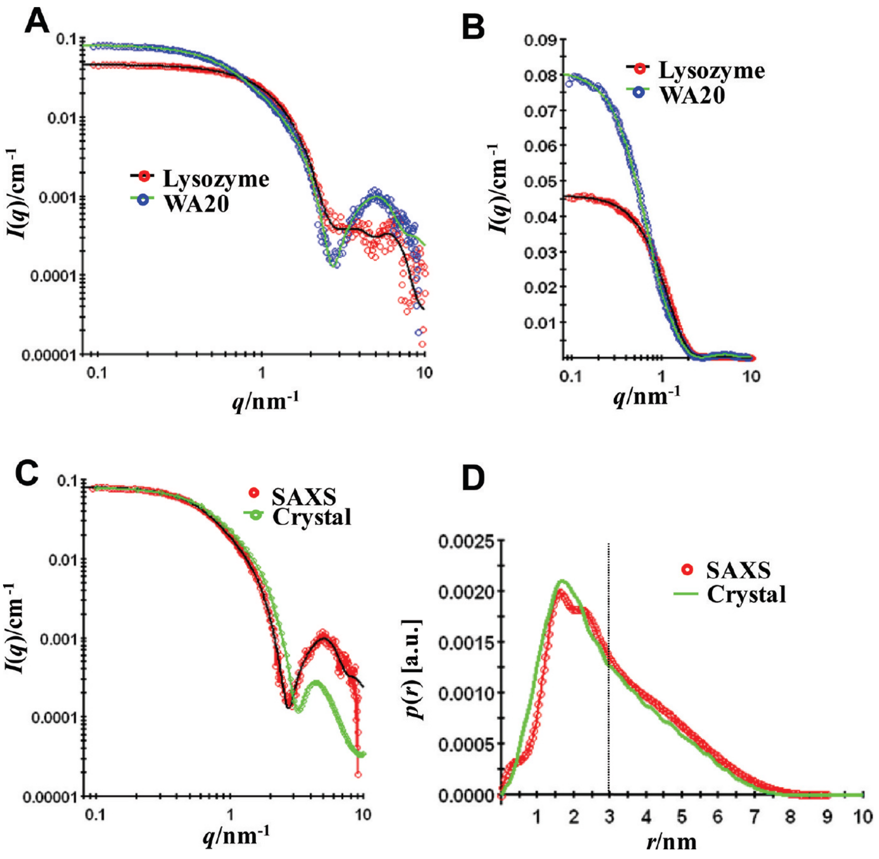


Figure 1-10. Small-angle X-ray scattering (SAXS) analyses. SAXS intensities, *I*(*q*), in (A) log scale and (B) linear scale, of WA20 (5.4 mg/mL) and chicken egg lysozyme (5.4 mg/mL) in HEPES buffer solution at 25 °C on absolute scale. (C) SAXS intensities, *I*(*q*), of WA20 solution and that simulated from the crystal structure of WA20. (D) The corresponding pair distance distribution function, *p*(*r*), of WA20 obtained as output of indirect Fourier transformation (IFT) analysis and that simulated from the crystal structure of WA20.

Concentration Independence of Dimerization in Refolding.

To examine the concentration dependence of dimerization, I tried refolding of WA20 at different concentrations, and analyzed the resulting protein by gel filtration chromatography. WA20 was denatured by 6 M GdnHCl and refolded by dialysis at different concentrations of protein (2.1 mg/mL, 0.21 mg/mL, and 21 μg/mL). There was no precipitation during refolding. The circular dichroism (CD) spectra (Figure 1-11) show that WA20 was denatured completely by 6 M GdnHCl, and the α-helical content of WA20 was recovered by refolding. Table 1-3 shows the elution volume and estimated molecular weight by gel filtration chromatography with the calibration curve (Figure 1-5). The dimer peak of WA20 was clearly detected by gel filtration chromatography in each tested concentration following refolding, and no monomer peak was detected (Figure 1-12). These data indicate that, in the range of concentrations tested, the refolding of WA20 into its dimeric structure is independent of concentration, thereby suggesting that the dimer form of WA20 is much more stable than the monomer.

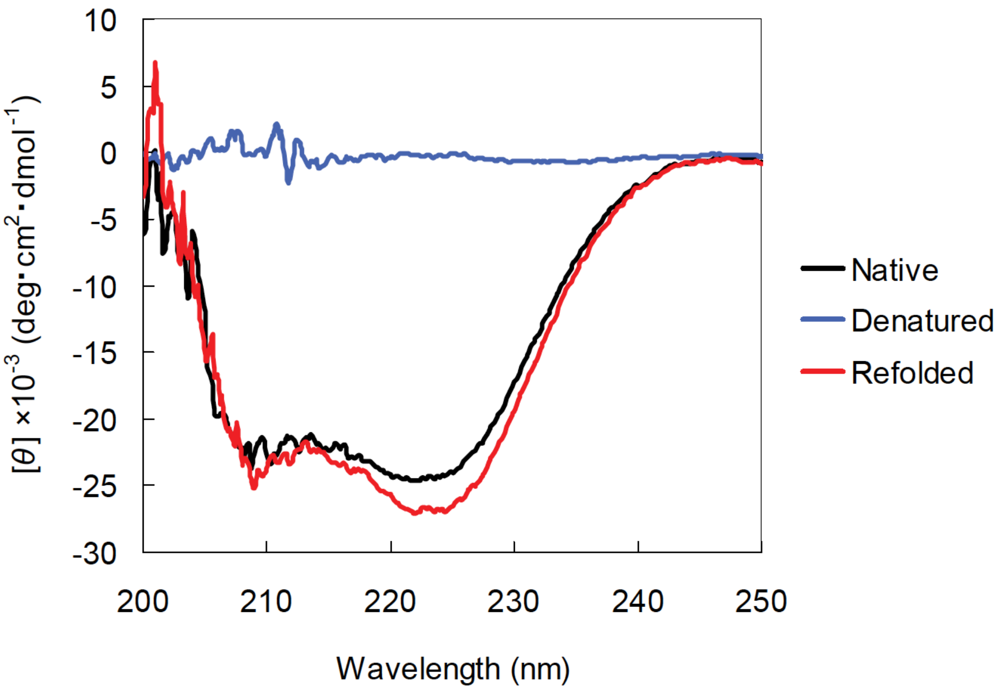


Figure 1-11. Circular dichroism (CD) spectra of WA20. CD spectra of native WA20 (0.22 mg/mL) (black line), WA20 (0.35 mg/mL) denatured by 6 M GdnHCl (blue line), and WA20 (0.18 mg/mL) refolded at 2.1 mg/mL (red line), were analyzed by J-600 CD spectrometer (JASCO, Tokyo, Japan).

Table 1-3. Elution Volume and Molecular Weight Estimated by Gel Filtration Chromatography

WA20 sample	V_e (mL)	K_{av}	M_w (kDa)	state
native	11.91	0.254	25.6	dimer
refolded at 2.1 mg/mL	11.83	0.249	26.4	dimer
refolded at 0.21 mg/mL	11.85	0.250	26.2	dimer
refolded at 21 μg/mL	11.87	0.251	26.0	dimer

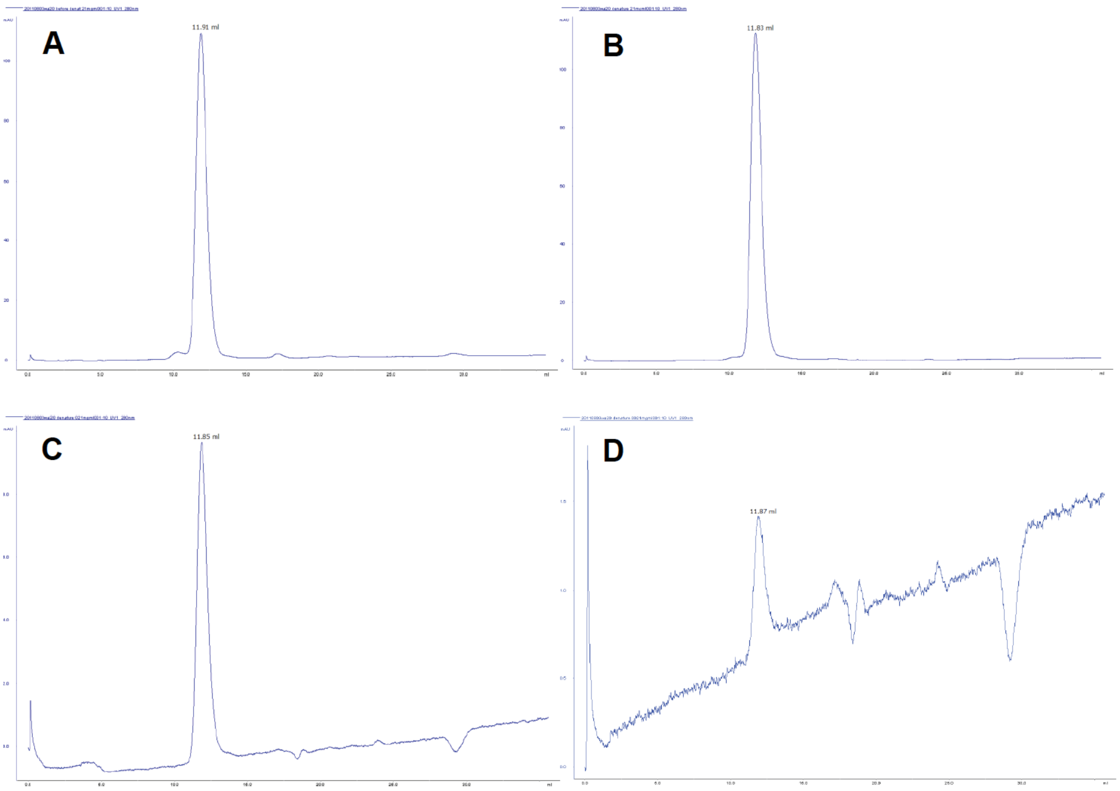


Figure 1-12. Gel filtration chromatograms of WA20 on Superdex 75 10/300 GL. (A) Native WA20 (2.1 mg/mL). (B) The sample refolded at 2.1 mg/mL WA20. (C) The sample refolded at 0.21 mg/mL WA20. (D) The sample refolded at 21 μg/mL WA20. This concentration was nearly detection limit of our chromatography system (AKTA explorer 10S, GE healthcare).

Thermal Denaturation.

To examine protein stability, I tested thermal denaturation by differential scanning fluorimetry (DSF) (Niesen et al., 2007; Vedadi et al., 2006) (Figure 1-13). The temperature at which a protein unfolds is measured by an increase in the fluorescence of the SYPRO orange dye with affinity for hydrophobic parts of the protein, which are exposed as the protein unfolds (Niesen et al., 2007; Vedadi et al., 2006). The melting temperature (T_m) of WA20 is about 70°C at various salt concentrations (0.1−1 M NaCl) (Table 1-4), indicating that the high stability of the dimeric structure is independent of salt. In addition, to test whether the temperature-induced unfolding of WA20 is reversible, I monitored the fluorescence by DSF again after the first thermal denaturation and cooling. The fluorescence curve of the second thermal denaturation is significantly different from that of the first denaturation (Figure 1-14), implying that the temperature-induced unfolding of WA20 is not reversible in the test condition.

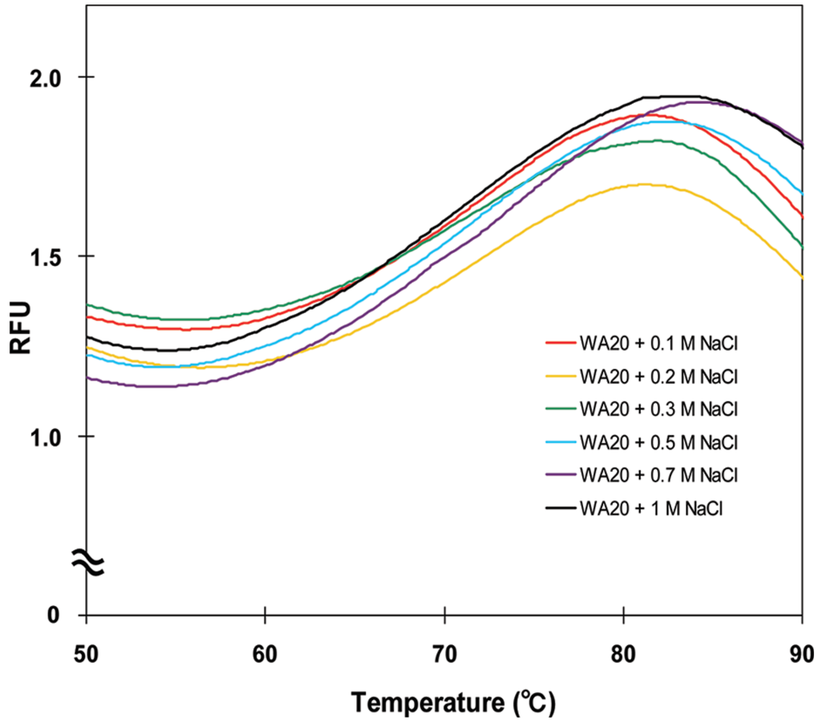


Figure 1-13. Thermal denaturation curves of WA20 by differentialscanning fluorimetry (DSF) at various salt concentrations (0.1−1 M NaCl) (RFU, Relative Fluorescence Units).

Table 3. Melting Temperature (T_m) of WA20 at Various Salt Concentrations, Analyzed by Differential Scanning Fluorimetry (DSF)

NaCl conc. (M)	T_m (°C)
0.1	70.1
0.2	70.3
0.3	69.9
0.5	69.7
0.7	70.8
1	69.7

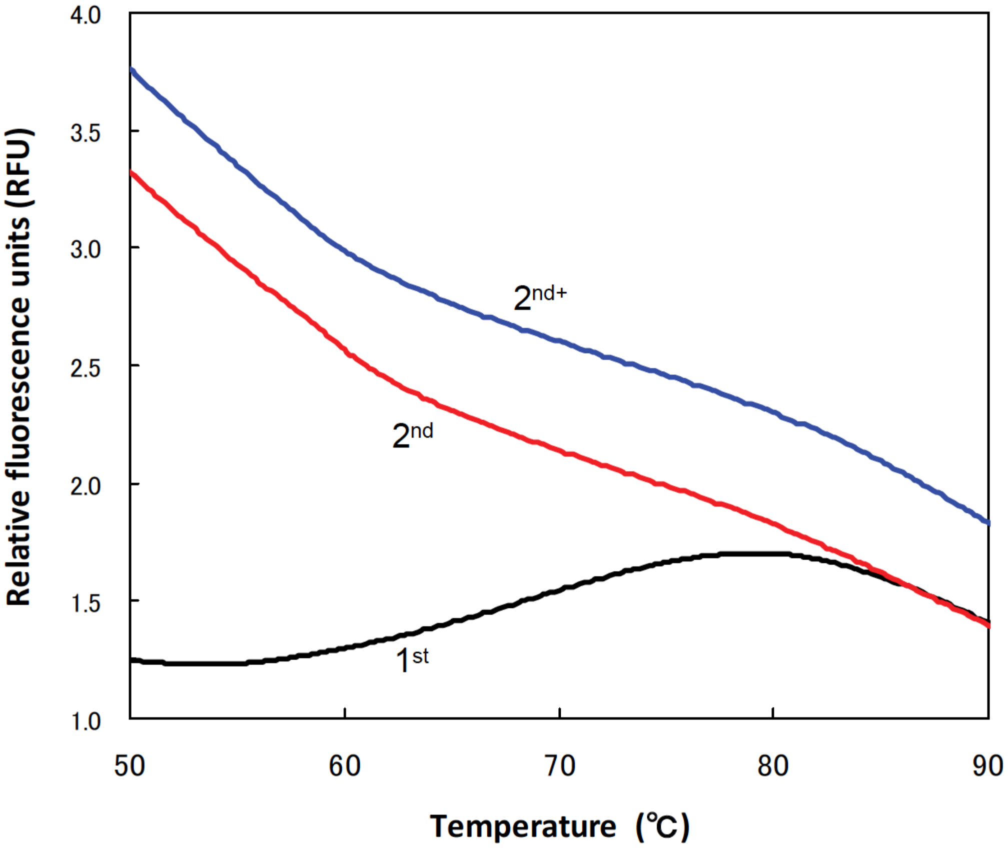


Figure 1-14. Thermal denaturation curves of WA20 by differential scanning fluorimetry (DSF). WA20 (1 mg/mL) protein samples (20 μL) with SYPRO Orange (5× concentration) in 25 mM HEPES buffer (pH 7.0) containing 100 mM NaCl, 10% glycerol and 1 mM DTT were analyzed in 48-well PCR microplates by MiniOpticon (Bio-Rad). After the first thermal denaturation (1st, black line) and cooling, I monitored

again the fluorescence in the second thermal denaturation (2^{nd}, red line). In the case of another second thermal denaturation curve (2^{nd+}, blue line), I added the SYPRO orange dye to the sample after the first thermal denaturation and cooling without the SYPRO orange dye. The thermal denaturation curves of the second thermal denaturation are significantly different from that of the first thermal denaturation, implying that the temperature-induced unfolding of WA20 is not reversible in the test condition.

Putative Primitive Active Site.

It was previously shown that WA20 binds heme, and that this protein/heme complex has rudimentary peroxidase activity (Patel et al., 2009). As shown previously (Patel et al., 2009) (and illustrated in Figure 1-15), the heme complex of WA20 produces an absorbance spectrum with a typical Soret peak at 410 nm. In natural heme proteins, histidine and methionine are generally used as axial ligand residues for heme (Reedy and Gibney, 2004). WA20 is relatively rich in these residues, with 26 histidine and 16 methionine residues in the dimer (twice the 102-residue primary sequence, Table 1-1). From this structure, several pairs of putative candidates for heme-ligand residues between chains A and B (e.g., His11-Met33, His24-His74, His31-His84, Met48-His101, His62-His97, His62-His101) are estimated by reference to the distance and geometry between axial ligand residues in cytochrome b_{562} (Lederer et al., 1981) and cytochrome *c′* (Finzel et al., 1985). Further studies are necessary to confirm the binding site.

It was also shown previously that WA20 in the absence of heme has low levels of esterase and lipase activity (Patel et al., 2009). Although substantially less active (~10000-fold) than natural enzymes, this novel protein produces rate enhancements ($k_{\mathrm{cat}}/k_{\mathrm{uncat}}$) that are ~400-fold and ~500-fold above background for esterase and lipase activities, respectively (Patel et al., 2009). To find putative substrate binding sites for hydrolase activities (esterase and lipase), I searched for pocket sites in the WA20 dimeric structure. Two relatively large pockets (volumes: 205 and 174 Å^3), comprised of Leu30, Met33, Asn34, Met37, His74, and Leu75 (in chain A/B) and Leu16, Leu19, Phe85, Leu89, and Leu92 (in chain B/A) are detected using the programs Pocket-Finder (Hendlich et al., 1997) (http://www.modelling.leeds.ac.uk/pocketfinder/) and Caver (Beneš et al., 2010) (http://www.caver.cz/) (Figure 1-16). Similar sized pockets were also found in the F64A mutant of monomeric protein S-824 (Das et al., 2011). I hypothesize that these hydrophobic pockets may serve as substrate binding sites. In natural hydrolases, a carboxyl peptidase family, Eqolisin, hydrolyzes peptide bonds using Glu and Gln as catalytic residues (Fujinaga et al., 2004). Although putative catalytic residues are still unclear, I speculate that candidates for primitive catalytic residues may possibly be Glu38, Glu91, Asn34, Gln35, and Asn95 around the putative substrate binding pockets, roughly similar to the Eqolisin family. The *de novo* proteins including WA20 showed similar K_{M} values to that of natural enzymes, but k_{cat} values of the *de novo* proteins were ~10000-fold lower than those of natural enzymes (Patel et al., 2009). This is consistent with our speculation that the binding pockets exist on the WA20 structure but the catalytic residues are unevolved and not optimized. I expect that the four-helix bundle dimeric structure of WA20 with the pockets may serve as a simple framework for the evolution of *de novo* enzymes.

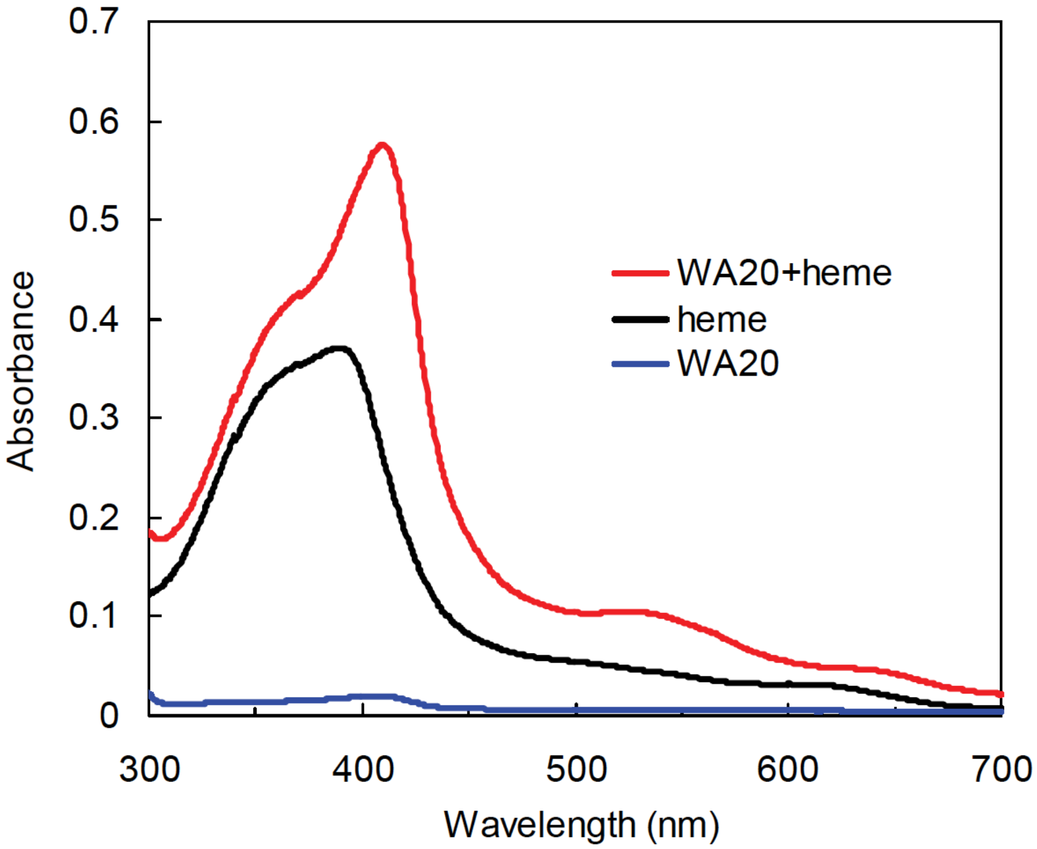


Figure 1-15. Absorbance spectra of WA20 with/without heme. The samples were 10 μM WA20 with 10 μM heme (WA20+heme), 10 μM WA20 (WA20), and 10 μM heme (heme), in 25 mM HEPES buffer (pH 7.0) containing 100 mM NaCl, 10% glycerol and 1 mM DTT. The absorbance spectra were analyzed by V-630BIO UV-Vis Spectrophotometer (JASCO, Tokyo, Japan). The spectrum of WA20 with heme shows a typical Soret peak (~410 nm) for general heme proteins, indicating that WA20 binds to heme.

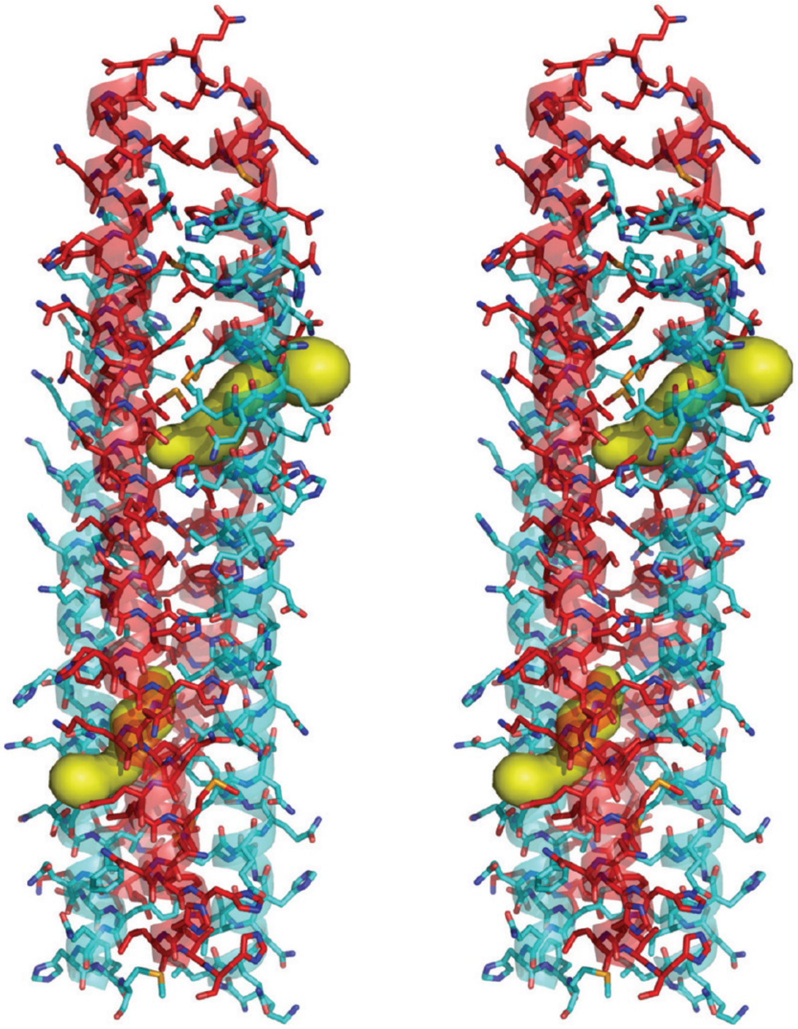


Figure 1-16. The pocket sites on the surface of WA20 (stereoview). There are two major pockets (yellow shapes) on the WA20 structure, depicted using the programs Caver (Beneš et al., 2010) and PyMOL. Chains A and B are shown in red and cyan, respectively.

Chapter 2 Self-Assembling Nano-Architectures Created from a Protein Nano-Building Block Using an Intermolecularly Folded Dimeric *De Novo* Protein

2.1 Introduction

Living organisms are maintained by various self-assembling biomolecules including proteins, nucleic acids, sugars, and lipids. The chemical reconstitution of living matter is one of the ultimate goals of chemistry and synthetic biology, and rational design of artificial biomacromolecules that self-assemble into supramolecular complexes is an important step toward achieving this goal.

In recent years, DNA origami has been developed to design and create various nanostructures (Endo et al., 2013; Rothemund, 2006; Seeman, 2003). The base complementarity of DNA can be used to design rationally artificial nanostructures with versatile two-dimensional (2D) and three-dimensional (3D) shapes, such as polyhedra (Ke, 2014). However, nucleic acids generally consist of only four types of bases, A, T, G, and C, and the limited number of combinations and chemical features might restrict the potential to produce advanced functions.

Proteins consist of 20 types of amino acids, allowing a greater number of chemical properties. Moreover, the enormous varieties of possible sequence combinations expand the probabilities to create diverse and advanced functions. In nature, proteins perform the complex and functional tasks in living organisms because proteins can form intricate and refined tertiary and quaternary structures with versatile chemical properties and functionalities. The design of *de novo* proteins is substantially complicated because of the contribution of many cooperative and long-range interactions. *De novo* protein design and engineering have been performed with mainly two motivations: (1) recapitulation of natural systems to ultimately test our understanding of the principles of protein structure and function and (2) construction of tailor-made proteins as an essential step toward applied biotechnology. Research on *de novo* protein design has progressed toward the construction of novel proteins emanated mainly from three approaches: (1) rational and computational design (Dahiyat and Mayo, 1997; Koga et al., 2012; Kuhlman et al., 2003), (2) combinatorial methods (Keefe and Szostak, 2001), and (3) semirational approaches, including elements of both rational design and combinatorial methods (Hecht et al., 2004; Kamtekar et al., 1993).

As a semirational approach, the binary code strategy has been developed to produce focused libraries of *de novo* proteins designed by the binary patterning of polar and nonpolar residues, and α-helix or β-sheet *de novo* proteins have been created (Hecht et al., 2004; Kamtekar et al., 1993; Wei et al., 2003b; West et al., 1999). From a third-generation library of *de novo* 4-helix bundle proteins designed by binary patterning, a stable and functional *de novo* protein called WA20 was obtained (Bradley et al., 2005; Patel et al., 2009). Recently, I described the crystal structure of the *de novo* protein WA20, revealing an unusual 3D-domain-swapped dimeric structure with a intermolecularly folded 4-helix bundle (Arai et al., 2012). (3D domain swapping is a mechanism of exchanging one structural domain of a protein monomer with that of the identical domain from a second monomer, resulting in an intertwined oligomer (Bennett et al., 1995; Bennett, 1994). Each WA20 monomer ("nunchaku"-like structure), which comprises two long α-helices, intertwines with the helices of another monomer (Figure 2-1A and Figure 2-2). This unusual intertwined topology was first found in the topology-changed structure of the RopA31P mutant (Glykos et al., 1999), which was thermodynamically destabilized (Peters et al., 1997). The structure of WA20 is stable (melting temperature, T_m = ~70°C) and forms a simple rod-like shape with ~8 nm length and ~3 nm diameter (Arai et al., 2012). The stable, simple, and unusual intermolecularly folded structure of the *de novo* protein WA20 raises the possibility of application to basic framework tools in nanotechnology and synthetic biology.

In recent years, several approaches to design artificial self-assembled protein complexes have been developed:

- 3D domain-swapped dimers and fibrous oligomers (Ogihara et al., 2001)
- Nanostructures including cages (Lai et al., 2012a; Lai et al., 2013; Padilla et al., 2001), filaments (Padilla et al., 2001), and lattices (Sinclair et al., 2011) constructed from fusion proteins designed by symmetric self-assembly
- Self-assembling fibers (Pandya et al., 2000; Papapostolou et al., 2007; Sharp et al., 2012), nanostructures (Boyle et al., 2012), and cages (Fletcher et al., 2013) constructed from designed coiled-coil peptide modules
- Metal-directed self-assembling protein complexes (Brodin et al., 2012; Salgado et al., 2007)
- A single-chain polypeptide tetrahedron assembled from coiled-coil segments (Gradisar et al., 2013)
- Computationally designed self-assembling protein nanomaterials with atomic level

accuracy (King et al., 2014; King et al., 2012)

• Other approaches (see refs (Armstrong et al., 2009; Bozic et al., 2013; King and Lai, 2013; Lai et al., 2012b; Radford et al., 2011; Salgado et al., 2010; Woolfson et al., 2012))

In this chapter, to apply the simple, stable, and characteristic intermolecularly folded dimeric structure of WA20 to construct supramolecular nanostructures, I designed and constructed a WA20-foldon a fusion protein of the dimeric *de novo* protein WA20 (Arai et al., 2012) and a trimeric foldon domain (Guthe et al., 2004; Tao et al., 1997) of fibritin from bacteriophage T4 as a simple and versatile nanobuilding block (Figure 2-1). In the nanoarchitectures, the parts of the WA20 and foldon domains resemble a rectilinear framework/edge and corner vertex/node, respectively (Figure 2-1C). Here, I describe the design and construction of the WA20-foldon fusion protein as a novel protein nanobuilding block (PN-Block) and demonstrate its characteristic self-assembling nanoarchitectures.

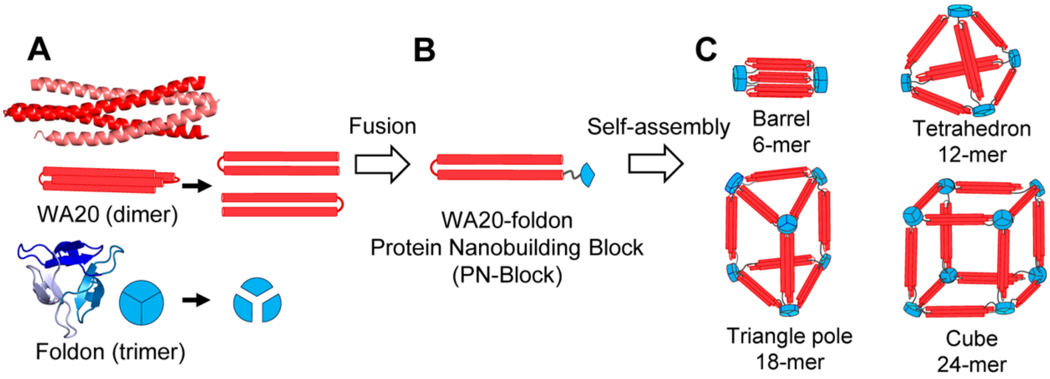


Figure 2-1. Schematics of construction and assemblies of the WA20-foldon fusion protein as a PN-Block. (A) Ribbon representation (see also Figure S1, Supporting Information) and schematics of the intermolecularly folded dimeric WA20 (PDB code 3VJF) (Arai et al., 2012) shown in red, and trimeric foldon domain of T4 phage fibritin (PDB code 1RFO) (Guthe et al., 2004) shown in blue. (B) Construction of the WA20-foldon fusion protein as a PN-Block. (C) Schematics of a nanoarchitecture design by expected self-assemblies of the WA20-foldon. In stable self-assembling complexes, the WA20-foldon is expected to form highly symmetric oligomers in multiples of 6-mer because of the combination of the WA20 dimer and foldon trimer.

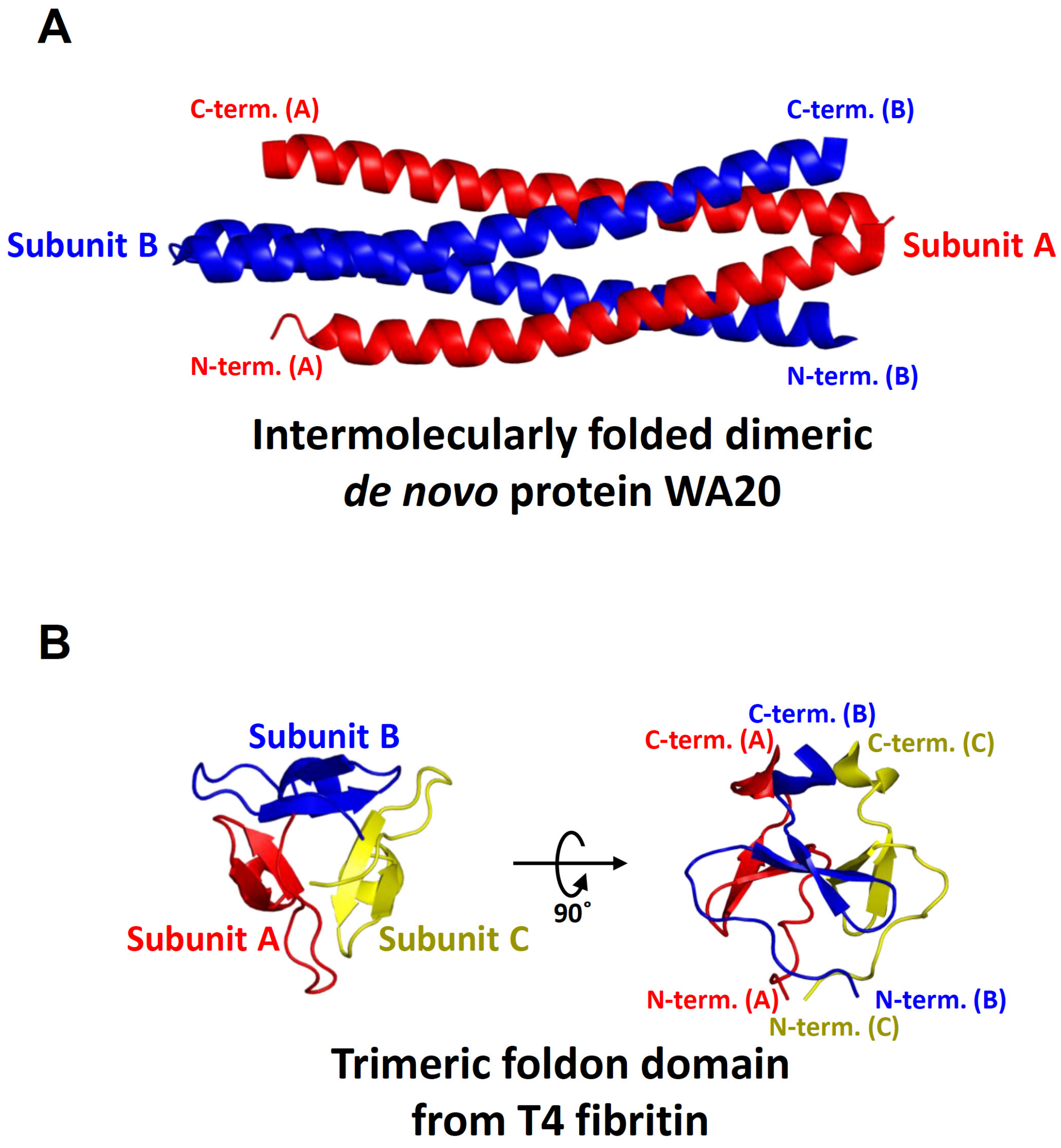


Figure 2-2. Ribbon representation of (A) the unusual intertwined intermolecularly-folded (domain-swapped) dimeric structure of WA20 (PDB code 3VJF) (Arai et al., 2012) and (B) the trimeric structure of the foldon domain of fibritin from bacteriophage T4 (PDB code 1RFO) (Guthe et al., 2004). The N- and C-terminal sites are shown in the structures.

2.2 Materials and methods

Construction of Expression Plasmid of WA20-Foldon.

The DNA fragment encoding the *de novo* protein WA20 was prepared from plasmid pET3a-WA20 (Arai et al., 2012; Patel et al., 2009) by polymerase chain reaction (PCR) using KOD-Plus-Neo DNA polymerase (Toyobo, Osaka, Japan) and primers, T7 promoter primer and WA20RV_*Hind*III (Table 2-1). The amplified fragment was digested by *Nde*I and *Hind*III and cloned into pET32b(+) (Merck Millipore, Darmstadt, Germany) between the *Nde*I and *Hind*III sites to construct the plasmid pET-WA20, i.e., the Trx tag was removed and replaced with WA20. The DNA fragment encoding the foldon domain (residues 458−483 in T4 phage fibritin) was prepared by annealing and extension reactions with the two synthesized oligonucleotides (Foldon_*Hind*III-*Not*I_FW and Foldon_*Xho*I_RV) (Table 2-1). The DNA fragment encoding the foldon domain digested *Hind*III and *Xho*I and cloned into pET-WA20 between the *Hind*III and *Xho*I sites to give the expression plasimid pET-WA20-foldon. The amino acid sequence of the WA20-foldon fusion protein with a His$_6$ tag at the C terminal is shown in Figure 2-3.

Table 2-1. Oligonucleotide Primer Sequences Used in This Study

Primer name	Sequence (5’→3’)
T7 promoter primer	TAATACGACTCACTATAGGG
WA20RV_*Hind*III	GCGGCAAAAGCTTGCGATGTACAAGGTGGTGGAAGT
Foldon_*Hind*III-*Not*I_FW	GGGGCAAAGCTTGCGGCCGCGTATATTCCTGAAGCTCCAAGAGATGGGCAAGCGTACGTTCGTAAAGATG
Foldon_*Xho*I_RV	GGGGCCCTCGAGAAAGGTAGAAAGTAATACCCATTCGCCATCTTTACGAACGTACGCT

MYGKLNKLVE HIKELLQQLN KNWHRHQGNL HDMNQQMEQL FQEFQHFMQG
NQDDGKLQNM IHEMQQFMNQ VDNHLQSESD TVHHFHNKLQ ELMNNFHHLV
HRKLAAAYIP EAPRDGQAYV RKDGEWVLLS TFLEHHHHHH

Figure 2-3. Amino acid sequence of the WA20-foldon protein with a C-terminal His$_6$ tag. The regions of WA20 and foldon domains are colored red and blue, respectively. The WA20-foldon protein consists of 140 amino-acid residues, and the theoretical average molecular mass of this polypeptide is 16959.82 Da.

Protein Expression and Purification of WA20-Foldon.

The WA20-foldon protein with a His_6 tag was expressed in *Escherichia coli* BL21 Star (DE3) (Invitrogen, Carlsbad, CA) harboring pET-WA20-foldon using LB broth, Lennox (Nacalai Tesque, Kyoto, Japan) with 50 μg/mL ampicillin sodium salt at 37 °C. The expression was induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside at OD_{600} (optical density at 600 nm) = ~0.8, and cells were further cultured for 3−4 h at 37 °C. The protein was extracted from the harvested cells by sonication in a lysis buffer (50 mM sodium phosphate buffer (pH 7.0), containing 300 mM NaCl, 10% glycerol). The protein was purified by immobilized metal ion affinity chromatography (IMAC) with a HisTrap HP column (GE healthcare, Little Chalfont, UK) and eluted using a linear gradient of imidazole (equilibration buffer: 20 mM sodium phosphate buffer (pH 7.4) containing 500 mM NaCl, 200 mM L-ArgHCl, 10% glycerol, and 20 mM imidazole; elution buffer: 20 mM sodium phosphate buffer (pH 7.4) containing 500 mM NaCl, 200 mM L-ArgHCl, 10% glycerol, and 500 mM imidazole). The protein samples were concentrated with Amicon ultra centrifugal filters (Merck Millipore). Each form of the WA20-foldon protein was further purified repeatedly by size exclusion chromatography (SEC) (20 mM HEPES buffer (pH 7.5) containing 100 mM NaCl, 200 mM L-ArgHCl, and 10% glycerol) with HiLoad 16/600 Superdex 200 pg and Superdex 200 Increase 10/300 GL columns (GE healthcare). SEC Multi-Angle Light Scattering (SEC-MALS). SEC-MALS experiments were performed using a 1260 Infinity HPLC system (Agilent Technologies, Santa Clara, CA) equipped with a Superdex 200 Increase 10/300 GL column, which was connected in line with a miniDAWN TREOS multiangle static light-scattering detector (Wyatt Technology, Santa Barbara, CA). The data were collected in phosphate buffered saline (PBS, pH 7.4:1 mM KH_2PO_4, 3 mM Na_2HPO_4, and 155 mM NaCl) at 20 °C and analyzed using ASTRA 6 software (Wyatt Technology). The d*n*/d*c* value (0.185 mL/g) was generally used for proteins, and the extinction coefficient (0.913 mL mg^{-1} cm^{-1}) for the WA20-foldon was calculated from the amino acid sequence (Pace et al., 1995).

Analytical Ultracentrifugation (AUC).

AUC experiments were performed at 20 °C using an analytical ultracentrifuge, Optima XL-I (Beckman Coulter, Brea, CA) with a An-50 Ti rotor. For sedimentation velocity experiments, cells with a standard Epon two channel centerpiece and sapphire windows were used. The sample (400 μL) and reference buffer (420 μL; 20 mM

HEPES buffer (pH 7.5) containing 100 mM NaCl, 200 mM L-ArgHCl, and 10% glycerol) were loaded into the cells. Absorbance scans at 280 nm were collected at 10 min intervals during sedimentation at 50 × 10^3 rpm. The sedimentation velocity experiments were performed at protein concentrations of 1.2, 0.6, and 0.3 mg/mL. Partial specific volume of the protein was calculated from standard tables using the SEDNTERP program (Laue et al., 1992). The solvent density (1.0472 g/cm^3) and solvent viscosity (1.5639 cP) were determined using DMA 4500 M and AMVn (Anton Paar, Graz, Austria), respectively. The resulting scans were analyzed using the continuous distribution ($c(s)$) analysis module in the SEDFIT program (version 14.7g) (Schuck, 2000). Sedimentation coefficient increments of 200 were used in the appropriate range for each sample. The frictional coefficient was allowed to float during fitting. The weight average sedimentation coefficient was obtained by integrating the range of sedimentation coefficients in which peaks were present. The values of sedimentation coefficient were corrected to 20 °C in pure water ($s_{20,\mathrm{w}}$).

Sedimentation equilibrium experiments were carried out in cells with a six-channel centerpiece and quartz windows at 20 °C. The sample concentrations were 0.6, 0.3, and 0.15 mg/mL. Data were obtained at 4, 9, and 20 × 10^3 rpm. A total equilibration time of 48 h was used for each speed, with absorbance scans at 280 nm taken every 4 h to ensure that equilibrium had been reached. Data analysis was performed by global analysis of data sets obtained at different loading concentrations and rotor speeds using SEDPHAT program (version 10.58d) (Vistica et al., 2004).

Small-Angle X-ray Scattering (SAXS).

SAXS measurements were performed on the WA20-foldon oligomers, chicken egg white lysozyme (Wako Pure Chemical Industries, Osaka, Japan), and WA20 (Arai et al., 2012) in 20 mM HEPES buffer (pH 7.5) containing 100 mM NaCl, 200 mM L-ArgHCl, 10% glycerol, and 1 mM dithiothreitol at 4 °C (Table 2-2). The SAXS measurements were performed by SAXSess mc^2 (Anton Paar) with a SAXSess camera (Anton Paar) attached to a sealed-tube anode X-ray generator (GE Inspection Technologies, Huerth, Germany). The line-shaped and monochromatic X-ray beams of CuKα radiation (λ = 0.1542 nm) were provided by a Göbel mirror and a block collimator. Liquid samples were filled into a temperature-controlled vacuum-tight quartz capillary cell. An imaging plate detector that recorded the primary beams attenuated by a semitransparent beam stop and the scattered X-rays was read out by a Cyclone Phosphor System (PerkinElmer, Waltham, MA). The 2D scattering patterns were integrated into one-dimensional scattering intensities, $I(q)$, as a function of the magnitude of the scattering vector, $q = (4\pi/\lambda)\ \sin(\theta/2)$, using the SAXSQuant program (Anton Paar), where θ is the total scattering angle. All $I(q)$ data were normalized to have a uniform primary intensity at $q = 0$ for transmission calibration. The background scattering contributions from the capillary and solvent were corrected. The absolute intensity calibration was performed by referring to water as a secondary standard (Orthaber et al., 2000).

Generally, the scattering intensity for a colloidal dispersion is given by the product of the form factor, $P(q)$, and the structure factor, $S(q)$:

$$I(q) = nP(q)S(q)$$

where n is the number density of the particle. If interparticle interactions including the excluded volume effect and electrostatic interaction can be neglected (i.e., $S(q) = 1$), the scattering intensity is proportional to the form factor. Our experimental condition can be regarded as a situation in which the structure factor is almost unity, i.e., $I(q) \approx nP(q)$, because of a low protein concentration and a high salt concentration of the solvent. The form factor is given by the Fourier transformation of the pair-distance distribution function, $p(r)$, which describes the size and shape of the particle:

$$P(q) = 4\pi \int_0^{D_{\max}} p(r) \frac{\sin qr}{qr} \mathrm{d}r$$

where $D_{\max}$ is the maximum intraparticle distance. To obtain $p(r)$ of the proteins and their self-assemblies using a virtually model-free routine, I used the indirect Fourier transformation (IFT) technique (Brunner-Popela and Glatter, 1997; Glatter, 1980b; Glatter and Kratky, 1982). The forward absolute scattering intensity, $I(q \rightarrow 0)$, was extrapolated from the data. The radius of gyration, R_g, was estimated by the Guinier approximation (Glatter and Kratky, 1982).

Table 2-2. Experimental Samples of Proteins for SAXS Measurements

Sample name	Protein concentration [mg/mL]
Lysozyme (as a M_w reference standard)	4.8
WA20 (as a control sample)	4.3
Small form (S form) of a WA20-foldon oligomer	4.3
Middle form (M form) of a WA20-foldon oligomer	5.1
Large form (L form) of a WA20-foldon oligomer	3.7
Huge form (H form) of a WA20-foldon oligomer	0.7

Rigid-Body Modeling of Oligomeric Structures of WA20-Foldon.

The rigid-body models of WA20-foldon oligomeric structures were constructed using the program COOT (Emsley et al., 2010) based on the crystal structure of WA20 [protein data bank (PDB) code, 3VJF] (Arai et al., 2012) and the solution structure of the foldon domain (PDB code, 1RFO) (Guthe et al., 2004) with a consideration of their N- and C-terminal directions and two- and three-fold symmetries. The rigid-body models were manually and iteratively refined to minimize differences in the *p*(*r*) and *I*(*q*) calculated from the models and those obtained from SAXS experiments (Figures 2-4−2-6).

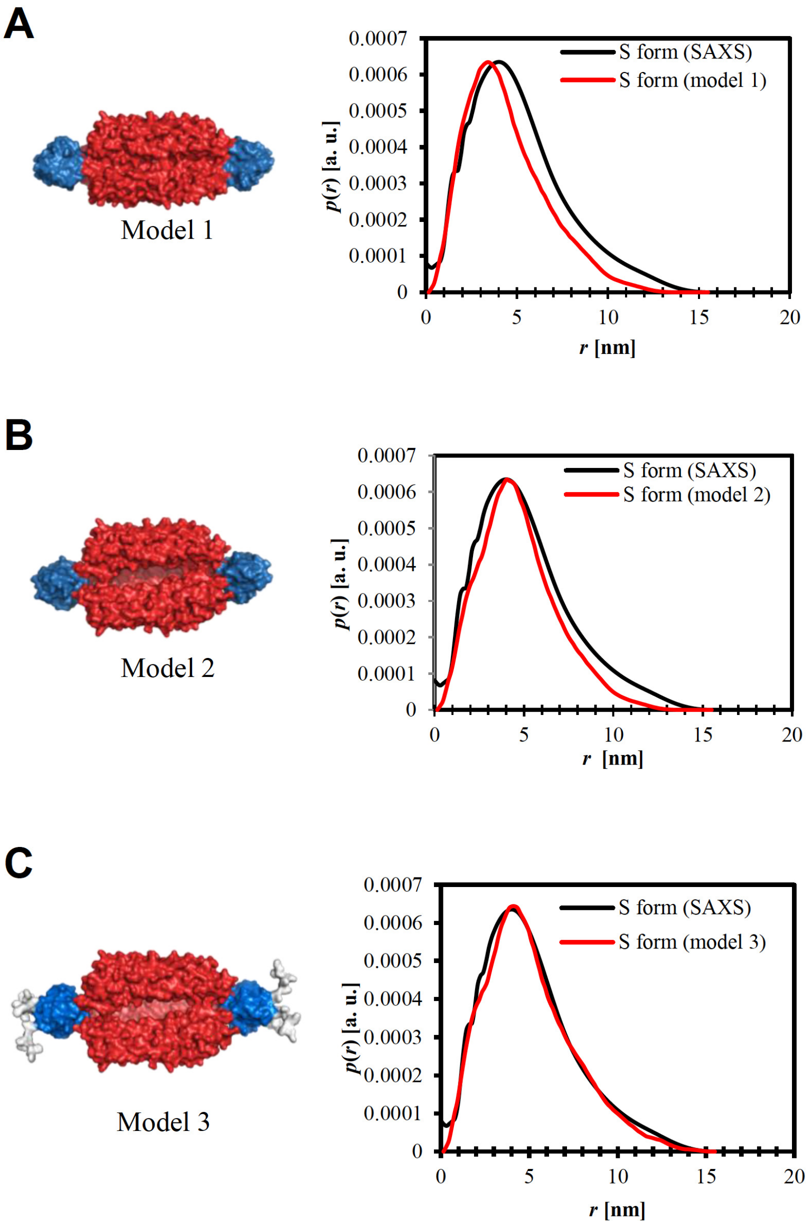


Figure 2-4. Typical process of the rigid-body modeling of the S form of the WA20-foldon hexamer. The rigid-body model structures were constructed based on the crystal structure of WA20 (PDB code 3VJF) (Arai et al., 2012) and the solution structure of foldon domain (PDB code 1RFO) (Guthe et al., 2004) with a consideration of their N- and C-terminal directions and 2- and 3-fold symmetries. The rigid-body models were manually and iteratively refined to minimize differences in the pair-distance distribution functions *p*(*r*) calculated from the models and *p*(*r*) obtained from the SAXS experiment. In the left panels, the domains of the WA20, foldon, and

His_6 tag are shown in red, blue, and light gray, respectively. The right panels show the pair distance functions, *p*(*r*), of the S form of the WA20-foldon as obtained by the SAXS experiment (black line) and that simulated from the rigid-body model structures of the S form (red line). (A) The early-stage rigid-body model (model 1) of the S form of the WA20-foldon without a His6 tag. There is no space in the center where three domains of WA20 dimers contact each other. (B) The middle-stage rigid-body model (model 2) of the S form of the WA20-foldon without a His_6 tag. The locations of three domains of WA20 dimers were adjusted to make little space in the central region. (C) The final-stage rigid-body model (model 3) of the S form of the WA20-foldon with His_6 tags. The His_6 tags were added to model 2. The model 3 was adopted as the final rigid-body model structure of the S form (Figure 2-16A).

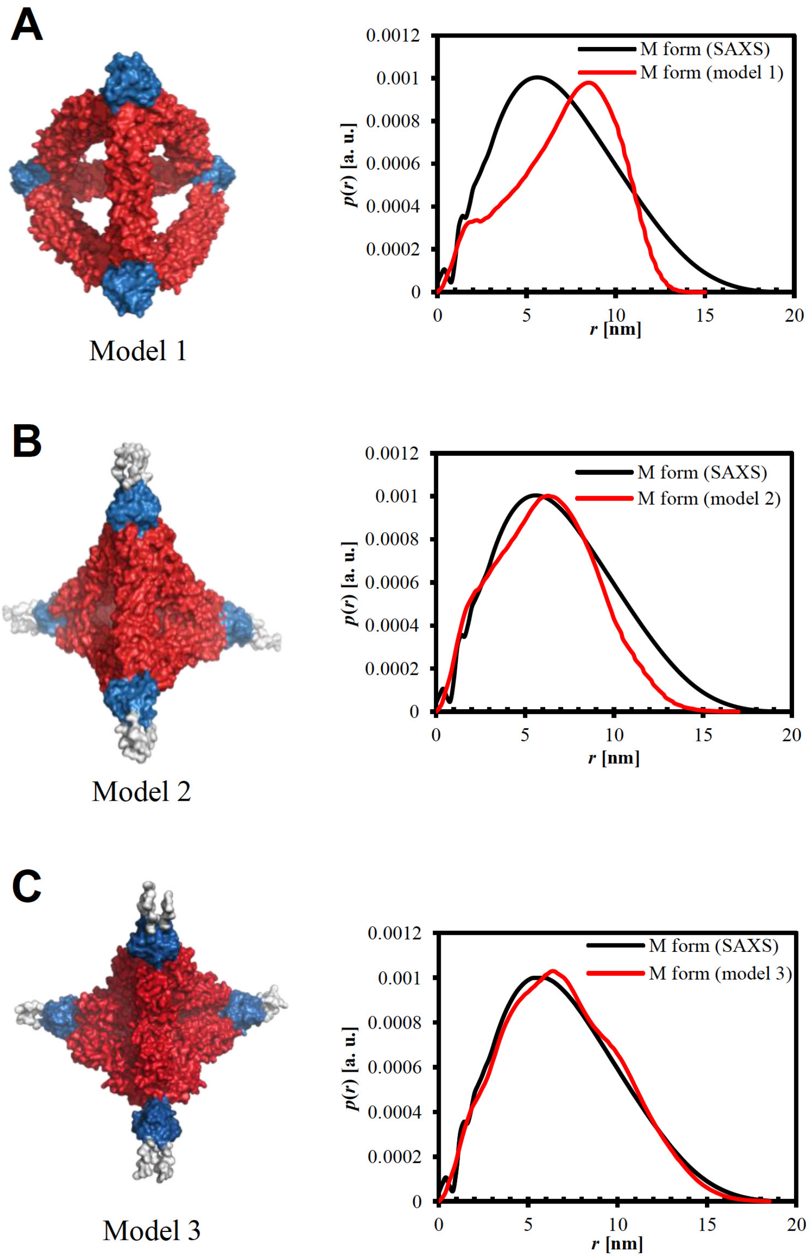


Figure 2-5. Typical process of the rigid-body modeling of the M form of the WA20-foldon dodecamer. The rigid-body model structures were constructed based on the crystal structure of WA20 (PDB code 3VJF) (Arai et al., 2012) and the solution structure of foldon domain (PDB code 1RFO) (Guthe et al., 2004) with a consideration of their N- and C-terminal directions and 2- and 3-fold symmetries. The rigid-body models were manually and iteratively refined to reduce differences in the pair-distance distribution functions *p*(*r*) calculated from the models and *p*(*r*) obtained from the SAXS experiment. In the left panels, the domains of the WA20, foldon, and His$_6$ tag are shown

in red, blue, and light gray, respectively. The right panels show the pair distance functions, *p*(*r*), of the M form of the WA20-foldon as obtained by the SAXS experiment (black line) and that simulated from the rigid-body model structures of the M form (red line). (A) The early-stage rigid-body model (model 1) of the M form of the WA20-foldon without a His$_6$ tag. There are large space in the center of the tetrahedron-like structure. (B) The middle-stage rigid-body model (model 2) of the M form of the WA20-foldon with His$_6$ tags. The locations and configurations of WA20 domains were changed to make smaller space in the center of the tetrahedron-like structure. (C) The final-stage rigid-body model (model 3) of the M form of the WA20-foldon with His$_6$ tags. The locations and configurations of WA20 and foldon domains were adjusted to fit the *p*(*r*) calculated from the model to the p(r) obtained from the SAXS experiment. The central regions of the helices of the WA20 domains, which were potentially-designed loop regions (Arai et al., 2012), were curved. The model 3 was adopted as the final rigid-body model structure of the M form (Figure 2-17A).

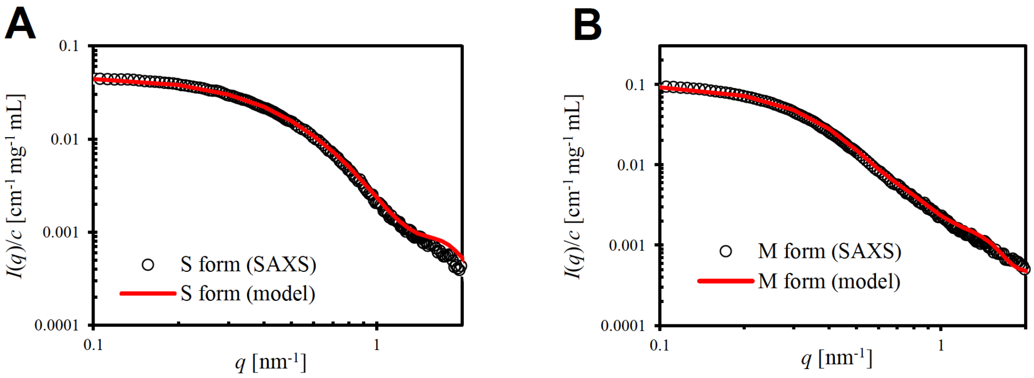


Figure 2-6. (A) Concentration-normalized SAXS intensity, $I(q)/c$, of the S form of the WA20-foldon as obtained by the SAXS experiment (black open circle) and that simulated from the rigid-body model structure (Figure 2-16A: Model 3 in Figure 2-4) of the S form (red line). (B) The $I(q)/c$ of the M form of the WA20-foldon as obtained by the SAXS experiment (black open circle) and that simulated from the rigid-body model structure (Figure 2-17A: Model 3 in Figure 2-5) of the M form (red line).

Calculation of *ab Initi*o Low-Resolution Shapes of WA20-Foldon.

Low-resolution supramolecular shapes were reconstructed from the SAXS intensity profiles using an *ab initio* procedure of the program DAMMIN (Svergun, 1999) in the ATSAS program package (Petoukhov et al., 2012). Using dummy atom minimization (DAMMIN), a protein molecule is approximated by densely packed small spheres (dummy atoms). Because the low-resolution *ab initio* modeling procedure does not consider the internal structure, relatively low-angle data ($qR_g < 7$) were used. Simulated annealing calculations were performed several times to determine a configuration that fits the SAXS data, starting from the dummy atoms placed at random coordinates within the search space, a sphere of diameter D_{max}, with/without a consideration of two- and/or three-fold symmetry constraints.

2.3 Results and Discussion

Design of WA20-Foldon as a PN-Block to Construct Self-assembling Nanostructures.

As shown in (Figure 2-1), to construct self-assembling nanostructures, the first PN-Block, the WA20-foldon, which utilized the unusual intermolecularly folded dimeric *de novo* protein WA20 as a framework, was designed by reference to the "nanohedra" strategy (Lai et al., 2012a; Padilla et al., 2001) using symmetry to design the nanostructures. The design of a dimer−trimer PN-Block is notably a versatile and powerful approach as a geometrically based building block to construct several polyhedra with three edges from one node, such as a tetrahedron, hexahedron, and dodecahedron. The WA20-foldon fusion protein was constructed as a fundamental PN-Block by fusing the dimeric *de novo* protein WA20 and trimeric foldon domain of T4 phage fibritin (Guthe et al., 2004; Tao et al., 1997) with an alanine-rich short linker (the amino-acid sequence: KLAAA) (Figure 2-1 and Figure 2-3). The residues in this linker have relatively high helix-forming propensities (Pace and Scholtz, 1998). The foldon domain consisting of only 26 residues is suitable for a trimeric connecting vertex/node part because it promotes stable trimerization by fast folding, (Guthe et al., 2004) and its application to the construction of engineered bionanotubes was reported (Yokoi et al., 2010). In the stable self-assembling complexes of the WA20-foldon, it is expected to form several oligomers in multiples of 6-mer because of the combination of the WA20 dimer and foldon trimer (Figure 2-1C).

Self-Assembling Oligomers of WA20-Foldon Produced in *E. coli*.

The WA20-foldon protein with a His_6 tag was expressed in a soluble fraction in *E. coli* and purified by IMAC. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) of the purified fraction shows almost a single band (Figure 2-7A). However, native PAGE of the same fraction shows several discrete bands (Figure 2-7B), suggesting that the WA20-foldon forms several homooligomeric states simultaneously in the soluble fraction in *E. coli*. The stable WA20-foldon oligomers were further purified by SEC, and the four discrete bands were separated in native PAGE (Figure 2-7D), but they showed the same single band in SDS-PAGE (Figure 2-7C). The major discrete bands of the WA20-foldon oligomers were named the small form (S form), middle form (M form), large form (L form), and huge form (H form) in order from the lower band in native PAGE (Figure 2-7D). In addition, Figure 2-8 shows that the quality of the proteins did not change practically in SDS-PAGE and native PAGE after at least ∼8 months of storage at 4 °C, suggesting that the WA20-foldon protein is very stable at the level of not only the polypeptide but also its oligomeric states (i.e., no observable exchange between the forms).

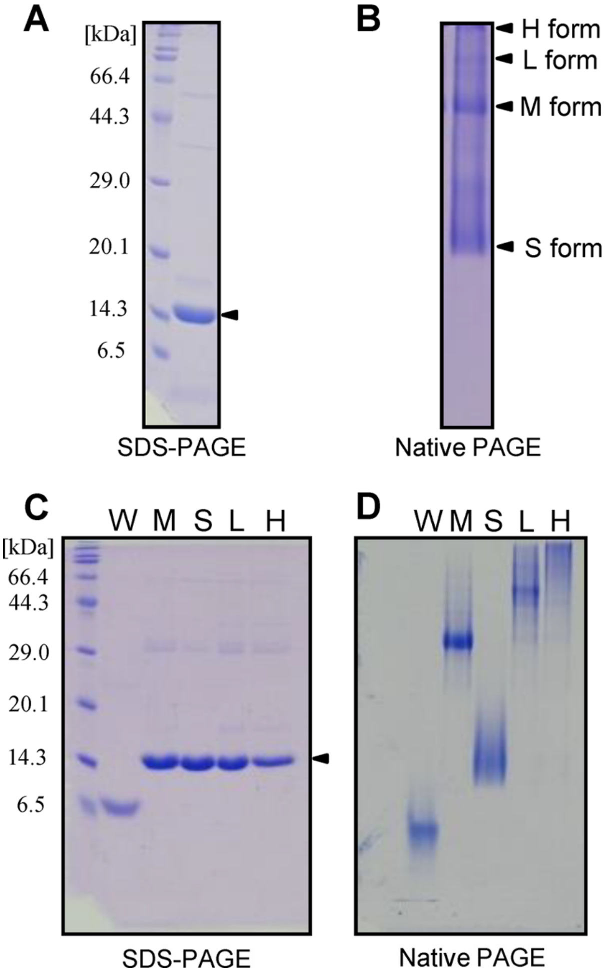


Figure 2-7. Polyacrylamide gel electrophoresis of WA20-foldon. (A) SDS-PAGE (17.5% polyacrylamide gel) and (B) native PAGE (7.5% polyacrylamide gel) of the WA20-foldon after IMAC purification. (C) SDS-PAGE (17.5% gel) and (D) native PAGE (7.5% gel) of each form of the WA20-foldon after SEC purification. S: S form; M: M form; L: L form; H: H form; W: WA20 as a control sample. Proteins were stained with Coomassie brilliant blue. The protein molecular weight marker (broad) (Takara Bio, Otsu, Japan) was used for SDS-PAGE.

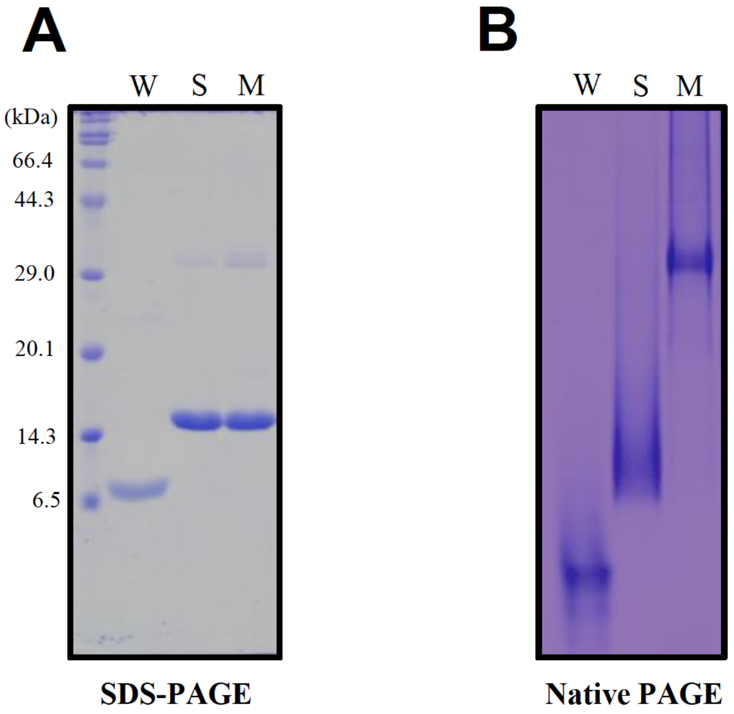


Figure 2-8. (A) SDS-PAGE (17.5% gel) and (B) Native PAGE (7.5% gel) of the S form (S) and M form (M) of the purified WA20-foldon after ~8 months of storage at 4℃. The WA20 (W) was used as a control sample. The protein molecular weight marker (broad) (Takara Bio) was used for SDS-PAGE. The proteins were stained with Coomassie brilliant blue.

Analyses of Oligomeric States of WA20-Foldon.

To estimate the molecular mass of the WA20-foldon oligomers, I first analyzed each form of the purified WA20-foldon using SEC. Figure 2-9 shows the chromatograms of each form of the WA20-foldon. The peaks of the S and M forms are sharp and nearly symmetrical-shaped single peaks, suggesting that they are almost monodispersed. In contrast, the peak shapes of the L and H forms are broad, suggesting that they are polydispersed. Table 2-3 shows the elution volume and estimated molecular mass using SEC with the standard calibration curve (Figure 2-10). The molecular masses of the individual forms of the WA20-foldon were estimated to be 84 kDa (S form), 195 kDa (M form), 284 kDa (L form), and 415 kDa (H form). However, the molecular mass of the H form might be underestimated because the H form peak overlapped with the column void volume (8.5 mL, elution volume of Blue Dextran 2000). Because the molecular mass of a WA20-foldon protomer is 17.0 kDa, oligomeric states of the individual forms of the WA20-foldon were roughly calculated to be 5-mer (S form), 11-mer (M form), 17-mer (L form), and 24-mer (H form) (Table 2-3). To determine the absolute molecular mass without reference standard samples, I performed SEC-MALS (Wyatt, 1993) experiments (Table 2-4; and Figure 2-11). The molecular masses of the individual forms of the WA20-foldon were determined to be 101 kDa (S form), 199 kDa (M form), 299 kDa (L form), 392 kDa (H1 form: the lower-mass refined component in the H form), and 543 kDa (H2 form: the higher mass refined component in the H form), and their oligomeric states were estimated to be 6-mer (S form), 12-mer (M form), 18-mer (L form), 23-mer (H_1 form), and 32-mer (H_2 form). These results suggest that the WA20-foldon forms regularly discrete oligomers in multiples of 6-mer because of the combination of the WA20 dimer and foldon trimer (Figure 2-1C).

As the third independent method to analyze molecular mass of the WA20-foldon oligomers, I performed AUC experiments. First, I performed sedimentation velocity experiments (Figure 2-12). The $c(s)$ distribution of the sample S shows the presence of a single species corresponding to the S form with a sedimentation coefficient ($s_{20,\mathrm{w}}$) of 5.20 (±0.02) S (Figure 2-12A). The molecular mass was estimated to be 91 (±2) kDa. The frictional ratio, f/f_0, for the S form was calculated to be 1.45 (±0.02). The value of the frictional ratio represents the degree of deviation, due to hydration, rugosity, asymmetry, and expansion of the molecule. The S form has a larger frictional ratio of 1.45 than the typical values of 1.05−1.30 for globular proteins (Tanford, 1961), implying that it has an atypical shape (e.g., an elongated shape).

The *c*(*s*) distribution of the sample M shows the presence of two main species corresponding to the M and S forms (Figure 2-12B). The large peak has a sedimentation coefficient ($s_{20,\mathrm{w}}$) of 7.61 (±0.01) S for the M form, and another small peak has that of 5.34 (±0.02) S for the S form. Because the *c*(*s*) distribution shape, the sedimentation coefficient value, and the ratio of the peak height did not significantly change with the protein concentration, the two species are independent molecules, not in equilibrium system (i.e., self-association system and/or subunit exchanging system), in the time scale of the experiments.

The *c*(*s*) distribution of the sample L shows that the solution contains various species in the broad range (Figure 2-12C). The major peak has a sedimentation coefficient ($s_{20,\mathrm{w}}$) of 9.93 (±0.04) S probably for the L form and another peak has that of 6.49 (±0.02) S. However, these values may possess lower reliability because the peaks are small and broad. Also, the *c*(*s*) distribution of the sample H shows that the solution contains various species (Figure 2-12D). It is difficult to analyze this because of the small and broad peaks.

Judging from the results of sedimentation velocity experiments, I further performed sedimentation equilibrium AUC experiments of the samples S and M to determine molecular mass of the S and M forms of the WA20-foldon (Figure 2-13). The molecular mass of the S form was determined to be 96 (±1) kDa from the sedimentation equilibrium experiments of the sample S (Figure 2-13A). The molecular masses of the two species in the sample M were determined to be 180 (±8) kDa for the M form and 100 (±5) kDa for the S form from the sedimentation equilibrium experiments using two species analysis model (Figure 2-13B). From these results, the oligomeric states of the S and M forms were estimated to be 6-mer and 11-mer, respectively.

In addition, the matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrum of the S form of the WA20-foldon cross-linked by glutaraldehyde shows the mass peak of m/z = 111449 assignable to 6-mer considering mass increase due to chemical modification (Figure 2-14). However, mass spectra peaks of the other forms of the WA20-foldon were not observed probably because of the higher molecular masses of the M, L, and H forms.

Furthermore, Figure 2-15A shows SAXS intensities of a series of the WA20-foldon oligomers, WA20, and chicken egg lysozyme. Assuming that these proteins have practically identical scattering length densities and specific volumes and that the structure factor $S(q) \approx 1$ for dilute samples, the forward scattering intensity normalized by protein concentration, $I(q\rightarrow 0)/c$, is proportional to the weight-average molecular mass (M_{w}). Lysozyme (M_{w} = 14.3 kDa) was used as a molecular mass reference standard. The average molecular masses of the individual forms of the WA20-foldon were estimated to be 97.1 kDa (S form), 224 kDa (M form), 331 kDa (L form), and 641 kDa (H form), and oligomeric states of the individual forms of the WA20-foldon were roughly calculated to be 6-mer (S form), 13-mer (M form), 19-mer (L form), and 38-mer (H form) (Table 2-5).

Table 2-6 summarizes the results of molecular mass and oligomeric states of the WA20-foldon oligomers, determined by the multifaceted experiments. Because the stable form of the WA20-foldon should form the oligomers in multiples of 6-mer in the light of the combination of the WA20 dimer and foldon trimer, overall these results indicate that the individual forms of the WA20-foldon exist as hexamer (6-mer) for the S form, dodecamer (12-mer) for the M form, octadecamer (18-mer) for the L form, and a mixture of 24-mer, 30-mer, and perhaps higher oligomers for the H form.

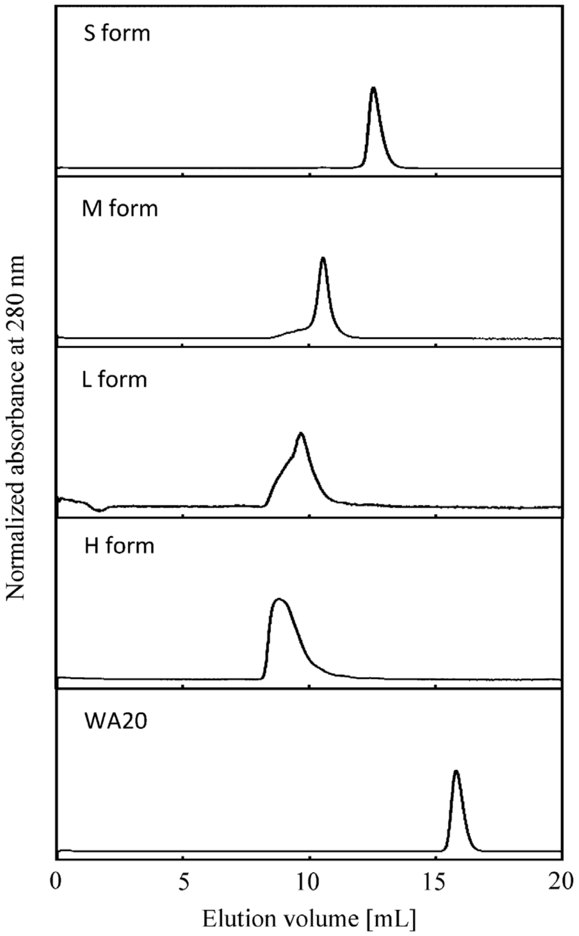


Figure 2-9. SEC chromatograms of each form of the purified WA20-foldon on Superdex 200 Increase 10/300 GL. WA20 was used as a control sample. The elution volume and molecular mass, estimated by the standard calibration curve (Figure 2-10) are summarized in Table 2-3.

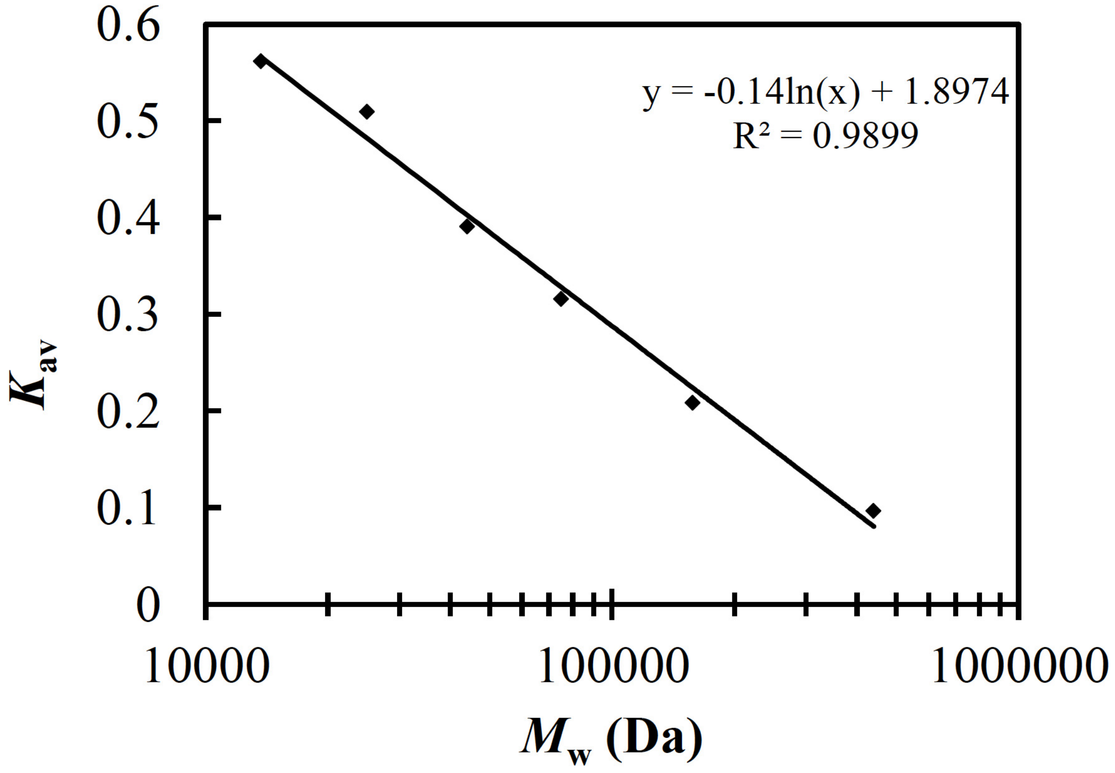
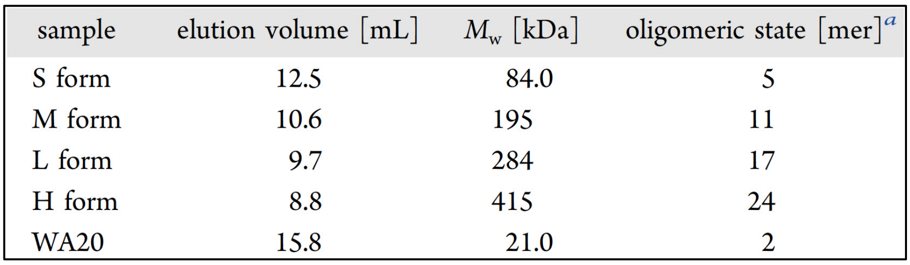


Figure 2-10. Size exclusion chromatography (SEC) calibration curve for standard proteins on a Superdex 200 Increase 10/300 GL column (GE Healthcare). The calibration protein samples are ferritin (440000 Da), aldolase (158000 Da), conalbumin (75000 Da), ovalbumin (44000 Da), chymotrypsinogen A (25000 Da), and ribonuclease A (13700 Da) in Gel Filtration Calibration kits LMW and HMW (GE Healthcare). The x-axis is molecular mass (M_w) in log scale. The partition coefficient K_{av} values for each standard protein were calculated using the equation, $K_{av} = (V_e - V_o)/(V_t - V_o)$, where V_e is elution volume for the protein, and V_o is column void volume (elution volume of blue dextran 2000, V_o = 8.5 mL), and V_t is total bed volume (V_t = 24 mL). The calibration curve was determined by linear least-squares analysis.

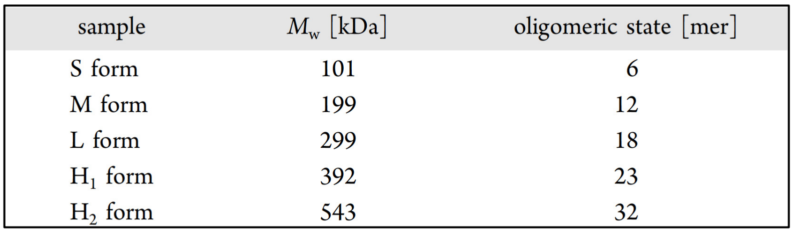
Table 2-3. Elution Volume and Estimated Molecular Mass of WA20-Foldon Oligomers in SEC Experiments



sample	elution volume [mL]	M_w [kDa]	oligomeric state [mer][a]
S form	12.5	84.0	5
M form	10.6	195	11
L form	9.7	284	17
H form	8.8	415	24
WA20	15.8	21.0	2

[a] The molecular mass of a WA20-foldon protomer is 17.0 kDa.

Table 2-4. Molecular Mass of WA20-Foldon Oligomers Determined by SEC-MALS Experiments



sample	M_w [kDa]	oligomeric state [mer]
S form	101	6
M form	199	12
L form	299	18
H_1 form	392	23
H_2 form	543	32

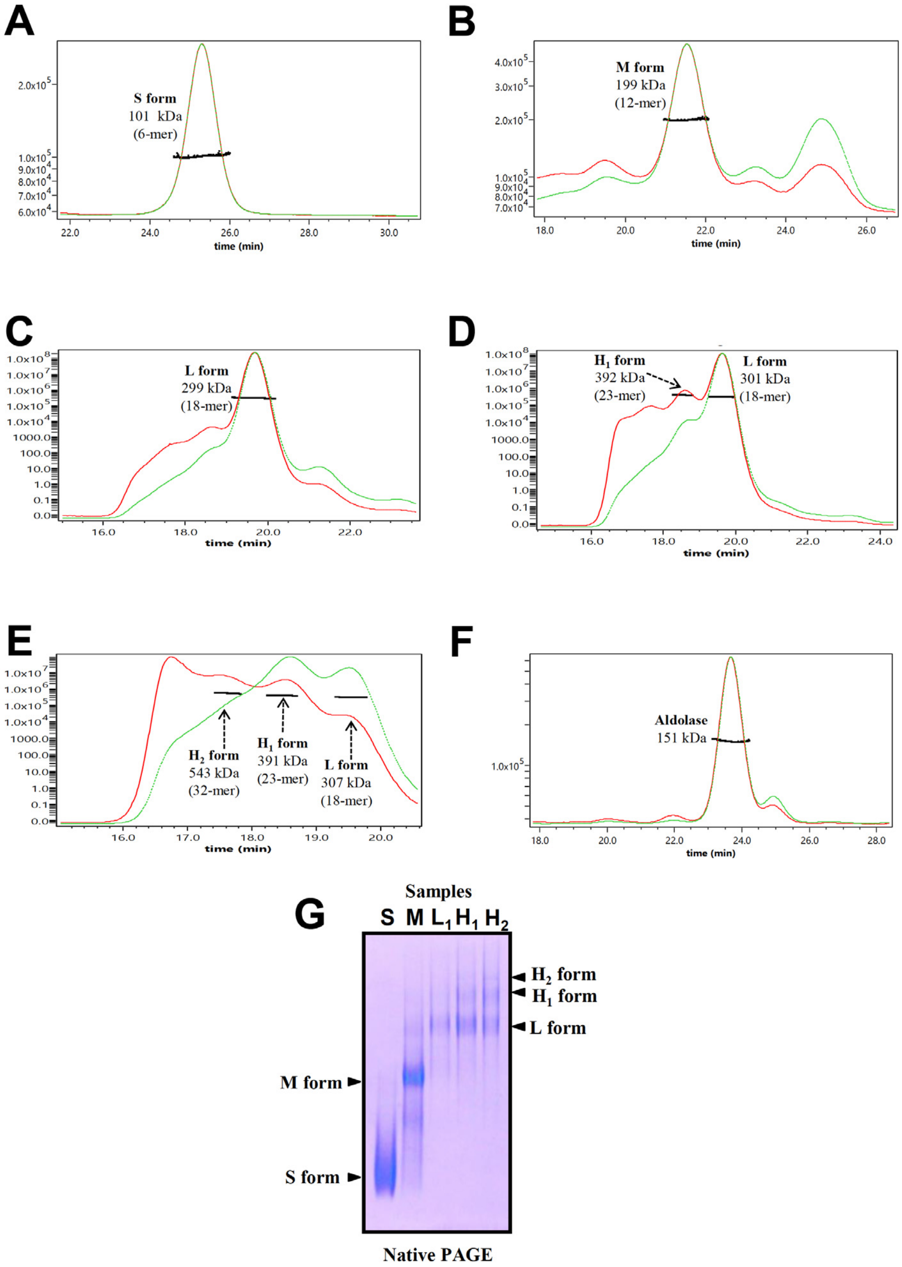


Figure 2-11. Results of SEC-MALS experiments of the WA20-foldon oligomers with the UV (green) and 90° light scattering (red) chromatograms [arbitrary unit]. (A) sample S, (B) sample M, (C) sample L_1, (D) sample H_1, (E) sample H_2, (F) aldolase (M_w = 158 kDa) as a control sample. Each molecular mass is shown as a black line across the elution peak. The samples L_1, H_1, and H_2 were further purified by SEC with a Superose 6 10/300 GL column (GE healthcare) before the SEC-MALS experiments. The molecular mass of a WA20-foldon protomer is 17.0 kDa. (G) Native PAGE (7.5% gel) of the WA20-foldon samples for the SECMALS experiments. The proteins were stained with Coomassie brilliant blue.

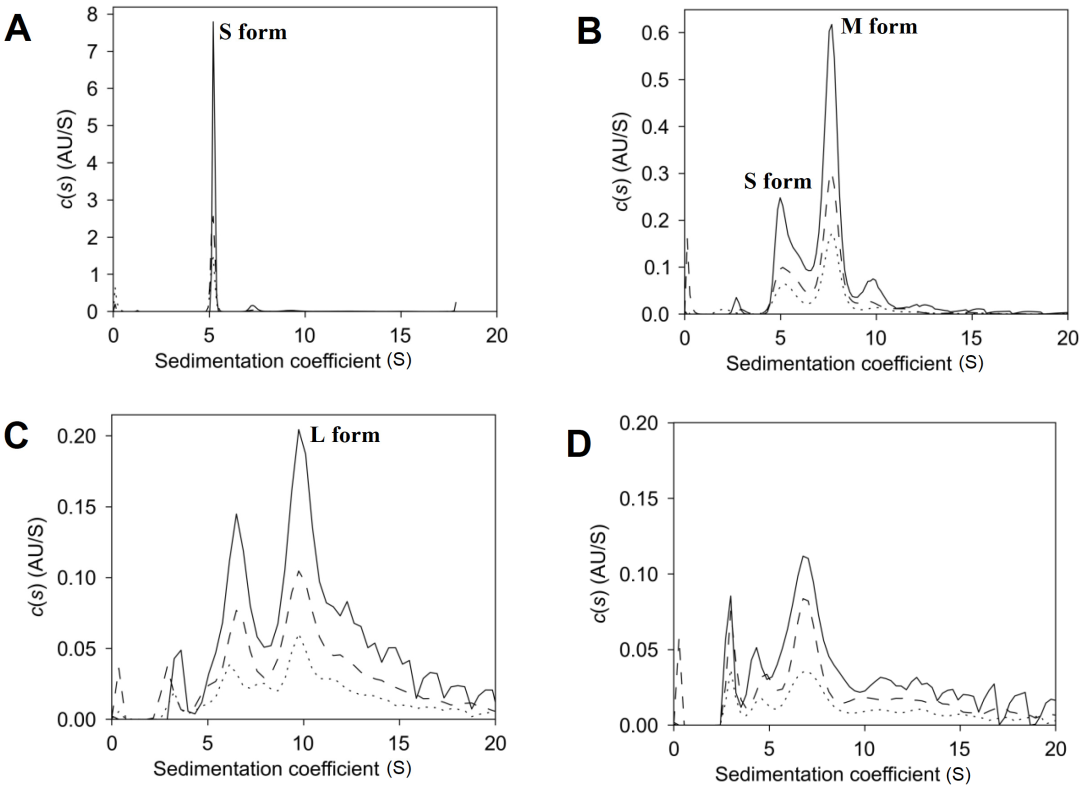


Figure 2-12. Sedimentation coefficient (*s*20,*w*) distribution, *c*(*s*), of the WA20-foldon oligomers. The values of sedimentation coefficient were corrected to 20 °C in pure water. (A) sample S, (B) sample M, (C) sample L, (D) sample H. The sedimentation velocity AUC experiments were performed at protein concentrations of 1.2 mg/mL (black line), 0.6 mg/mL (broken line), and 0.3 mg/mL (dotted line).

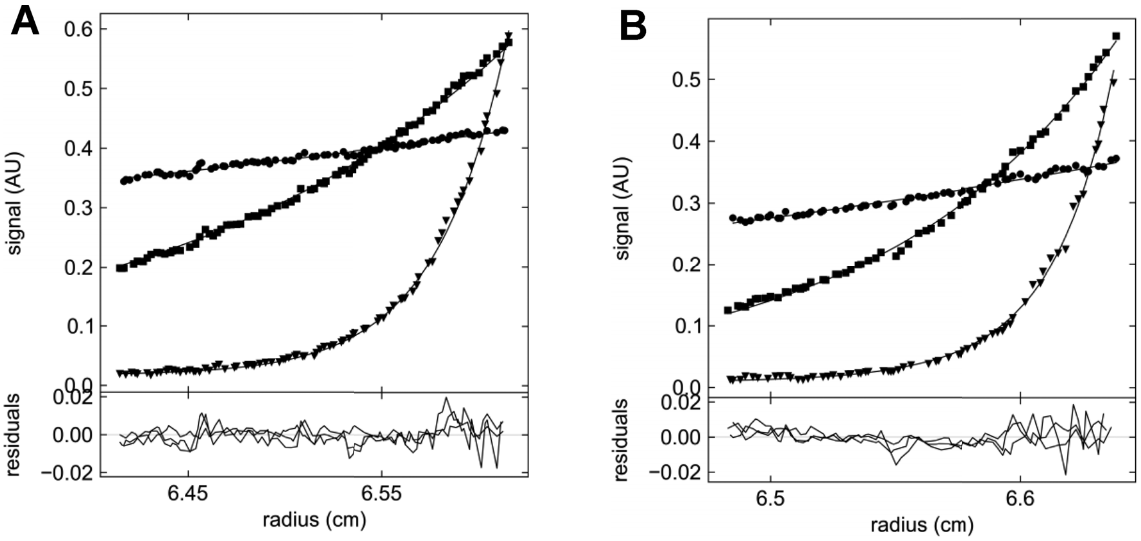


Figure 2-13. Sedimentation equilibrium AUC experiments of the WA20-foldon. Scans from three different rotor speeds (●: 4000 rpm; ■: 9000 rpm; ▼: 20000 rpm) monitored at 280 nm. (Protein concentration: 0.3 mg/mL). (A) Sample S: the lines represent fits to the data of a single species model that yield a calculated protein mass of 96 kDa. (B) Sample M: the lines represent fits to the data of two species model that yield calculated protein mass of 100 and 180 kDa.

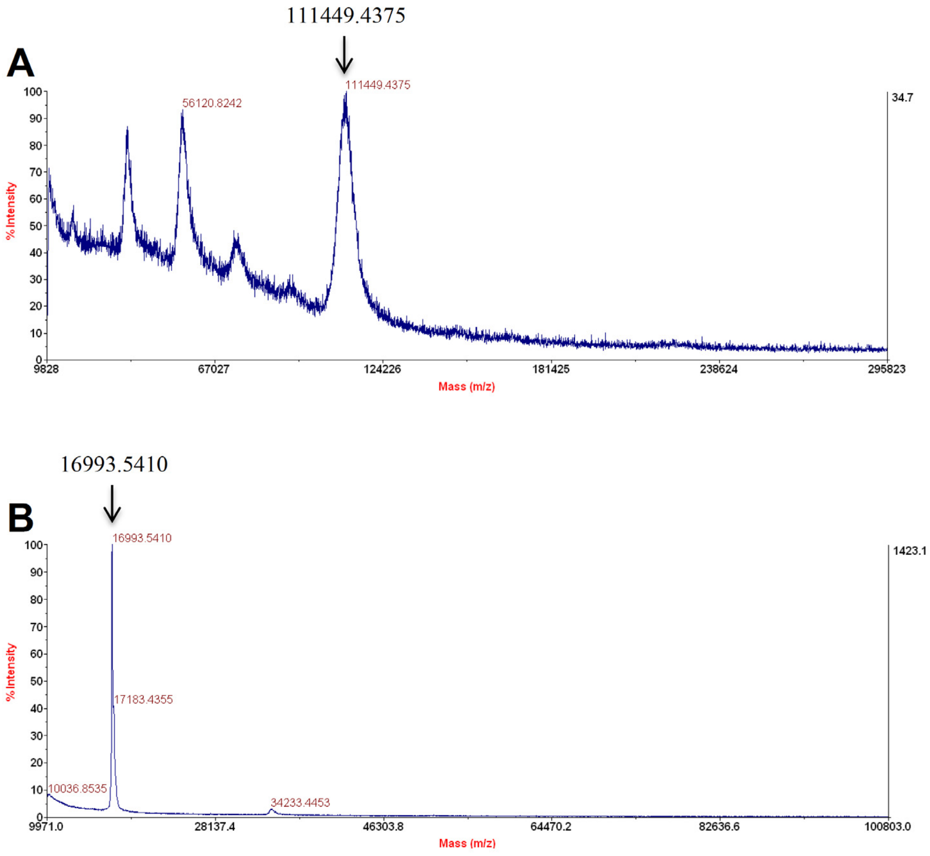


Figure 2-14. (A) MALDI-TOF mass spectrum of the S form of the WA20-foldon cross-linked by glutaraldehyde. For the cross-linking reaction, the S-form protein solution (0.1 mg/mL in 20 mM HEPES buffer (pH 7.5) containing 100 mM NaCl, and 10% glycerol) was treated with 0.24% (final concentration) glutaraldehyde for 4.5 hours at 37 °C. The reaction was terminated by addition of 87 mM (final concentration) Tris-HCl (pH 8.0). The sample was mixed with an equal volume of 6 M guanidine hydrochloride and 4% trifluoroacetic acid, and was desalted on a ZipTip C4 (Merck Millipore). The MALDI-TOF mass spectrum was recorded on an AB SCIEX TOF/TOF 5800 (Applied Biosystems) with sinapinic acid matrix. Because the theoretical molecular mass of a WA20-foldon protomer is 16959.82 Da, the mass peak of m/z = 111449.4375 is assignable to the WA20-foldon hexamer considering mass increase due to chemical modification with glutaraldehyde. (B) MALDI-TOF mass spectrum of the WA20-foldon without cross-linking. The mass peak corresponds to the theoretical molecular mass of a WA20-foldon protomer.

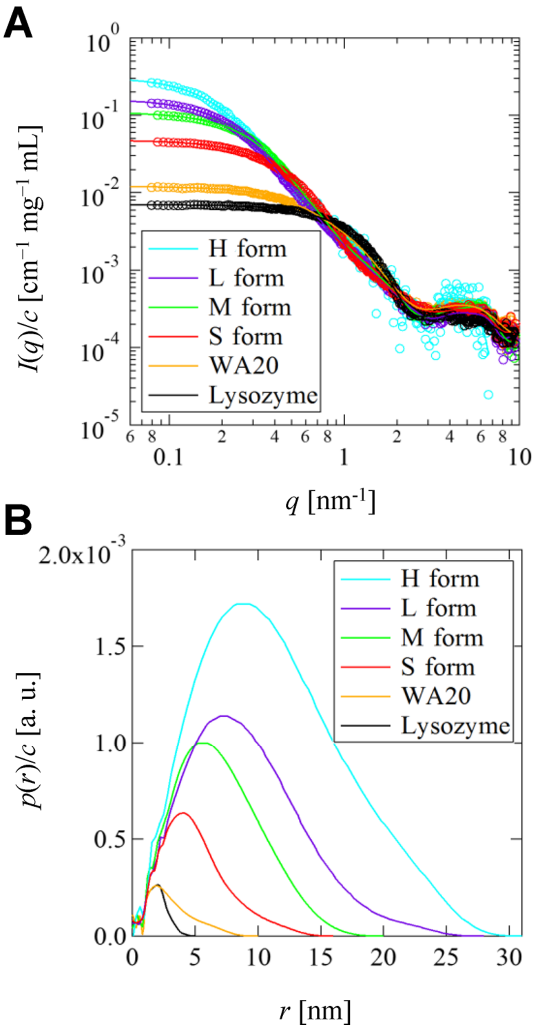
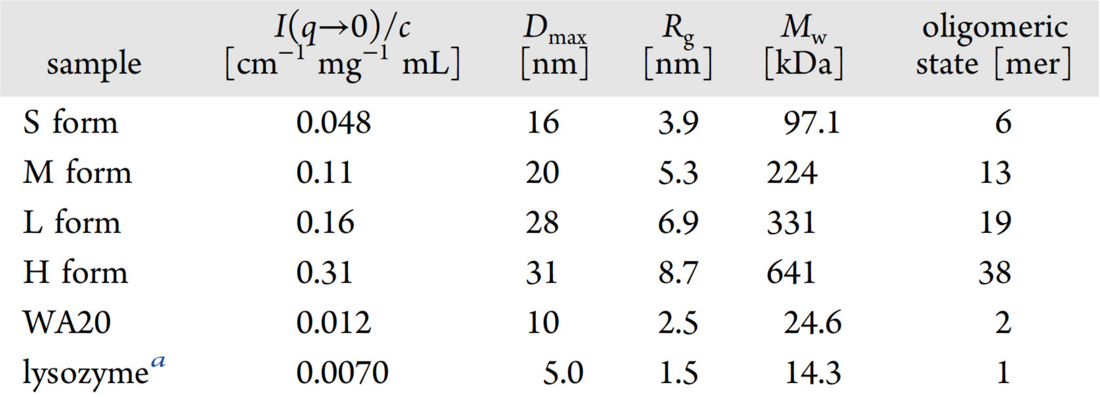


Figure 2-15. SAXS analysis of the WA20-foldon. (A) Concentrationnormalized absolute scattering intensities, $I(q)/c$, of the WA20-foldon oligomers, lysozyme, and WA20. (B) Their real-space information, pair-distance distribution functions normalized by the concentration, $p(r)/c$, as obtained by IFT.

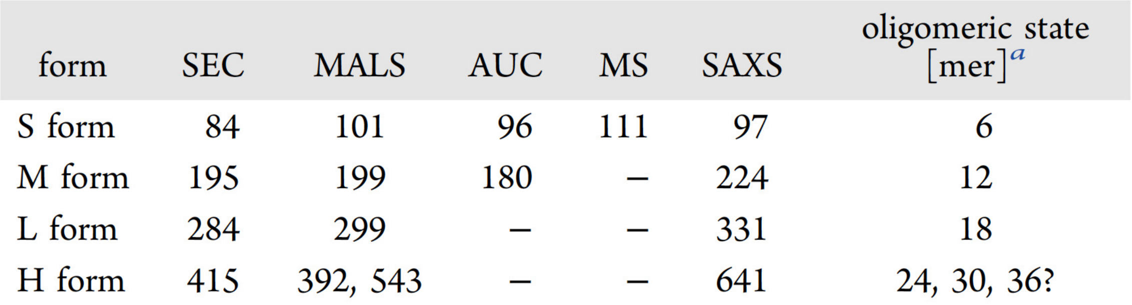
Table 2-5. SAXS Analysis Data and Estimated Molecular Mass of WA20-Foldon Oligomers



sample	$I(q\rightarrow 0)/c$ [cm^{-1} mg^{-1} mL]	D_{max} [nm]	R_g [nm]	M_w [kDa]	oligomeric state [mer]
S form	0.048	16	3.9	97.1	6
M form	0.11	20	5.3	224	13
L form	0.16	28	6.9	331	19
H form	0.31	31	8.7	641	38
WA20	0.012	10	2.5	24.6	2
lysozyme[a]	0.0070	5.0	1.5	14.3	1

[a] Lysozyme is as a molecular mass reference standard.

Table 2-6. Summary of Molecular Mass and Oligomeric State of WA20-Foldon Oligomers



form	SEC	MALS	AUC	MS	SAXS	oligomeric state [mer][a]
S form	84	101	96	111	97	6
M form	195	199	180	–	224	12
L form	284	299	–	–	331	18
H form	415	392, 543	–	–	641	24, 30, 36?

[a] The molecular mass of a WA20-foldon protomer is 17.0 kDa.

Shape Analysis of WA20-Foldon from SAXS Data.

To extract intuitive real-space information from the SAXS data, the pair-distance distribution functions, $p(r)$, reflected by the shapes of the WA20-foldon oligomers were obtained using the IFT technique (Brunner-Popela and Glatter, 1997; Glatter, 1980b; Glatter and Kratky, 1982) (Figure 2-15B). The integral of $p(r)/c$ from $r = 0$ to $r = D_{\mathrm{max}}$ is equal to the extrapolated forward absolute scattering intensity normalized by concentration, $I(q \to 0)/c$, and therefore, it is proportional to the weight-average molecular mass, M_{w}. The shape of $p(r)$ of the S form is characterized by an extended tail in the high-r regime, which is approximated by an ellipsoid (Glatter and Kratky, 1982). In contrast, $p(r)$ of the M form shows a comparatively symmetrical bell-like shape, suggesting a sphere-like structure. The shapes of $p(r)$ of the L and H forms show a higher similarity to that of the M form than that of the S form. These suggest that the L and H forms exist in larger sphere-like structures. However, the high-r residual of $p(r)$ of the L form implies that the L form contains a larger-sized complex as a minor component because of its polydispersity, as suggested by the SEC experiment (Figure 2-9). Because of further purification by repeated SEC, the S and M forms of the WA20-foldon can be considered as practically monodispersed particles. Therefore, further analysis for the S and M forms was performed to obtain more structural insights from the SAXS data utilizing high-resolution structures of the WA20 and foldon domains with geometrical and symmetrical restrictions. The rigid-body model structures of the S form hexamer and the M form dodecamer were iteratively constructed based on the crystal structure of the dimeric WA20 (PDB code 3VJF) (Arai et al., 2012) and the NMR structure of the trimeric foldon domain (PDB code 1RFO) (Guthe et al., 2004) to reproduce the experimental $p(r)$ and $I(q)$ (Figures 2-4−2-6). The models were constructed with a consideration of linking their C and N terminals with two- and threefold symmetries (Figures 2-4 and 2-5). The models show the extended barrel-like structure for the S form (Figure 2-16A) and the distinctive tetrahedron-like structure for the M form (Figure 2-17A). The $p(r)$ and $I(q)$ simulated from the models of the S and M forms closely resemble those obtained from the SAXS experiments (Figures 2-16B and 2-17B and Figure 2-6). Moreover, the hydrodynamic properties of the S and M forms are predicted from the rigid-body model structures using the program HYDROPRO (Ortega et al., 2011). The predicted radii of gyration (R_{g}) are 3.9 nm (S form) and 5.0 nm (M form), and the predicted sedimentation coefficients ($s_{20,\mathrm{w}}$) are 5.4 S (S form) and

7.9 S (M form). These values are highly consistent with the experimental results, 3.9 nm (S form) and 5.3 nm (M form) of R_g from SAXS and 5.2 S (S form) and 7.6 S (M form) of $s_{20,w}$ from AUC, supporting our rigid-body models.

Furthermore, the low-resolution shapes of the S and M forms were reconstructed from the SAXS data using the *ab initio* modeling program DAMMIN (Svergun, 1999). The protein models were composed of small beads (dummy atoms). The shapes were estimated using nonlinear least-squares that fit the experimental SAXS data (Figures 2-16D and 2-17D). I performed calculations several times with and without symmetry constraints (*P*2 and/or *P*3, derived from the dimer and/or trimer domains, respectively), and a majority of the calculations led to similar results. The typical and major results are shown in Figures 2-16C and 2-17C. The S form shows an elongated barrel-like shape (Figure 2-16C). The M form shows a less-elongated, sphere-like shape with four humps, similar to a tetrahedron, with a consideration of a *P*23 symmetry constraint (Figure 2-17C).

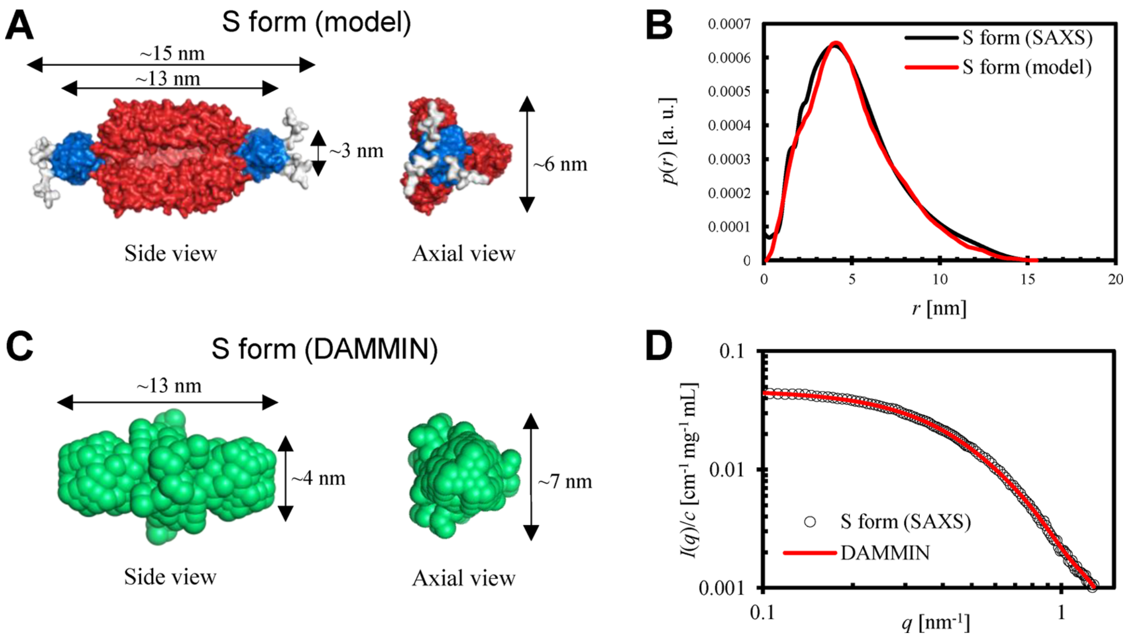


Figure 2-16. Three-dimensional model structures of the S form of the WA20-foldon derived from SAXS analysis. (A) The rigid-body model structure of the S form hexamer of the WA20-foldon. The domains of the WA20 (PDB code 3VJF) (Arai et al., 2012), foldon (PDB code 1RFO) (Guthe et al., 2004), and His$_6$ tag are shown in red, blue, and light gray, respectively. (B) The pair-distance distribution function, $p(r)$, of the S form of the WA20-foldon as obtained by the SAXS experiment (black line) and that simulated from the rigid-body model structure (red line). (C) The dummy atom model shape of the S form of the WA20-foldon reconstructed from the SAXS data using the *ab initio* modeling program DAMMIN with a *P*32 symmetry constraint (Svergun, 1999). (D) The concentration-normalized SAXS intensity, $I(q)/c$, of the S form of the WA20-foldon (black open circle) and that optimized by the DAMMIN procedure (red line).

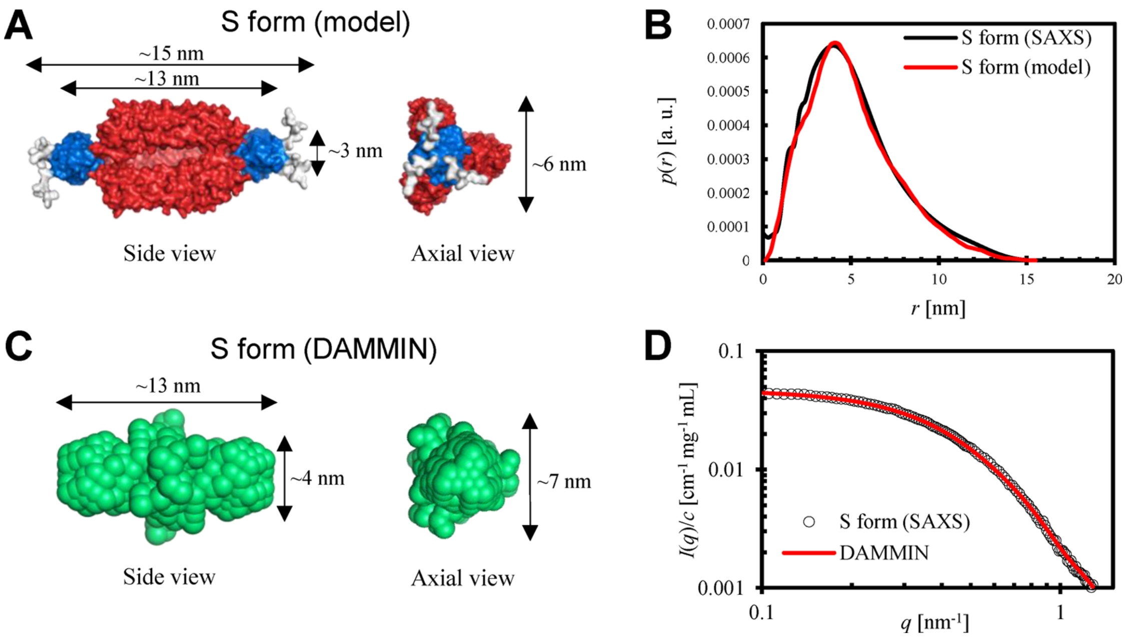


Figure 2-17. Three-dimensional model structures of the M form of the WA20-foldon derived from SAXS analysis. (A) The rigid-body model structure of the M form dodecamer of the WA20-foldon. The domains are shown in the same colors as in Figure 2-16A. (B) The pair-distance distribution function, *p*(*r*), of the M form of the WA20-foldon as obtained by the SAXS experiment (black line) and that simulated from the rigid-body model structure (red line). (C) The dummy atom model shape of the M form of the WA20-foldon reconstructed from the SAXS data using the program DAMMIN with a *P*23 symmetry constraint. (D) The concentration-normalized SAXS intensity, *I*(*q*)/*c*, of the M form of the WA20-foldon (black open circle) and that optimized by the DAMMIN procedure (red line).

Perspectives of PN-Block Approach.

I demonstrated that the WA20-foldon fusion protein as a PN-Block formed several types of self-assembling oligomeric nanostructures. The results illustrated the concept of the "PN-Block approach": various self-assembling nanostructures are created from a few types of simple and fundamental PN-Blocks (Figure 2-1). PN-Blocks using the intermolecularly folded dimeric *de novo* protein (e.g., WA20) have some advantages: (1) the simple, stable, and intertwined rod-like structure of the *de novo* protein makes it easy to use PN-Blocks to design and construct simple and stable frameworks of nanoarchitectures, and (2) the PN-Blocks using the *de novo* protein, based on the simple binary patterning, have great potential for redesigning functional nanostructures using our functional binary-patterned *de novo* protein library (Fisher et al., 2011; Patel et al., 2009). ("PN" in PN-Block also has a different meaning from the polar and nonpolar abbreviations used in the binary code strategy for protein design.) Moreover, the PN-Block approach can be further enhanced by adding cofactors and/or synthetic ligands such as metal-directed protein self-assemblies (Brodin et al., 2012; Salgado et al., 2007) and the protein encapsulation in synthetic self-assembled coordination cages (Fujita et al., 2012) and by using computational methods for protein design such as Rosetta software suite for macromolecular modeling (Baker, 2014; Leaver-Fay et al., 2013).

Chapter 3 Self-Assembling Supramolecular Nanostructures Created by *de Novo* Extender Protein Nanobuilding Blocks

3.1 Introduction

Living organisms are maintained by various self-assembling biomolecules including proteins, nucleic acids, sugars, and lipids. The chemical reconstitution of living matter is one of the ultimate goals of chemistry and synthetic biology. Rational design of artificial biomacromolecules that self-assemble into supramolecular complexes is an important step toward achieving the goal.

Proteins are the most versatile biomacromolecules performing the complex and functional tasks in the living organisms. Protein functions are essentially determined by three-dimensional (3D) structures of proteins. There are four hierarchical levels in protein structures. The amino acid sequence of a protein's polypeptide chain is called its primary structure. The secondary structure can take the local regular form either of α-helices or of β-strands. In globular form of proteins, elements of α-helices and/or β-sheets as well as loops are folded into a tertiary structure. Furthermore, many proteins are formed by self-assembling the folded chains of more than one polypeptide; this constitutes the quaternary structure of a protein. The complex and refined quaternary structures create versatile functionalities of proteins.

The design of *de novo* proteins is substantially very complicated to explore enormous amino-acid sequence space because the contribution of many cooperative and long-range interactions causes a significant gap between the primary structure and the tertiary and quaternary structure. Research on *de novo* protein design has progressed toward the construction of novel proteins emanated mainly from three approaches: (1) rational and computational design (Dahiyat and Mayo, 1997; Huang et al., 2016; Koga et al., 2012; Kuhlman et al., 2003), (2) combinatorial methods (Keefe and Szostak, 2001; Urvoas et al., 2012), and (3) semirational approaches, including elements of both rational design and combinatorial methods (Hecht et al., 2004; Kamtekar et al., 1993; Urvoas et al., 2012). As a semirational approach, the binary code strategy has been developed to produce primary structure libraries for *de novo* protein tertiary structures using secondary structure units designed by the binary patterning of polar and non-polar residues (Hecht et al., 2004; Kamtekar et al., 1993). Alpha-helix and β-sheet *de novo* proteins have been successfully created (Hecht et al., 2004; Kamtekar et al., 1993). From a third-generation library of *de novo* 4-helix bundle proteins designed by binary patterning (Bradley et al., 2005; Patel et al., 2009), several *de novo* proteins have actually functions *in vitro* (Cherny et al., 2012; Patel et al., 2009; Patel and Hecht, 2012) and *in vivo* (Digianantonio and Hecht, 2016; Fisher et al., 2011; Hoegler and Hecht, 2016; Smith et al., 2015). Recently, I described the crystal structure of the *de novo* protein WA20 (Arai et al., 2012), a stable and functional *de novo* protein from the third-generation library (Patel et al., 2009). WA20 has an unusual dimeric structure with an intermolecularly folded (domain-swapped) 4-helix bundle. Each WA20 monomer ("nunchaku"-like structure), which comprises two long α-helices, inter-twines with the helices of another monomer. The structure of WA20 is stable (melting temperature, T_m = ~70 °C) and forms a simple rod-like shape (Arai et al., 2012). The stable, simple, and unusual intermolecularly folded structure of the *de novo* protein WA20 raises the possibility of application to basic framework tools in nanotechnology and synthetic biology.

In recent years, several approaches to design artificial self-assembling protein complexes have been developed: 3D domain-swapped oligomers (Hirota et al., 2010; Miyamoto et al., 2015; Ogihara et al., 2001); Nanostructures constructed from fusion proteins designed by symmetric self-assembly (Bai et al., 2013; Lai et al., 2012a; Lai et al., 2014; Padilla et al., 2001; Sinclair et al., 2011); Self-assembling nanostructures constructed from designed coiled-coil peptide modules (Boyle et al., 2012; Fletcher et al., 2013; Pandya et al., 2000; Papapostolou et al., 2007; Sciore et al., 2016; Sharp et al., 2012); Metal-directed self-assembling protein complexes (Brodin et al., 2012; Salgado et al., 2007; Sontz et al., 2015); Computationally designed self-assembling protein nanocages with atomic level accuracy (Bale et al., 2016; Hsia et al., 2016; Huang et al., 2016; King et al., 2014; King et al., 2012) and computationally designed β-propeller proteins (Voet et al., 2014); Other approaches (see references (Bozic et al., 2013; King and Lai, 2013; Lai et al., 2012b; Luo et al., 2016; Radford et al., 2011; Woolfson et al., 2012; Yeates et al., 2016)).

More recently, I have designed and created a "protein nanobuilding block (PN-Block)": an artificial protein that can form supramolecular complexes by self-assembly as bulding blocks, in nanometer scale. A polyhedral PN-Block, called WA20-foldon (Kobayashi et al., 2015), was constructed by fusing an intermolecularly folded dimeric *de novo* protein WA20 (Arai et al., 2012) and a trimeric foldon domain of T4 phage fibritin (Guthe et al., 2004). The WA20-foldon formed several distinctive types of self-assembling nanoarchitectures in multiples of 6-mer (6-, 12-, 18-, 24-, and 30-mer) because of the combination of dimer and trimer. The basic concept of the "PN-Block Strategy": various self-assembling nanostructures are created from a few types of simple and fundamental PN-Blocks. PN-Blocks using the intermolecularly folded dimeric *de novo* protein (e.g., WA20) as a component have some advantages: (1) the simple, stable, and intertwined rod-like structure of the *de novo* protein makes it easy to use PN-Blocks to design and construct simple and stable frameworks of nano-architectures (Kobayashi et al., 2015), and (2) the PN-Blocks using the *de novo* protein, based on the simple binary patterning, have great potential for redesigning functional protein nanostructures using binary-patterned *de novo* protein libraries functionable *in vitro* (Cherny et al., 2012; Patel et al., 2009; Patel and Hecht, 2012) and *in vivo* (Digianantonio and Hecht, 2016; Fisher et al., 2011; Hoegler and Hecht, 2016; Smith et al., 2015). Further designing and creating new types of PN-Blocks and reconstructing various PN-Blocks are essential steps to developing the PN-Block strategy.

In this study, I designed and created *de novo* extender protein nanobuilding blocks (ePN-Blocks), constructed by fusing tandemly two *de novo* WA20 proteins with various linkers (Arai et al., 2001; Arai et al., 2004), as a new series of PN-Blocks to construct self-assembling extended or cyclized chain nanostructures (Figure 3-1). Moreover, I successfully reconstructed quaternary structural heteromeric complexes from extender and stopper PN-Blocks by denaturation and refolding. Furthermore, I demonstrate that the complexes can further self-assemble into supramolecular nanostructures on mica surface as a "supra-quaternary structure," expanding further possibilities of the PN-block strategy.

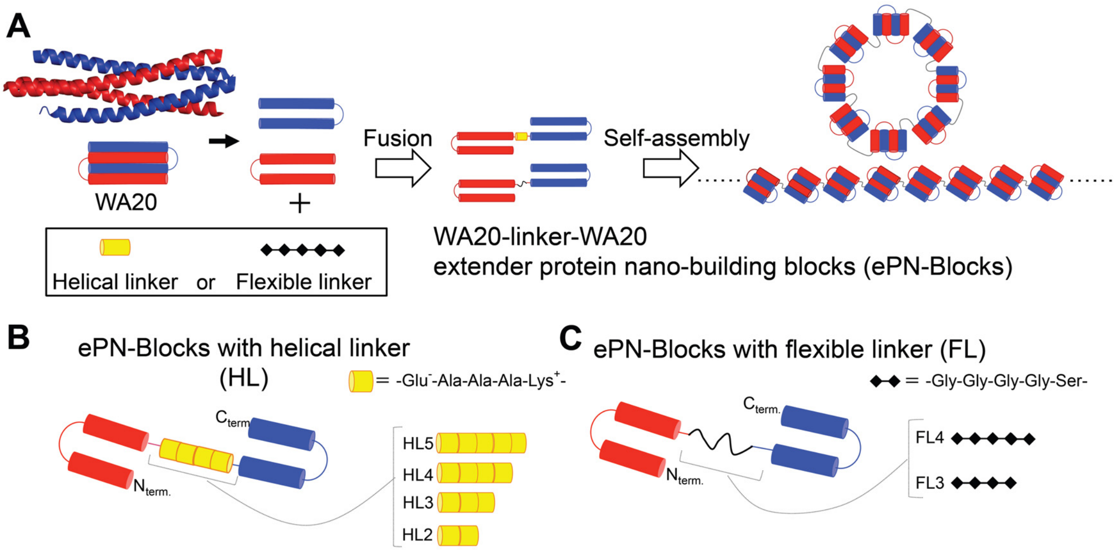

A
WA20
+
Fusion
Self-assembly
WA20-linker-WA20
extender protein nano-building blocks (ePN-Blocks)
Helical linker or Flexible linker
B
ePN-Blocks with helical linker (HL)
= -Glu$^-$-Ala-Ala-Ala-Lys$^+$-
C$_{term}$
N$_{term.}$
HL5
HL4
HL3
HL2
C
ePN-Blocks with flexible linker (FL)
= -Gly-Gly-Gly-Gly-Ser-
C$_{term.}$
N$_{term.}$
FL4
FL3

Figure 3-1. Schematics of extender protein nano-building blocks (ePN-blocks). (A) Construction and assemblies of the ePN-blocks. Ribbon representation and schematics of the intermolecularly folded dimeric *de novo* protein WA20 (PDB code 3VJF) (Arai et al., 2012) are shown in red and blue. Helical and flexible linkers (Arai et al., 2001) are shown in yellow rod and black line, respectively. (B) Detailed construct of ePN-Blocks with helical linkers. (C) Detailed construct of ePN-Blocks with flexible linkers.

3.2 Materials and methods

Construction of the Expression Plasmids.

The DNA fragment encoding the *de novo* protein WA20 was prepared from plasmid pET3a-WA20 (Arai et al., 2012) by polymerase chain reaction (PCR) using KOD-Plus-Neo DNA polymerase (Toyobo, Osaka, Japan) and primers, T7 promoter primer and WA20RV_*Hind*III (Table 3-1). The amplified fragment was digested by *Nde*I and *Hind*III and cloned into pET32/EBFP-HL5-EGFP (Arai et al., 2001), between the *Nde*I and *Hind*III sites to construct the plasmid pET-WA20-HL5-GFP (i.e., the Trx tag and the EBFP gene was removed and replaced with the WA20 gene). Another DNA fragment encoding the WA20 was prepared from plasmid pET3a-WA20 by PCR with primers, WA20FW_*Not*I and WA20RV_*Xho*I (Table 3-1). The amplified fragment was digested by *Not*I and *Xho*I and cloned into pET-WA20-HL5-EGFP, between the *Not*I and *Xho*I sites to give the expression plasmid pET-WA20-HL5-WA20 for an extender PN-Block (ePN-Block). The DNA fragments encoding the other linker genes (HL2, HL3, HL4, FL3 and FL4) were prepared by digestion of plasmids pET32/EBFP-linker-EGFP49 with *Hind*III and *Not*I and cloned into pET-WA20-HL5-WA20 between the *Hind*III and *Not*I sites to give the expression plasmid pET-WA20-HL2-WA20, pET-WA20-HL3-WA20, pET-WA20-HL4-WA20, pET-WA20-FL3-WA20, and pET-WA20-FL4-WA20. The amino acid sequences of the ePN-Block proteins (WA20-HL2-WA20, WA20-HL3-WA20, WA20-HL4-WA20, WA20-HL5-WA20, WA20-FL3-WA20, and WA20-FL4-WA20) are shown in Figure 3-2.

Table 3-1. Oligonucleotide Primer Sequences Used in This Study

Primer name	Sequence (5’→3’)
T7 promoter primer	TAATACGACTCACTATAGGG
WA20RV_HindIII	GCGGCAAAAGCTTGCGATGTACAAGGTGGTGGAAGT
WA20FW_NotI	GCTACGGGGCGGCCGCAATGTATGGCAAGTTGAACAAGCTG
WA20RV_XhoI	GCAGCCCCTCGAGCAGCCGGATCCTATTAGCG

```
>extender PN-Block (HL2): WA20-HL2-WA20; 220aa; 26.50 kDa; pI 6.93
MYGKLNKLVEHIKELLQQLNKNWHRHQGNLHDMNQQMEQLFQEFQHFMQGNQDDGKLQNMIHEMQQFMNQVDNHLQSE
SDTVHHFHNKLQELMNNFHHLVHRKLAEAAAKEAAAKAAAMYGKLNKLVEHIKELLQQLNKNWHRHQGNLHDMNQQME
QLFQEFQHFMQGNQDDGKLQNMIHEMQQFMNQVDNHLQSESDTVHHFHNKLQELMNNFHHLVHR


>extender PN-Block (HL3): WA20-HL3-WA20; 225aa; 26.97 kDa; pI 6.93
MYGKLNKLVEHIKELLQQLNKNWHRHQGNLHDMNQQMEQLFQEFQHFMQGNQDDGKLQNMIHEMQQFMNQVDNHLQSE
SDTVHHFHNKLQELMNNFHHLVHRKLAEAAAKEAAAKEAAAKAAAMYGKLNKLVEHIKELLQQLNKNWHRHQGNLHDM
NQQMEQLFQEFQHFMQGNQDDGKLQNMIHEMQQFMNQVDNHLQSESDTVHHFHNKLQELMNNFHHLVHR


>extender PN-Block (HL4): WA20-HL4-WA20; 230aa; 27.43 kDa; pI 6.93
MYGKLNKLVEHIKELLQQLNKNWHRHQGNLHDMNQQMEQLFQEFQHFMQGNQDDGKLQNMIHEMQQFMNQVDNHLQSE
SDTVHHFHNKLQELMNNFHHLVHRKLAEAAAKEAAAKEAAAKEAAAKAAAMYGKLNKLVEHIKELLQQLNKNWHRHQG
NLHDMNQQMEQLFQEFQHFMQGNQDDGKLQNMIHEMQQFMNQVDNHLQSESDTVHHFHNKLQELMNNFHHLVHR


>extender PN-Block (HL5): WA20-HL5-WA20; 236aa; 28.01 kDa; pI 6.93
MYGKLNKLVEHIKELLQQLNKNWHRHQGNLHDMNQQMEQLFQEFQHFMQGNQDDGKLQNMIHEMQQFMNQVDNHLQSE
SDTVHHFHNKLQELMNNFHHLVHRKLAEAAAKEAAAKEAAAKEAAAKEAAAKAAAMYGKLNKLVEHIKELLQQLNKNW
HRHQGNLHDMNQQMEQLFQEFQHFMQGNQDDGKLQNMIHEMQQFMNQVDNHLQSESDTVHHFHNKLQELMNNFHHLVH
R


>extender PN-Block (FL3): WA20-FL3-WA20; 224aa; 26.43 kDa; pI 6.93
MYGKLNKLVEHIKELLQQLNKNWHRHQGNLHDMNQQMEQLFQEFQHFMQGNQDDGKLQNMIHEMQQFMNQVDNHLQSE
SDTVHHFHNKLQELMNNFHHLVHRKLGGGGSGGGGSGGGGSAAAMYGKLNKLVEHIKELLQQLNKNWHRHQGNLHDMN
QQMEQLFQEFQHFMQGNQDDGKLQNMIHEMQQFMNQVDNHLQSESDTVHHFHNKLQELMNNFHHLVHR


>extender PN-Block (FL4): WA20-FL4-WA20; 230aa; 26.83 kDa; pI 6.93
MYGKLNKLVEHIKELLQQLNKNWHRHQGNLHDMNQQMEQLFQEFQHFMQGNQDDGKLQNMIHEMQQFMNQVDNHLQSE
SDTVHHFHNKLQELMNNFHHLVHRKLSGGGGSGGGGSGGGGSGGGGSAAAMYGKLNKLVEHIKELLQQLNKNWHRHQG
NLHDMNQQMEQLFQEFQHFMQGNQDDGKLQNMIHEMQQFMNQVDNHLQSESDTVHHFHNKLQELMNNFHHLVHR

>stopper PN-Block (WA20); 102aa; 12.55 kDa; pI 6.93
MYGKLNKLVEHIKELLQQLNKNWHRHQGNLHDMNQQMEQLFQEFQHFMQGNQDDGKLQNMIHEMQQFMNQVDNHLQSE
SDTVHHFHNKLQELMNNFHHLVHR
```

Figure 3-2. Amino acid sequences of protein nanobuilding blocks (PN-Blocks) used in this study.

Protein Expression and Purification.

All ePN-Block proteins were expressed in *E. coli* BL21 Star(DE3) (Invitrogen, Carlsbad, CA) harboring pET-WA20-Linker-WA20 using 2 L of LB broth, Lennox (Nacalai Tesque, Kyoto, Japan) with 50 µg/mL ampicillin sodium salt at 37 °C. The expression was induced with 0.2 mM β-D-1-thiogalactopyranoside at OD_{600} (optical density at 600 nm) = ~0.8, and cells were further cultured for 3–4 h at 37 °C. The protein was extracted from the harvested cells by sonication in a lysis buffer (50 mM sodium phosphate buffer (pH 7.0) containing 300 mM NaCl, 10% glycerol). The protein was purified by immobilized metal ion affinity chromatography (IMAC) with TALON metal affinity resin (Clontech, Takara Bio, Mountain View, CA) and according to the manufacturer's protocols (equilibration/wash buffer: 50 mM sodium phosphate buffer (pH 7.0) containing 300 mM NaCl; elution buffer: 50 mM sodium phosphate buffer (pH 7.0) containing 300 mM NaCl, 10% glycerol, and 250 mM imidazole). Even without a His-tag, domains of the WA20 proteins ePN-Block proteins can bind to a TALON metal affinity resin, probably because of the relatively high percentage (12.7%) of histidine residues in the amino acid sequence of the WA20 domains (Arai et al., 2012). Protein expression and purification of WA20, the stopper PN-Block (sPN-Block), were performed as previously described (Arai et al., 2012).

Sodium dodecyl sulfate polyacrylamide gel electro-phoresis (SDS-PAGE) analysis was performed according to the standard Laemmli procedure. The stacking gel (4.5% polyacrylamide gel, 125 mM Tris–HCl (pH6.8), and 0.1% SDS), separatory gel (17.5% polyacrylamide gel, 125 mM Tris–HCl (pH8.8), and 0.1% SDS), and the running buffer (25 mM Tris, 192 mM glycine, and 0.1% SDS) were used. The running buffer was. For native PAGE analysis the stacking gel, separatory gel (5% or 7.5% polyacryla-mide gel), and the running buffer were prepared in the same way as the SDS-PAGE, except that no SDS was used. Proteins in the gels were stained with Coomassie brilliant blue.

Denaturation, Refolding, and Further Purificarion of PN-Block Proteins.

The ePN-Block (HL4 or FL4) protein mixed with the stopper PN-Block (sPN-Block) protein (WA20) was denatured by 6 M guanidine hydrochloride (GdnHCl) for 3 h at 25 °C in 20 mM HEPES buffer (pH 7.5) containing 100 mM NaCl, 10% glycerol. For refolding, the denatured proteins were dialyzed for ~4 h three times against 20 mM HEPES buffer (pH 7.5) containing 100 mM NaCl, 10% glycerol, 200 mM L-arginine hydrochloride (ArgHCl) using BioTech oscillatory microdialysis system

(BM Equipment, Tokyo, Japan). The concentrated samples of the refolded extender and stopper PN-Block complexes (esPN-Blocks) of ePN-Block and sPN-Block were concentrated by Amicon Ultra centrifugal filters (Merck Millipore) and were separated and purified by gel filtration chromatography (20 mM HEPES buffer (pH 7.5) containing, 100 mM NaCl, 10% glycerol, and 200 mM ArgHCl) on a Superdex 200 increase 10/300 GL column (GE healthcare, Little Chalfont Buckinghamshire, UK). All ePN-Block proteins were further purified by size exclusion chromatography (SEC) (20 mM HEPES buffer (pH 7.5) containing 100 mM NaCl, 200 mM ArgHCl, and 10% glycerol) with a Superdex 200 Increase 10/300 GL columns (GE healthcare).

SEC Multi-Angle Light Scattering (SEC-MALS).

SEC-MALS experiments were performed using a 1260 Infinity HPLC system Agilent Technologies, Santa Clara, CA) equipped with a Superdex 200 Increase 10/300 GL column, which was connected in line with a miniDAWN TREOS multiangle static light-scattering detector (Wyatt Technology, Santa Barbara, CA). The data were collected in phosphate buffered saline (PBS, (pH 7.4): 1 mM KH_2PO_4, 3 mM Na_2HPO_4, and 155 mM NaCl) at 20 °C and analyzed using ASTRA 6 software (Wyatt Technology). The d*n*/d*c* value (0.185 mL/g) was generally used for proteins, and the extinction coefficients (0.507 mL mg^{-1} cm^{-1} and 0.519 mL mg^{-1} cm $^{-1}$ for the esPN-Block (HL4) and esPN-Block (FL4), respectively) were calculated from the amino acid sequences.

Small-Angle X-ray Scattering (SAXS).

SAXS measurements were performed on the several fractions of esPN-Block (HL4) complex and esPN-Block (FL4) complex separated by SEC, chicken egg white lysozyme (Wako Pure Chemical Industries, Osaka, Japan), and WA20 (Arai et al., 2012) in HEPES buffer (pH 7.5) containing 100 mM NaCl, 200 mM ArgHCl, 10% glycerol at 20°C (Table 3-2). The measurements were performed using synchrotron radiation (λ = 0.1488 nm) with an instrument for SAXS installed at the BL-10C beamline (Igarashi et al., 2011) of Photon Factory (KEK, Tsukuba, Japan) using PILATUS3 2M detector (Dectris, Baden, Switzerland).

Generally, the scattering intensity for a colloidal dispersion is given by the product of the form factor, $P(q)$, and structure factor, $S(q)$,

$I(q) = n\, P(q)\, S(q)$,

where n is the number density of the particle. If interparticle interactions such as the excluded volume effect and electrostatic interaction can be neglected (i.e., $S(q) = 1$), the scattering intensity is proportional to the form factor. Our experimental condition can be regarded as a situation in which the structure factor is almost unity, i.e., $I(q) \approx n\,P(q)$, because of a low protein concentration and high salt concentration of the solvent. The form factor is given by the Fourier transformation of the pair-distance distribution function, $p(r)$, which describes the size and shape of the particle,

$$P(q) = 4\pi \int_0^{D_{\max}} p(r) \frac{\sin qr}{qr} \mathrm{d}r$$

where $D_{\max}$ is the maximum intraparticle distance. To obtain $p(r)$ of the particle using a virtually model-free routine, the indirect Fourier transformation (IFT) technique was used (Brunner-Popela and Glatter, 1997; Glatter, 1980b; Glatter and Kratky, 1982). The forward absolute scattering intensity, $I(q\rightarrow 0)$, was extrapolated from the data. The radius of gyration, R_g, was estimated by the Guinier approximation (Glatter and Kratky, 1982).

Assuming that these proteins have practically identical scattering length densities and specific volumes and that the structure factor $S(q) \approx 1$ for dilute samples, the forward-scattering intensity normalized by protein concentration, $I(q\rightarrow 0)/c$, is proportional to the weight-average molecular mass (M_w). Lysozyme (M_w = 14.3 kDa) was used as a molecular mass reference standard.

Table 3-2. Experimental Samples and Protein Concentration for SAXS Meas urements

Samples	Fractions (Figures S3 or S5)	Protein concentration [mg/mL]
esPN-Block (HL4) complex sample I	8–11	7.0
esPN-Block (HL4) complex sample II	16–18	5.7
esPN-Block (HL4) complex sample III	21–24	5.7
esPN-Block (HL4) complex sample IV	27–31	5.7
esPN-Block (HL4) complex sample V	32–34	2.6
esPN-Block (FL4) complex sample I	9–12	5.8
esPN-Block (FL4) complex sample II	15–18	6.1
esPN-Block (FL4) complex sample III	22–25	6.6
esPN-Block (FL4) complex sample IV	28–32	6.6
esPN-Block (FL4) complex sample V	33–34	3.5
Hen egg lysozyme		7.5
sPN-Block (WA20)		6.1

Rigid-Body Modeling.

The rigid-body models of esPN-Block oligomeric structures were constructed using the program COOT (Emsley et al., 2010) based on the crystal structure of the *de novo* protein WA20 dimer (protein data bank (PDB) code, 3VJF) (Arai et al., 2012) with a consideration of their N- and C-terminal directions and 2-fold symmetries. The rigid-body models were manually and iteratively refined to minimize differences in the $p(r)$ calculated from the models and those obtained from SAXS experiments.

Atomic Force Microscopy Imaging.

Solution of The esPN-Block complex samples were diluted with PBS buffer to a concentration of ~10 μg /mL. 5 mM $NiCl_2$ solution was deposited onto a freshly cleaved mica surface (φ12 mm). Five minutes later, the surface was gently rinsed 10 times with super pure water and dried out by blowing with nitrogen gas. The esPN-Block complex samples were diluted with PBS buffer to a concentration of ~10 μg/mL. For frequency modulation atomic force microscopy (FM-AFM) imaging, the esPN-Block complex sample solution (100 μL) was deposited onto the nickel ion-coated mica surface. The sample was incubated at room temperature (25°C) for 5 min for self-assembling nanostructures of esPN-Block(HL4) or esPN-Block(FL4) complex on mica surface, and rinsed with PBS buffer. AFM measurements was performed using an in-house-built, ultralow-noise FM atomic force microscope (Fukuma et al., 2005) combined with a commercially available AFM controller (Nanonis RC-4, SPECS Zurich GmbH, Zurich, Switzerland). All AFM experiments were performed at room temperature (25°C) in PBS buffer solution. A commercially available silicon cantilever (PPP-NCH; Nanoworld, Headquarters, Switzerland) with a nominal spring constant of 42 Nm^{-1} and a resonance frequency of 150 kHz in liquid was used. A phase-locked loop circuit (Nanonis OC-4; SPECS) was used to detect the frequency shift and oscillate the cantilever at its resonance frequency with a constant amplitude.

3.3 Results and Discussion

Design of *de Novo* Extender PN-Blocks with Various Linkers to Construct Self-Assembling Chain-like Nanostructures.

As shown (Figure 3-1A), to construct self-assembling chain-like extended nanostructures, I designed and constructed *de novo* extender protein nanobuilding blocks (ePN-Blocks), fusion proteins of tandem *de novo* WA20 proteins with various linkers which have different types and length. The linkers between two WA20 domains adopted a helical linker (HL), $(EAAAK)_n$ (n=2–5) (Figure 3-11B) or a flexible linker (FL), $(GGGGS)_n$ (n=3, 4) (Figure 3-1C). The helical linkers derived from stably helix-forming *de novo* designed peptides (Marqusee and Baldwin, 1987) are able to separate two domains and control the distance between two domains (Arai et al., 2001; Arai et al., 2004). Since WA20 forms stably an intermolecularly folded dimeric structure (Arai et al., 2012), the ePN-Blocks are expected to self-assemble into cyclized or extended chain-like oligomers (Figure 3-1). The ePN-Blocks with different type and length linkers can affect nanostructures of the complexes.

In addition, it is noteworthy that the second series of the PN-Blocks are fully "*de novo*" proteins, which have no sequences derived from any natural proteins. The *de novo* ePN-Block proteins are designed by tandemly linking two *de novo* WA20 proteins (Arai et al., 2012) created by the binary code strategy (Hecht et al., 2004; Kamtekar et al., 1993), with the artificial peptide linker sequences (Arai et al., 2001).

Self-Assembling Oligomers of ePN-Blocks produced in *E. coli*.

The ePN-Block proteins with various linkers were expressed in *E. coli*. The cells were disrupted by sonication, and soluble and insoluble fractions were prepared by centrifugation. Figure 3-3A shows SDS-PAGE analysis of the fractions. The ePN-Block proteins with the long helical linkers (HL4, HL5) and flexible linkers (FL3, FL4) were expressed mainly in soluble fractions in *E. coli*. However, the ePN-Block proteins with the short helical linkers (HL2, HL3) were expressed mainly in insoluble fractions. The ePN-Block proteins were purified from the soluble fractions by IMAC. SDS-PAGE of the purified samples shows almost a single band (Figure 3-3B). However, native PAGE of the same samples shows the ePN-Block proteins with the long helical linkers (HL4, HL5) and the flexible linkers (FL3, FL4) were migrated as ladder bands in native PAGE. In contrast, the ePN-Blocks with the short helical linkers (HL2, HL3) were migrated as a few bands (Figure 3-3C). These results suggest that the ePN-Blocks with the long helical linkers (HL4 and HL5) and the flexible linkers (FL3 and FL4) stably formed several homooligomeric states in the soluble fraction in *E. coli*. In contrast, the PN-Blocks with short helical linkers (HL2 and HL3) mainly precipitated in the insoluble fraction and formed only a few limited oligomeric states in the soluble fraction, possibly due to characteristics of the short and rigid linkers (HL2 and HL3).

Thereafter, I performed further experiments using soluble samples, especially ePN-Blocks with HL4 and FL4 for comparison between typical helical and flexible linkers at the same length, because insoluble samples have difficulty in general analysis techniques for protein solution.

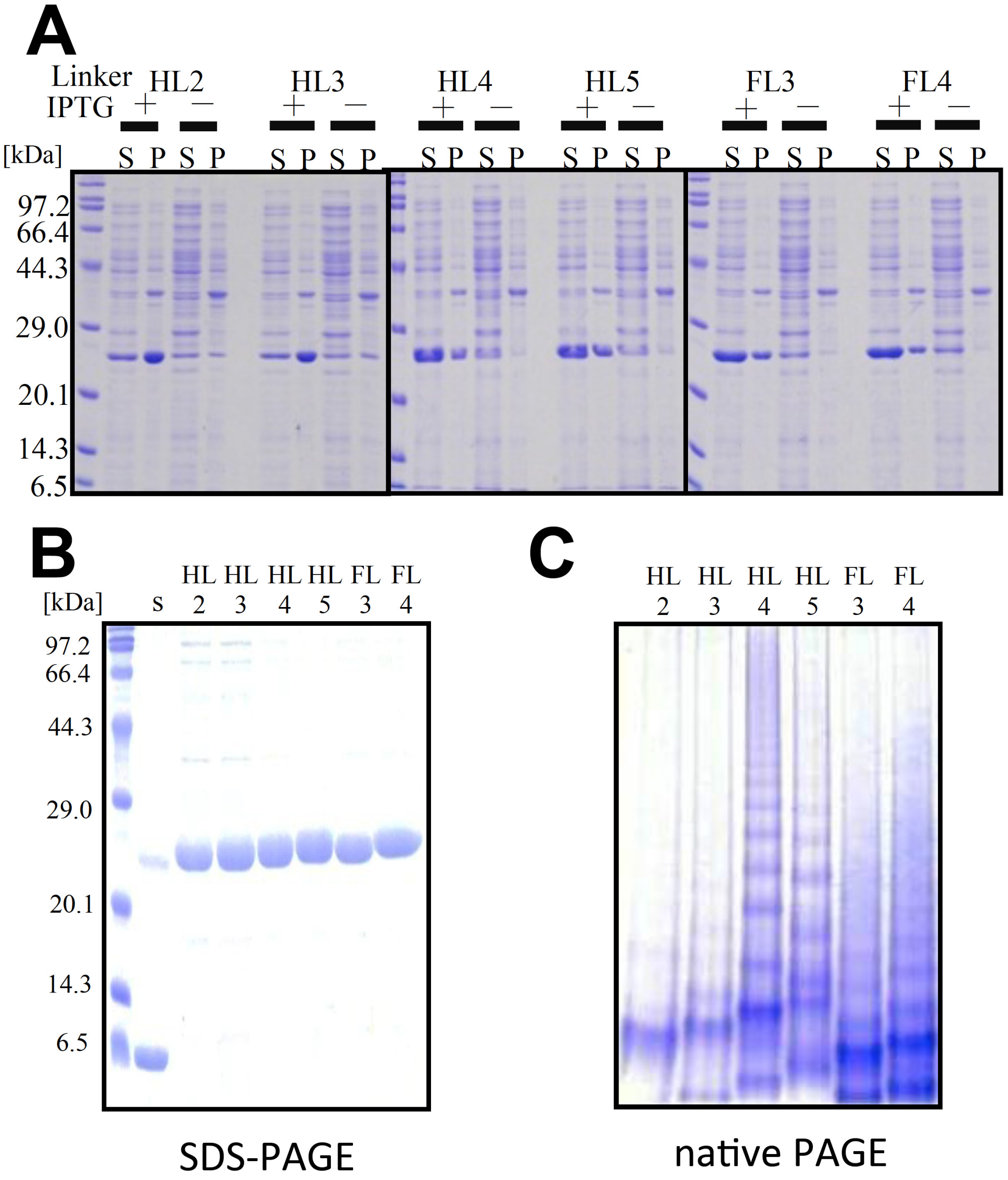


Figure 3-3. Polyacrylamide gel electrophoresis (PAGE) of extender PN-Blocks (ePN-Blocks) with various linkers. (A) SDS-PAGE (17.5% polyacrylamide gel) of ePN-Blocks expressed in *E. coli*. S: supernatant (soluble fraction); P: pellet (insoluble fraction); +: addition of 0.2 mM IPTG; -: no addition of IPTG. (B) SDS-PAGE (17.5% polyacrylamide gel) and (C) native PAGE (5.0% polyacrylamide gel) of the ePN-Blocks after IMAC purification. s: stopper PN-Block (WA20). Proteins were stained with Coomassie brilliant blue. The protein molecular weight marker (broad) (Takara Bio, Otsu, Japan) was used for SDS-PAGE.

Reconstruction of Extender and Stopper PN-Blocks by Denaturation and Refolding.

To expand possibilities of the PN-Block strategy, I reconstructed multi-component PN-Block complexes from extender PN-Block (ePN-Block) and stopper PN-Block (sPN-Block, i.e., WA20) proteins, by denaturation and refolding (Figure 3-4). Figure 3-4A shows native PAGE analysis of reconstruction of the ePN-Block (HL4) and sPN-Block (WA20) proteins. Before denaturation, band patterns of the ePN-Block (HL4) in a mixture with the sPN-Block did not change. After denaturation and refolding, new-pattern bands and diminished bands appeared with increasing the sPN-Block (the black and gray arrowheads, respectively, in Figure 3-4A), indicating that the ePN-Block (HL4) and sPN-Block formed several heteromeric complexes (esPN-Block complexes). Figure 3-4B shows reconstruction of ePN-Block (FL4) and sPN-Block proteins. After denaturation and refolding, a stronger band and weaker bands were seen with increasing the sPN-Block (the black and gray arrowheads, respectively, in Figure 3-4B), suggesting that the ePN-Block (FL4) and sPN-Block proteins formed heteromeric esPN-Block complexes. These results suggest that several esPN-Block complexes of extended chain-like heterooligomers were reconstructed from ePN-Blocks and sPN-Blocks by denaturation and refolding (Figure 3-4C).

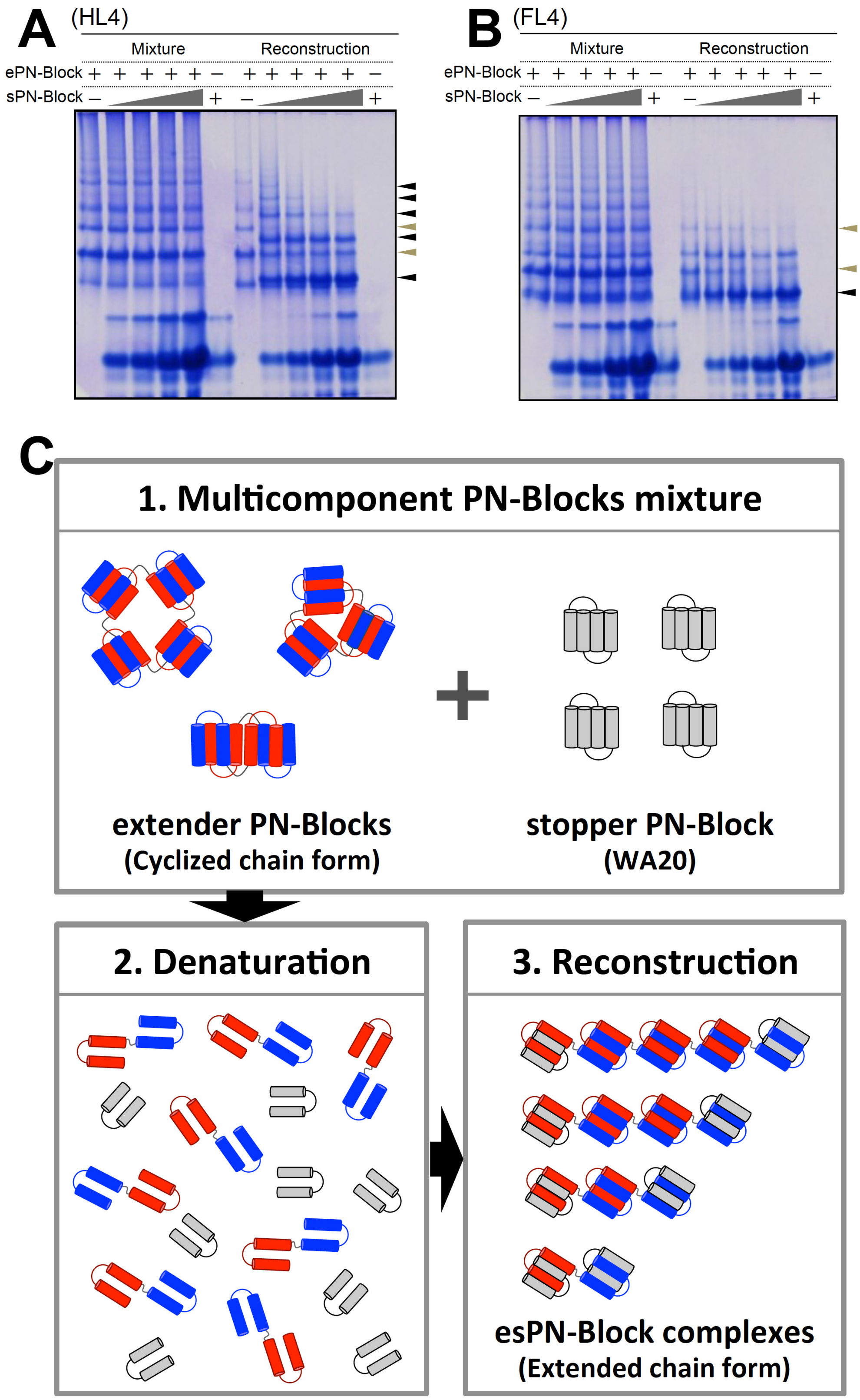


Figure 3-4. Reconstruction of heterooligomeric complexes from multicomponent

PN-Blocks, extender and stopper, by denaturation and refolding. Native PAGE (7.5% polyacrylamide gel) analysis of reconstruction of the extender PN-Block (ePN-block) and stopper PN-Block (sPN-Block): (A) for ePN-block (HL4); (B) for ePN-block (FL4). In the left half, samples were just mixed. In the right half, samples were denatured and refolded after mixing. The sPN-block (WA20) was added in stepwise increase of the ratio of sPN-Block/ePN-Block, 1, 2, 4, and 8.   Proteins were stained with Coomassie brilliant blue. (C) Schematics of the reconstruction process of the ePN-Blocks (red and blue) and sPN-Blocks (gray).

Oligomeric State Analysis of esPN-Block Complexes.

To analyze the oligomeric states of the esPN-Block complexes, I fractionated the esPN-Blocks before/after reconstruction by size exclusion chromatography (SEC) (Figures 3-5–3-8). The chromatograms and SDS-PAGE analysis before reconstruction (Figures 3-5 and 3-7) show that the ePN-Block (HL4 or FL4) and sPN-Block (WA20) were just mixed and did not form heteromeric complexes because the sPN-Block were eluted in the last part fractions only at low molecular size corresponding a homodimer of WA20. In contrast, the chromatograms and SDS-PAGE analysis after reconstruction by denaturation and refolding (Figures 3-6 and 3-8) clearly show that the ePN-Block (HL4 or FL4) and sPN-Block (WA20) formed the heteromeric esPN-Block complexes because sPN-Block (WA20) were eluted in wide range of fractions together with ePN-Block (HL4 or FL4).

I focused on some samples derived from several fractions of the esPN-Block complexes for SAXS analysis (Table 3-2). Some typical samples were also analyzed by SEC-MALS. Table 3-3 shows the summarized results of SEC-MALS analysis (Figures 3-9 and 3-10). The molecular masses of the esPN-Block complexes reveal existence of supramolecular species of the esPN-Block complexes of one ePN-Block and two sPN-Blocks (e1s2), two ePN-Blocks and two sPN-Blocks (e2s2), and three ePN-Blocks and two sPN-Blocks (e3s2), and four ePN-Blocks and two sPN-Blocks (e4s2), corresponding to native PAGE bands 1, 2, 3, and 4, respectively, in Figure 3-11.

In addition, Figures 3-13A and 3-13B show SAXS intensities of the esPN-Block complex samples, sPN-Block (WA20), and chicken egg lysozyme as a molecular mass reference standard (M_{w} = 14.3 kDa). Assuming that these proteins have practically identical scattering length densities and specific volumes and that the structure factor $S(q) \approx 1$ for dilute samples, the forward-scattering intensity normalized by protein concentration, $I(q \rightarrow 0)/c$, is proportional to the weight-average molecular mass (M_{w}). The samples M_{w} are shown in Table 3-4. The results are roughly consistent with the SEC-MALS analysis, in consideration of sample purity of complex components (Figure 3-11 and Table 3-4) and experimental errors.

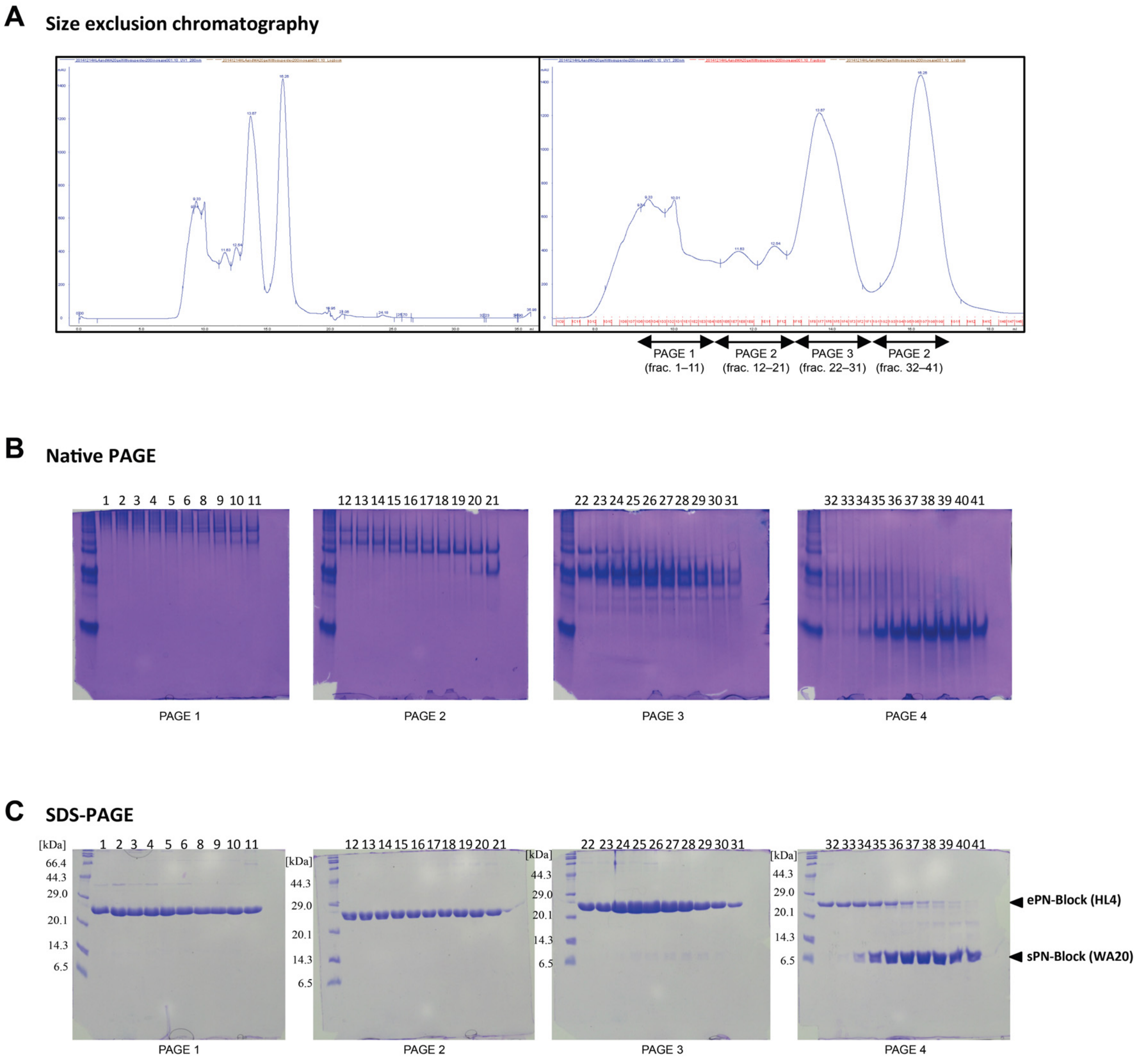


Figure 3-5. Size exclusion chromatography (SEC) of the mixture of ePN-Block (HL4) and sPN-Block (WA20) without reconstruction. (A) The SEC chromatogram of the mixture sample of ePN-Block (HL4) and sPN-Block (WA20) without reconstruction. The right panel is the magnified view of the elution peaks of the chromatogram with fraction numbers. (B) Native PAGE of the eluted fractions of the SEC experiments. The mixture sample before SEC separation was loaded into the far-left lane in native PAGE. (C) SDS-PAGE of the eluted fractions of the SEC experiments. The protein molecular weight marker (broad) (Takara Bio) was loaded into the far-left lane in SDS-PAGE. The proteins were stained with Coomassie brilliant blue.

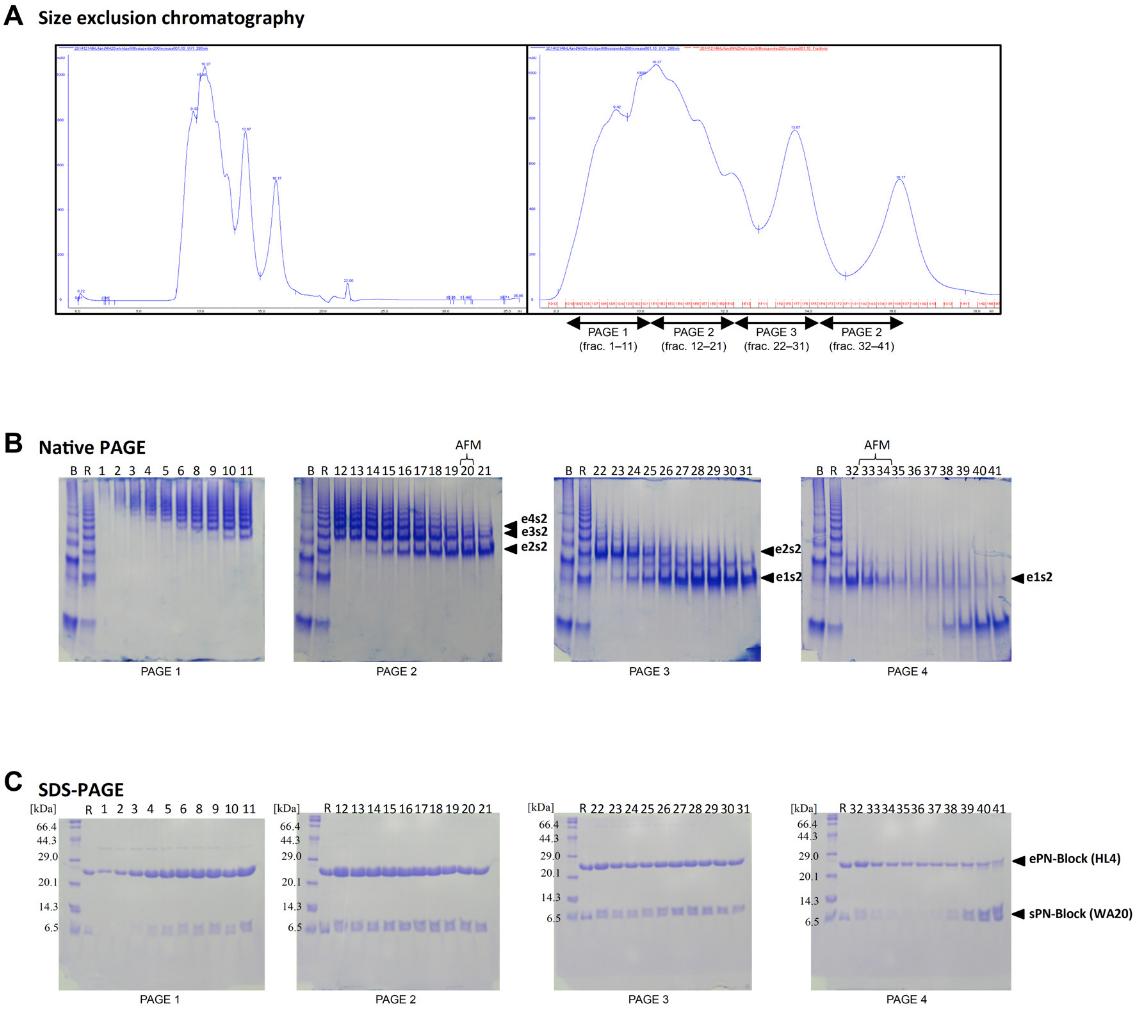


Figure 3-6. Size exclusion chromatography (SEC) of the esPN-Block (HL4) complexes after reconstruction by denature and refolding. (A) The SEC chromatogram of the sample of the esPN-Block (HL4) complexes after reconstruction. The right panel is the magnified view of the elution peaks of the chromatogram with fraction numbers. (B) Native PAGE of the eluted fractions of the SEC experiments. The sample before SEC separation was loaded into the far-left lane in native PAGE. B: the sample, before SEC separation, of the mixture sample before reconstruction. R: the sample, before SEC separation, of the esPN-Block (HL4) complex after reconstruction. (C) SDS-PAGE of the eluted fractions of the SEC experiments. The protein molecular weight marker (broad) (Takara Bio) was loaded into the far-left lane in SDS-PAGE. R: the esPN-Block (HL4) complex sample before SEC separation. The proteins were stained with Coomassie brilliant blue.

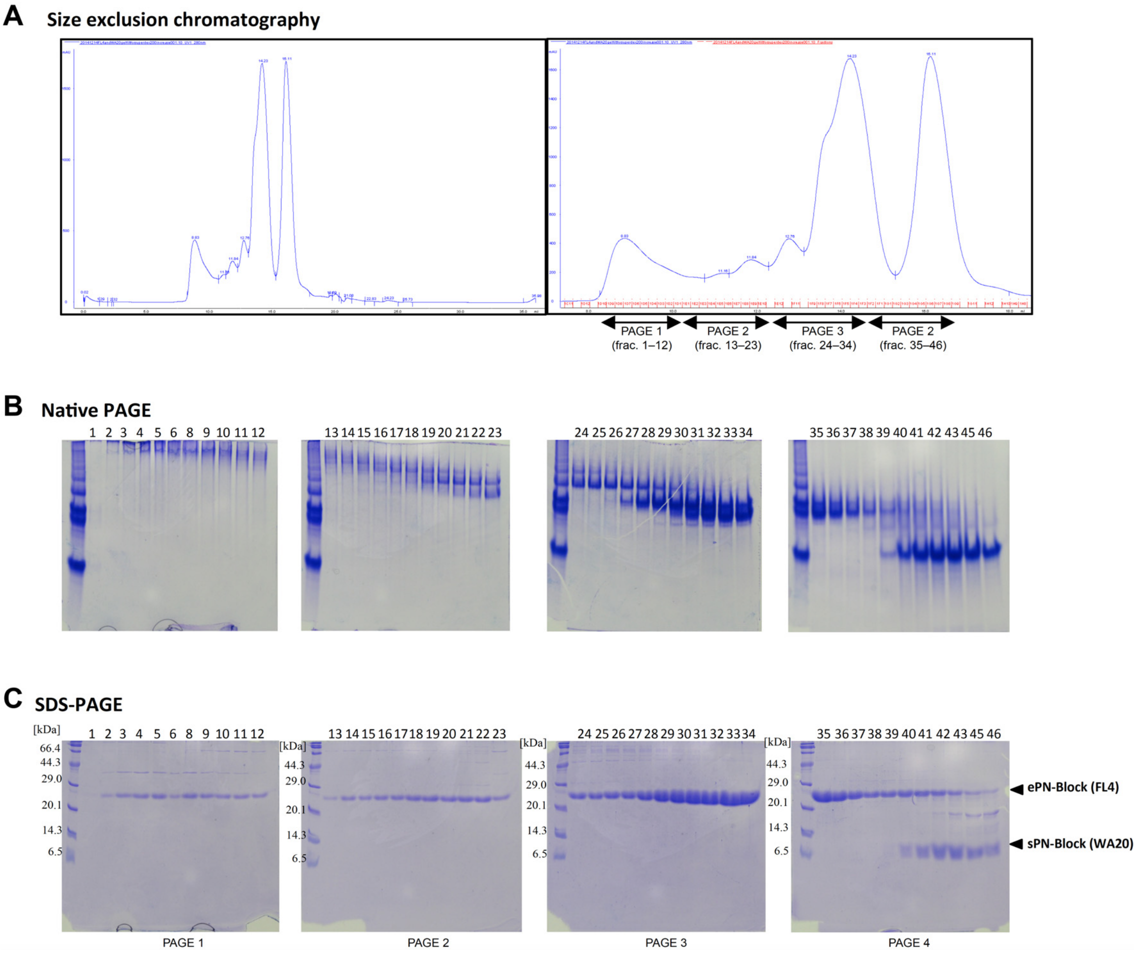


Figure 3-7. Size exclusion chromatography (SEC) of the mixture of ePN-Block (FL4) and sPN-Block (WA20) without reconstruction. (A) The SEC chromatogram of the mixture sample of ePN-Block (FL4) and sPN-Block (WA20) without reconstruction. The right panel is the magnified view of the elution peaks of the chromatogram with fraction numbers. (B) Native PAGE of the eluted fractions of the SEC experiments. The mixture sample before SEC separation was loaded into the far-left lane in native PAGE. (C) SDS-PAGE of the eluted fractions of the SEC experiments. The protein molecular weight marker (broad) (Takara Bio) was loaded into the far-left lane in SDS-PAGE. The proteins were stained with Coomassie brilliant blue.

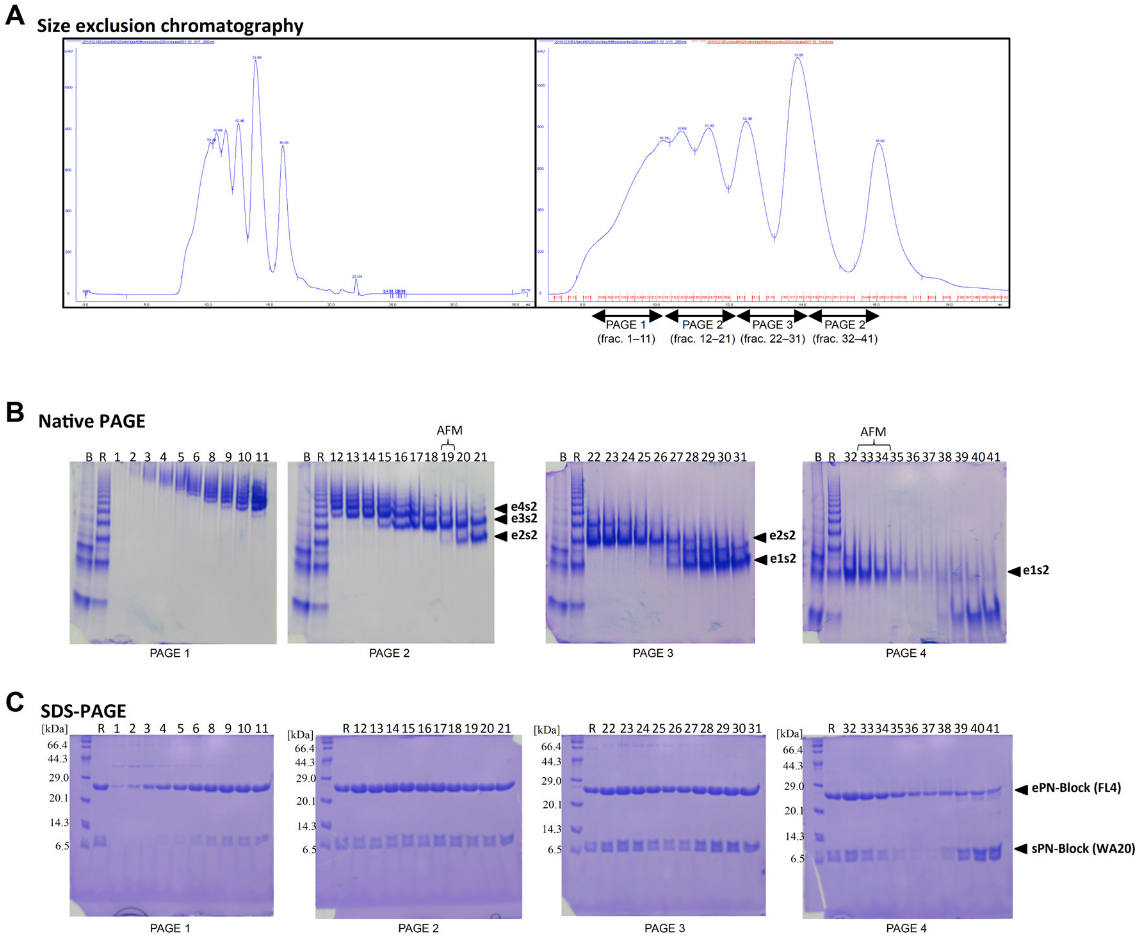


Figure 3-8. Size exclusion chromatography (SEC) of the esPN-Block (FL4) complexes after reconstruction by denature and refolding. (A) The SEC chromatogram of the sample of the esPN-Block (FL4) complexes after reconstruction. The right panel is the magnified view of the elution peaks of the chromatogram with fraction numbers. (B) Native PAGE of the eluted fractions of the SEC experiments. The sample before SEC separation was loaded into the far-left lane in native PAGE. B: the sample, before SEC separation, of the mixture sample before reconstruction. R: the sample, before SEC separation, of the esPN-Block (FL4) complex after reconstruction. (C) SDS-PAGE of the eluted fractions of the SEC experiments. The protein molecular weight marker (broad) (Takara Bio) was loaded into the far-left lane in SDS-PAGE. R: the esPN-Block (FL4) complex sample before SEC separation. The proteins were stained with Coomassie brilliant blue.

Table 3-3 Summary of SEC-MALS analysis

esPN-Block complex	Sample and peak	Molecular mass [kDa]	ePN-Block: sPN-Block	Native PAGE band (Figure 3-11)
HL4	Sample IV, Peak 1	49.5	1:2 (e1s2)	Band 1
HL4	Sample II, Peak 1	83.4	2:2 (e2s2)	Band 2
HL4	Sample II, Peak 2	111.9	3:2 (e3s2)	Band 3
HL4	Sample II, Peak 3	149	4:2 (e4s2)	Band 4
FL4	Sample IV, Peak 1	46.8	1:2 (e1s2)	Band 1
FL4	Sample III, Peak 1	75.8	2:2 (e2s2)	Band 2
FL4	Sample II, Peak 1	106.2	3:2 (e3s2)	Band 3
FL4	Sample II, Peak 2	132.2	4:2 (e4s2)	Band 4

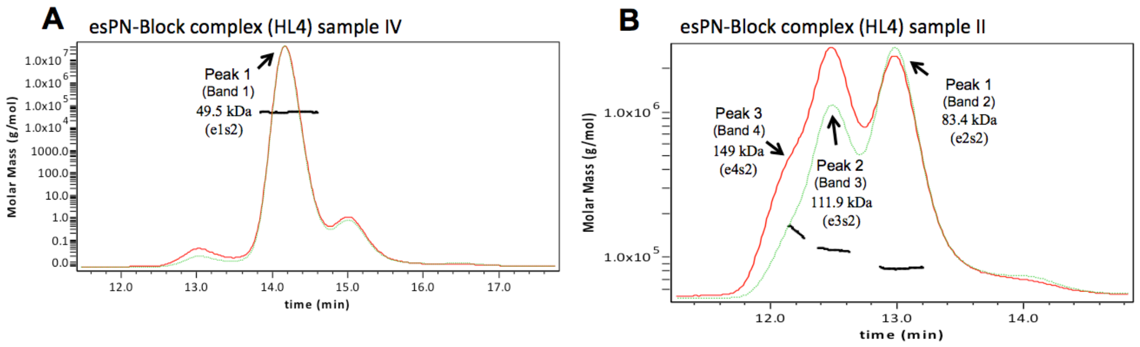


Figure 3-9. Results of SEC-MALS experiments of the esPN-Block complex (HL4) samples with the UV (green) and 90° light scattering (red) chromatograms [arbitrary unit]. Each molecular mass is shown as a black line across the elution peak. (A) Sample IV; the main peak (peak 1) corresponds to the band 1 of the native PAGE in Figure 3-11A. (B) Sample II; the main peaks (peak 1 and peak 2) correspond to the band 2 and band 3 of the native PAGE in Figure 3-11A, respectively; the minor shoulder peak (peak 3) corresponds to the band 4 of the native PAGE in Figure 3-11A.

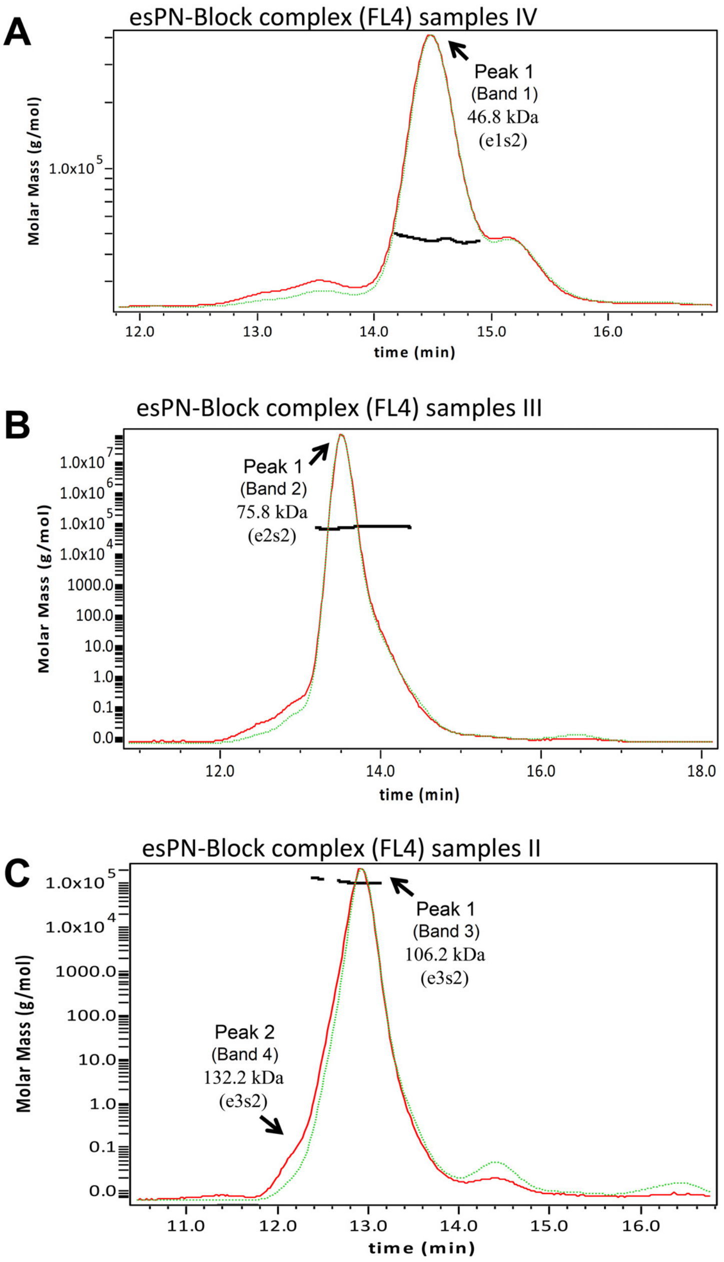


Figure 3-10. Results of SEC-MALS experiments of the esPN-Block complex (FL4) samples with the UV (green) and 90 degree light scattering (red) chromatograms [arbitrary unit]. Each molecular mass is shown as a black line across the elution peak. (A) Sample IV; the main peak (peak 1) corresponds to the band 1 of the native PAGE in Figure 3-11C. (B) Sample III; the main peaks (peak 1) corresponds to the band 2 of the native PAGE in Figure 3-11C. (C) Sample II; the main peaks (peak 1) corresponds to the band 3 of the native PAGE in Figure 3-11C; the minor shoulder peak (peak 2) corresponds to the band 4 of the native PAGE in Figure 3-11C.

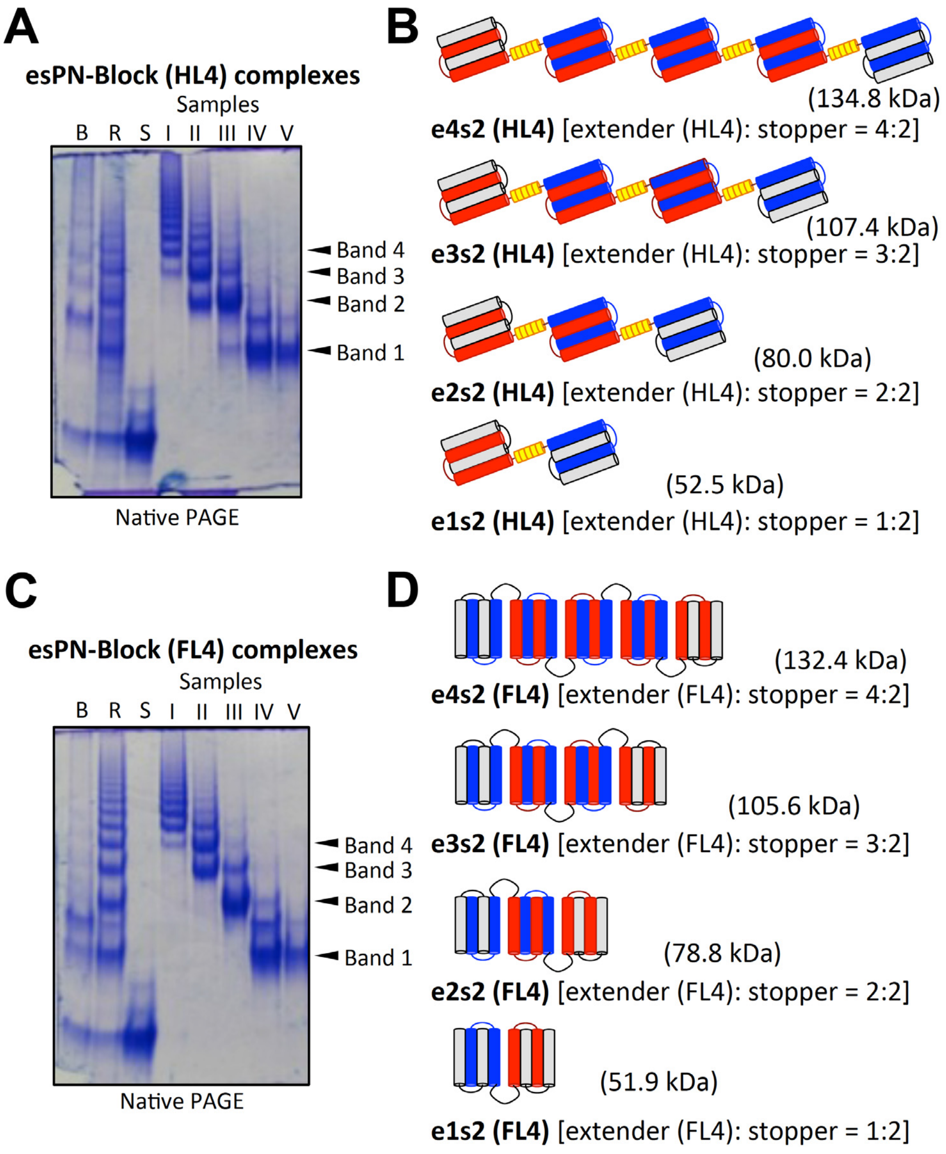

A
esPN-Block (HL4) complexes
Samples
B R S I II III IV V
Band 4
Band 3
Band 2
Band 1
Native PAGE
B
(134.8 kDa)
e4s2 (HL4) [extender (HL4): stopper = 4:2]
(107.4 kDa)
e3s2 (HL4) [extender (HL4): stopper = 3:2]
(80.0 kDa)
e2s2 (HL4) [extender (HL4): stopper = 2:2]
(52.5 kDa)
e1s2 (HL4) [extender (HL4): stopper = 1:2]
C
esPN-Block (FL4) complexes
Samples
B R S I II III IV V
Band 4
Band 3
Band 2
Band 1
Native PAGE
D
(132.4 kDa)
e4s2 (FL4) [extender (FL4): stopper = 4:2]
(105.6 kDa)
e3s2 (FL4) [extender (FL4): stopper = 3:2]
(78.8 kDa)
e2s2 (FL4) [extender (FL4): stopper = 2:2]
(51.9 kDa)
e1s2 (FL4) [extender (FL4): stopper = 1:2]

Figure 3-11. Native PAGE analysis of the esPN-Block complex samples (Table 3-2) separated by size exclusion chromatography. (A) Native PAGE (7.5% polyacrylamide gel) of the esPN-Block complexes (HL4). B: the mixture sample of ePN-Block and sPN-Block before reconstruction and SEC separation. R: the sample, before SEC separation, of the esPN-Block complexes after reconstruction. S: sPN-Block (WA20). (B) Schematics of the esPN-Block complexes (HL4). (C) Native PAGE (7.5% polyacrylamide gel) of the esPN-Block complexes (FL4). (D) Schematics of esPN-Block (FL4). Theoretical molecular masses of the esPN-Block complexes are described in parentheses. The proteins were stained with Coomassie brilliant blue. SDS-PAGE of these samples is also shown in Figure 3−12.

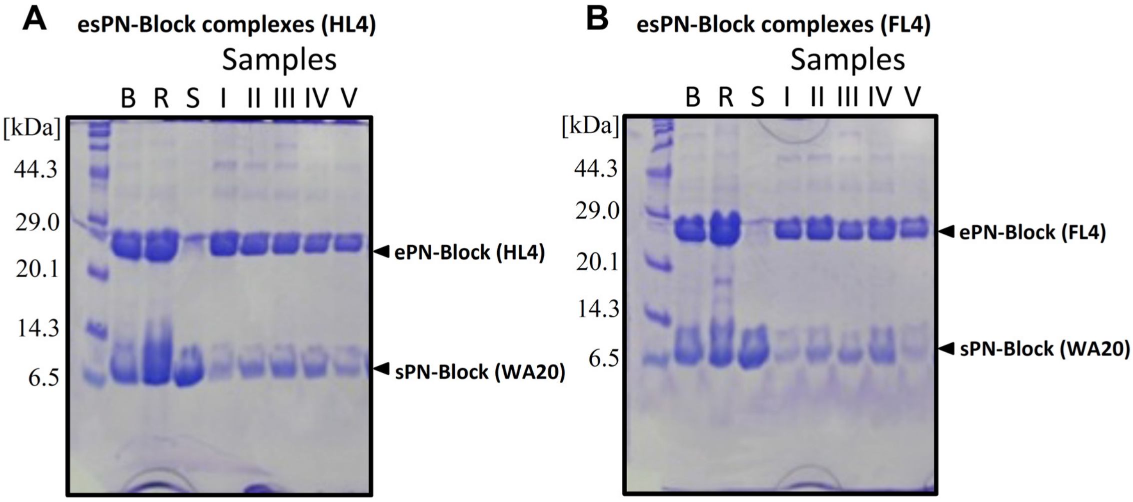


Figure 3-12. SDS-PAGE of the esPN-Block complex samples (Table 3-2) separated by size exclusion chromatography. (A) SDS-PAGE (17.5% gel) of the esPN-Block complexes (HL4). (B) SDS-PAGE (17.5% gel) of the esPN-Block complexes (FL4). B: the mixture sample of ePN-Block and sPN-Block before reconstruction and SEC separation. R: the sample, before SEC separation, of the esPN-Block complexes after reconstruction. S: sPN-Block (WA20). The proteins were stained with Coomassie brilliant blue. Native PAGE of these samples is shown in Figures 3-11A and 3-11C.

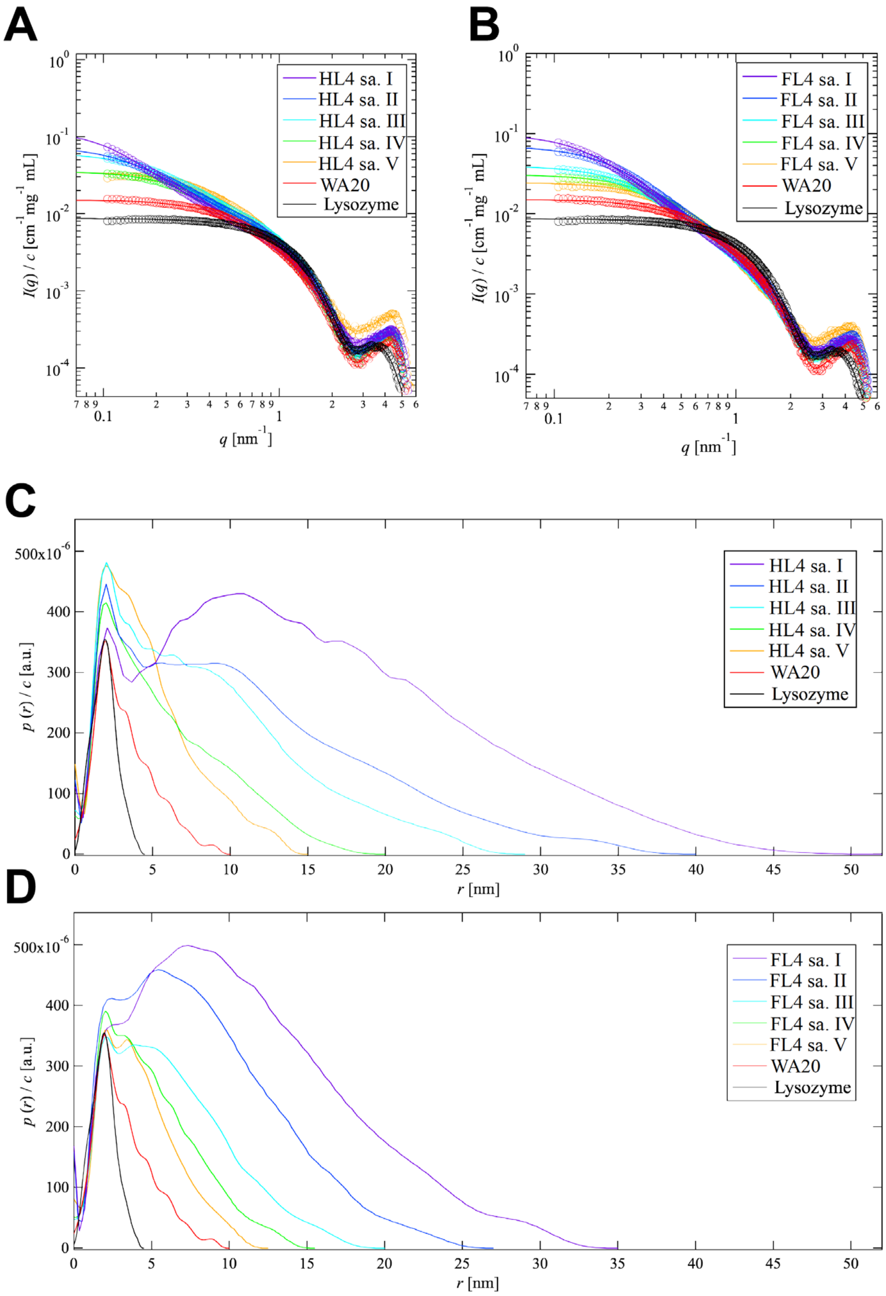


Figure 3-13. SAXS analysis of the esPN-Block complex samples. Concentration-normalized absolute scattering intensities, $I(q)/c$, of (A) the esPN-Block complex (HL4) samples and (B) the esPN-Block complex (FL4) samples. WA20 is a control sample and chicken egg lysozyme is a molecular mass reference standard. Their real-space function, pair-distance distribution functions normalized by the concentration, $p(r)/c$, of (C) the HL4 samples and (D) the FL4 samples, as obtained by IFT.

Table 3-4 Summary of SAXS analysis.

esPN-Block complex samples	$I(q)/c$ [cm^{-1}mg^{-1}mL]	R_g [nm]	D_{max} [nm]	M_w [kDa]	esPN-Block complexes [a]
HL4, Sample I	0.133	11.0	52	219.4	e3s2, **e4s2, higher**
HL4, Sample II	0.076	8.6	40	125.0	e2s2, **e3s2, e4s2**, higher
HL4, Sample III	0.062	6.9	29	101.8	**e2s2**, e3s2
HL4, Sample IV	0.036	4.7	20	59.0	**e1s2**, e2s2
HL4, Sample V	0.035	3.6	15	58.1	**e1s2**
FL4, Sample I	0.102	8.8	35	167.3	e4s2, **higher**
FL4, Sample II	0.071	6.4	27	117.2	**e3s2, e4s2**, higher
FL4, Sample III	0.040	4.8	20	65.9	**e2s2**, e3s2
FL4, Sample IV	0.031	3.6	15	51.0	**e1s2**, e2s2
FL4, Sample V	0.025	3.3	12	40.8	**e1s2**
sPN-Block (WA20)	0.015	2.6	10	25.0	
Lysozyme	0.009	1.5	4.5	14.3	

[a] Main components are indicated by boldface.

Shape Analysis of esPN-Block Complexes.

To extract intuitive real-space information from the SAXS data, the pair-distance distribution functions, $p(r)$, reflected by shapes of the esPN-Block complex samples were obtained using the indirect Fourier transformation technique (Brunner-Popela and Glatter, 1997; Glatter, 1980b; Glatter and Kratky, 1982) (Figures 3-13C and 3-13D). The integral of $p(r)/c$ from $r = 0$ to $r = D_{\max}$ is equal to the extrapolated forward absolute scattering intensity normalized by concentration, $I(q\rightarrow0)/c$, and therefore, it is proportional to M_{w}. The $p(r)$ series of the esPN-Block complex (HL4) samples (Figure 3-13C) is characterized by an extended tail in the high-r regime, suggesting that the esPN-Block (HL4) complexes form extended shapes. In contrast, the $p(r)$ series of the esPN-Block (FL4) complexes show shorter $D_{\max}$ than those of the corresponding samples (HL4), suggesting that the esPN-Block (FL4) complexes form relatively more compact shapes than the corresponding esPN-Block complexes (HL4).

Further analyses for the esPN-Block complexes was performed to obtain more detailed structural insights from the SAXS data. As shown in native PAGE (Figure 3-11A), the esPN-Block complex (HL4) sample V was mostly composed of the esPN-Block complex e1s2 (HL4). Therefore, a rigid-body model structure of the esPN-Block complex e1s2 (HL4) was constructed based on the crystal structure of the WA20 dimer (PDB code 3VJF) (Arai et al., 2012) to explain the experimental $p(r)$ with a consideration of the helical linker rigidity in linking their C and N terminals. The model of the e1s2 (HL4) complex shows an extended "Z" shape (Figure 3-14A). The $p(r)$ simulated from the rigid-body model resembles that obtained from the SAXS experiment (Figure 3-14B). To interpret the $p(r)$ of the esPN-Block complex (HL4) sample III comprising mainly the e2s2 and e3s2 complexes (HL4), the rigid body model structures of the e2s2 and e3s2 complexes (HL4) were constructed by appropriately connecting the rigid-body models of the e1s2 complex (HL4) (Figure 3-14A) as a basic structural unit (Figure 3-15). The models (Figure 3-14C and 3-14D) resemble repeated chain-like structures of extended Z shapes. Assuming a ratio of the structures (e2s2:e3s2 = 5:3), the experimental $p(r)$ is roughly explained (Figure 3-14E). The differences between the experimental and calculated $p(r)$ are possibly due to dynamically multi-conformational structures and/or impurity with minor components of the complexes.

In the other case, a rigid-body model structure of the esPN-Block complex e1s2

(FL4) was constructed based on the WA20 structure (Figure 3-16A). The “V” form rigid-body model for the e1s2 (FL4) complex is shown in Figure 3-16A. The *p*(*r*) calculated from only the V form model poorly resembles that from the SAXS experiment (Figure 3-17A). An additional rigid-body model of a compact form (C form) was also constructed with a consideration of the linker flexibility and domain interaction. The *p*(*r*) of the esPN-Block complex (FL4) sample V composing mainly the e1s1 (FL4) complex can be simulated apparently well with an equal ratio of the V and C forms (Figure 3-16B), probably suggesting that dynamic structures of the e1s1 (FL4) complex represent structural ensemble including the V form, C form, and many transient forms between them. To simulate the *p*(*r*) of the esPN-Block complex (FL4) sample III including e2s2 and e3s2 complexes mainly, an ensemble of multicomponents and multi-conformation structures should be considered. For example, three rigid-body models for the e2s2 complex (FL4), a compact form (C form) and two forms as a letter "N" (N_1 form and N_2 form), were constructed (Figure 3-16C). A rigid-body model of the e3s2 complex (FL4), a form like a letter "W" (W form) was also constructed (Figure 3-16D). Assum-ing a ratio of the forms ($C:N_1:N_2:W = 1:1:3:2$), the *p*(*r*) is apparently explained as a composite *p*(*r*) function from *p*(*r*) of these models (Figure 3-17B). These results imply that the e2s2 and e3s2 complexes (FL4) in the sample III have multicomponents and dynamic structural ensemble including those forms (Figure 3-16D) and various intermediate forms.

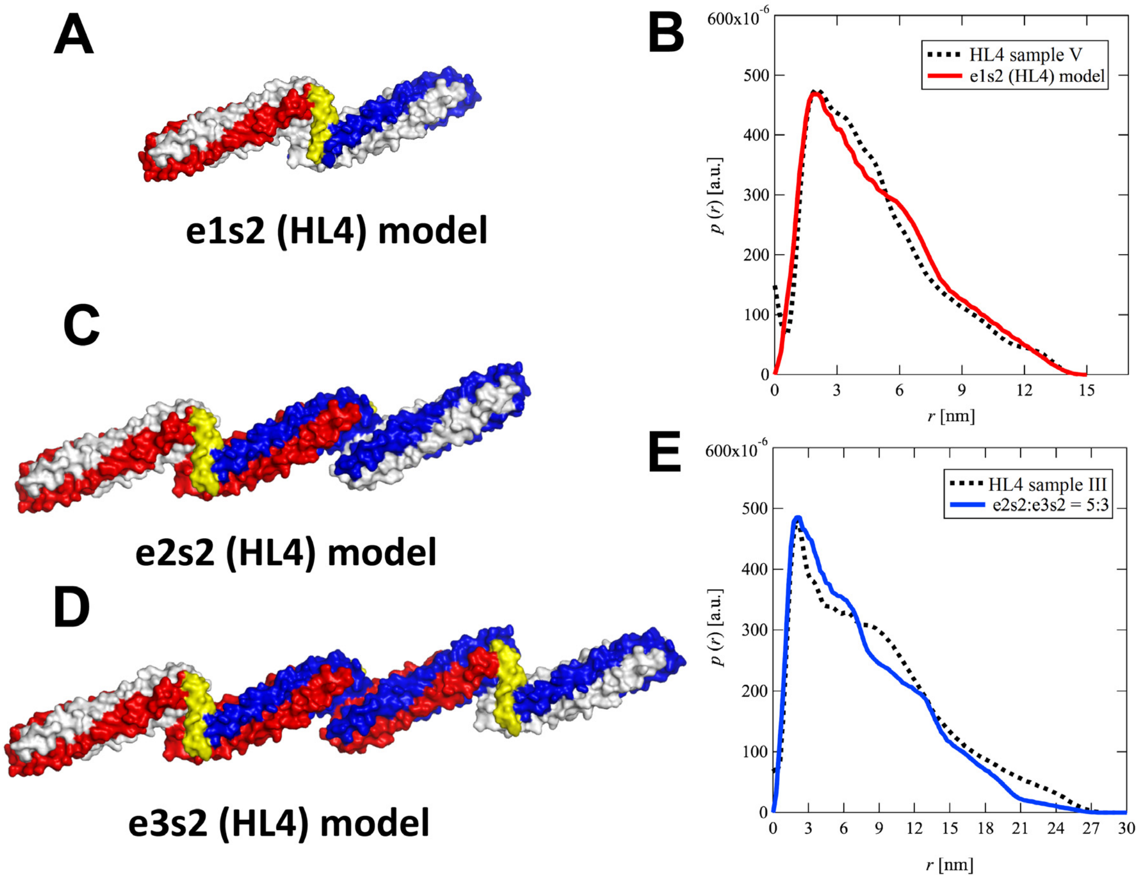


Figure 3-14. Rigid-body models of the esPN-Block complexes with the helical linker (HL4), derived from SAXS analysis. (A) The rigid-body model structure of one extender and two stopper (e1s2) of the esPN-Block complex (HL4). The model is shown in the same colors in Figure 3-2: the first WA20 domain (red), the helical linker (yellow), and the second WA20 domain (blue) of ePN-Block; sPN-Block (gray). (B) The pair-distance distribution function, $p(r)$, of the esPN-Block complex (HL4) sample V as obtained by the SAXS experiment (black dash line) and $p(r)$ simulated from the rigid-body model structure (red line). The rigid-body model structures of e2s2 (C) and e3s2 (D) of the esPN-Block complexes (HL4). The models are shown in the same colors as above. (E) The pair-distance distribution function, $p(r)$, of the esPN-Block complex (HL4) sample III as obtained by the SAXS experiment (black dash line) and the composite $p(r)$ simulated from the rigid-body model structures (e2s2:e3s2 = 5:3) (blue line).

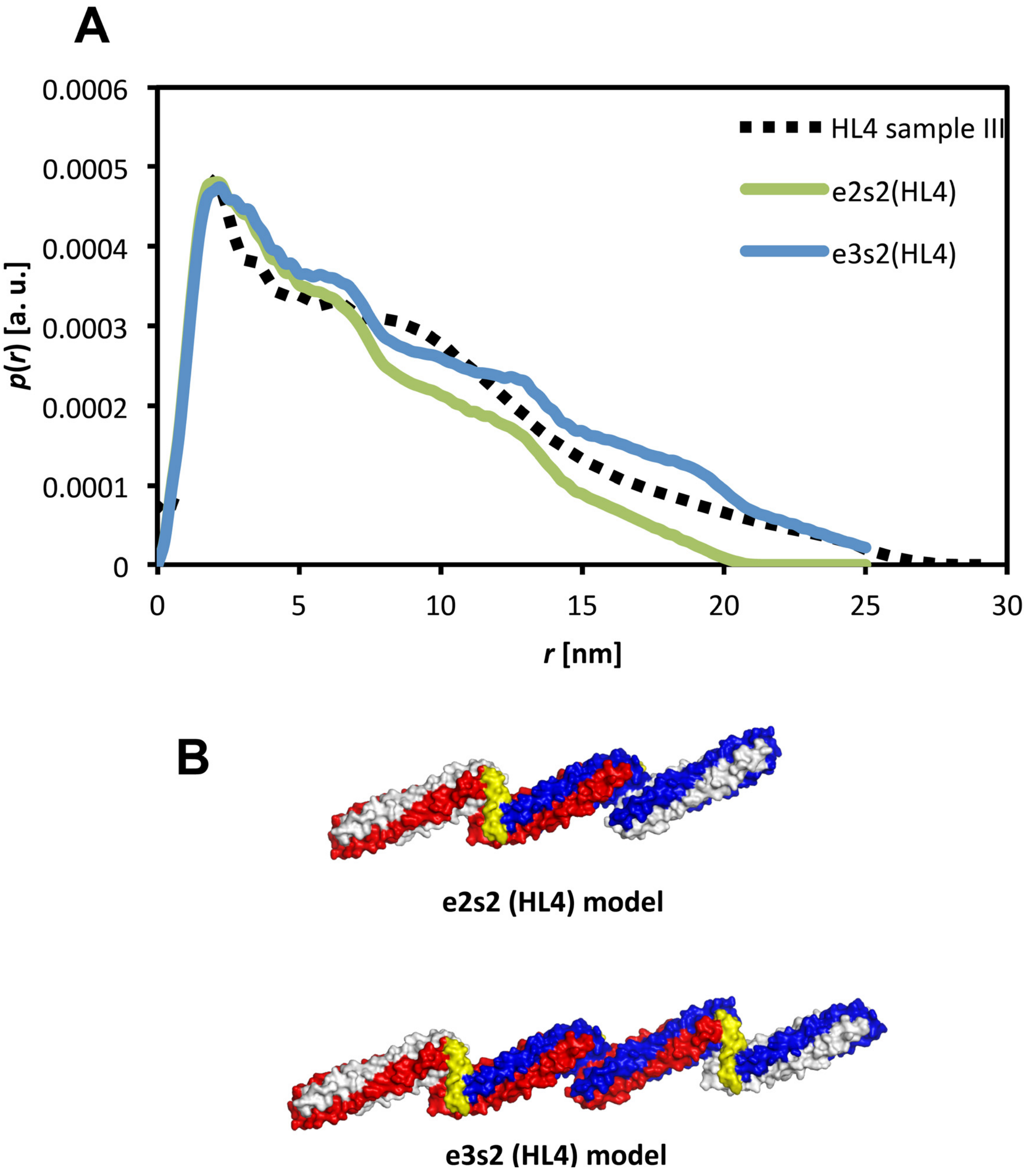


Figure 3-15. (A) The pair-distance distribution function, *p*(*r*), of the esPN-Block complex (HL4) sample III as obtained by the SAXS experiment (black dash line) and each *p*(*r*) simulated from the rigid-body model structures of e2s2 (HL4) (green line) and e3s2 (HL4) (blue line). (B) The rigid-body model structures of e2s2 and e3s2 of the esPN-Block complexes (HL4). The model is shown in the same colors in Figure 3-14: the first WA20 domain (red), the helical linker (yellow), and the second WA20 domain (blue) of ePN-Block; sPN-Block (gray).

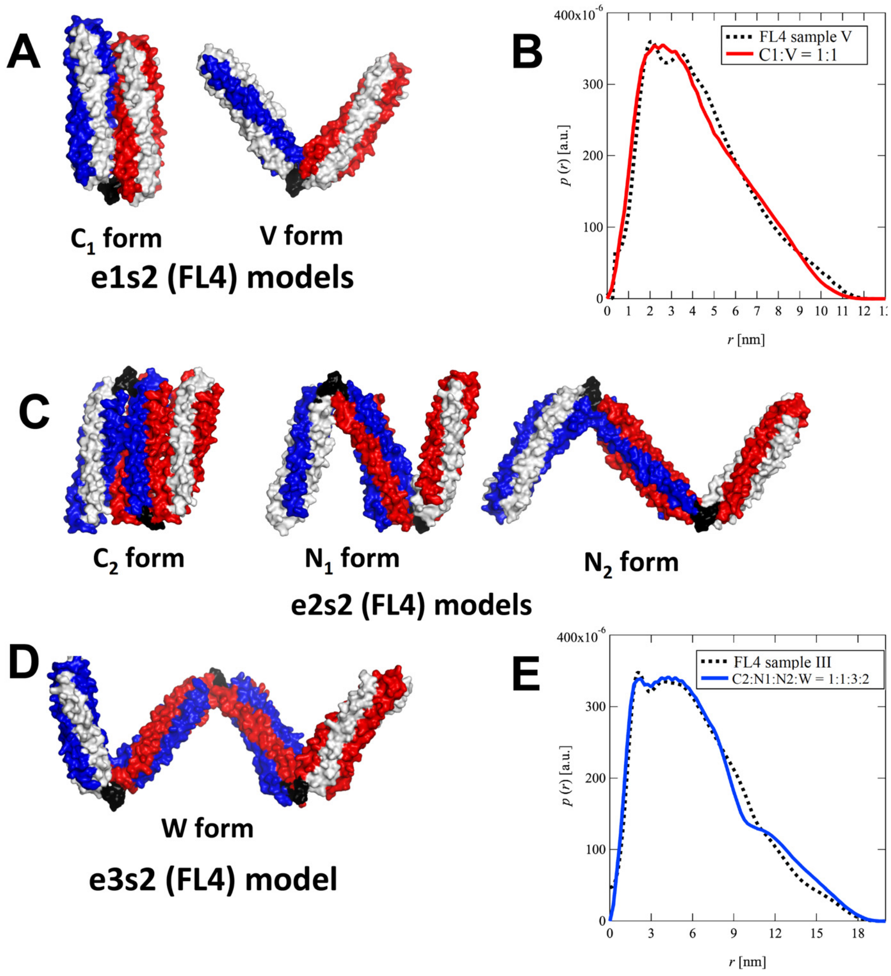


Figure 3-16. Rigid-body models of the esPN-Block complexes with the flexible linker (FL4), derived from SAXS analysis. (A) The rigid-body model structures of one extender and two stopper (e1s2) of the esPN-Block complex (FL4). The models are shown in the same colors in Figure 3-2: the first WA20 domain (red), the flexible linker (black), and the second WA20 domain (blue) of ePN-Block; sPN-Block (gray). (B) The pair-distance distribution function, *p*(*r*), of the esPN-Block complex (FL4) sample V as obtained by the SAXS experiment (black dash line) and the composite *p*(*r*) simulated from the rigid-body model structures of e1s2 (C1 form:V form = 1:1) (red line). The rigid-body model structures of e2s2 (C) and e3s2 (D) of the esPN-Block complexes (FL4). The models are shown in the same colors as above. (E) The pair-distance distribution function, *p*(*r*), of the esPN-Block complex (FL4) sample III as obtained by the SAXS experiment (black dash line) and the composite *p*(*r*) simulated from the rigid-body model structures of e2s2 and e3s2 (forms C_2:N_1:N_2:W = 1:1:3:2) (blue line).

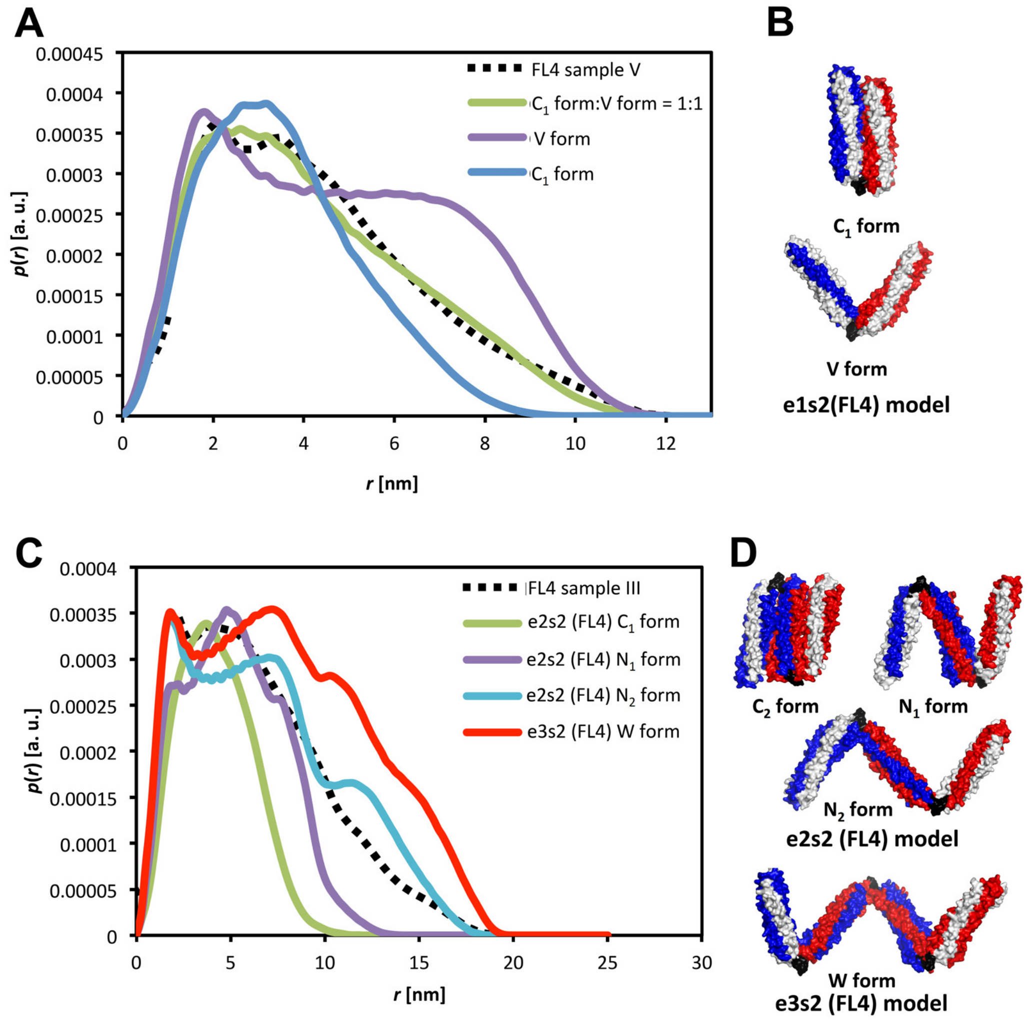


Figure 3-17. (A) The pair-distance distribution function, *p*(*r*), of the esPN-Block complex (FL4) sample III as obtained by the SAXS experiment (black dash line) and each *p*(*r*) simulated from the rigid-body model structures of the esPN-Block complexes (FL4), e1s2 C_1 form (blue line), e1s2 V form (purple line), and the composite function (C_1 form:V form = 1:1) (green line). (B) The rigid-body model structures of C_1 form and V form of the esPN-Block complexes (FL4). The model is shown in the same colors in Figure 3-16: the first WA20 domain (red), the flexible linker (black), and the second WA20 domain (blue) of ePN-Block; sPN-Block (gray). (C) The pair-distance distribution function, *p*(*r*), of the esPN-Block complex (FL4) sample III as obtained by the SAXS experiment (black dash line) and each *p*(*r*) simulated from the rigid-body model structures of the esPN-Block complexes (FL4), e2s2 C_2 form (green line), e2s2 N_1 form (purple line), e2s2 N_2 form (blue line), and e3s2 W form (red line). The models are shown in the same colors as above.

Self-Assembling Supra-Quaternary Nanostructures of esPN-Block Complexes Observed by AFM in Liquid.

To expand possibilities of the PN-Block strategy for construction of supra-quaternary structures, the esPN-Block complexes self-assembled into supramolecular nanostructures with nickel ion (Ni^{2+}) on mica surface, and they were observed by frequency modulation atomic force microscopy (FM-AFM) in liquid (Figure 3-18 and Figure 3-19). Figure 3-18A shows an AFM image of self-assembling supramolecular nanostructures of the esPN-Block complex (HL4) fraction 20 (Figure 3-6) comprising mainly the e2s2 and e3s2 complexes (HL4). As shown in Figure 3-14A, a number of bundles of rod-like structures with a length of ~10 nm and a width of ~3 nm are found in the AFM image, representing the WA20 structure as a structural domain unit significantly consistent with the crystal structure (PDB code 3VJF) of the WA20 dimer (Arai et al., 2012). Several domains of WA20 units align laterally and they line up in the longitudinal axis direction (Figures 3-18A and 3-18B). In contrast, Figure 3-18C shows an AFM image of self-assembling nanostructures of the esPN-Block complexes (FL4) fraction 19 (Figure 3-8) comprising the e2s2, e3s2, and e4s2 (FL4) complexes. A number of bundles of rod-like structures with a length of ~10 nm and a width of ~3 nm, representing the WA20 structural domain, are also observed. However, more than several numbers of the rod-like structural domains line up and extend in the lateral axis direction (Figures 3-18C and 3-18D). These contrasting observations probably reflect the different structural properties of the esPN-Block complexes due to differences in rigidity and flexibility of the linkers, the helical linker (HL4) and the flexible linker (FL4).

In contrast, Figure 3-19A shows an AFM image of self-assembling supramolecular nanostructures of the esPN-Block complex (HL4) fractions 33 and 34 (Figure 3-6) comprising mainly the e1s2 complex (HL4). Several domains of WA20 units may be found but obscure. Figure 3-19C shows an AFM image of self-assembling nanostructures of the esPN-Block complex (FL4) fractions 33 and 34 (Figure 3-8) comprising the e1s2 (FL4) complexes. A number of bundles of rod-like structures with a length of ~10 nm and a width of ~3 nm, representing the WA20 structural domain, are observed.

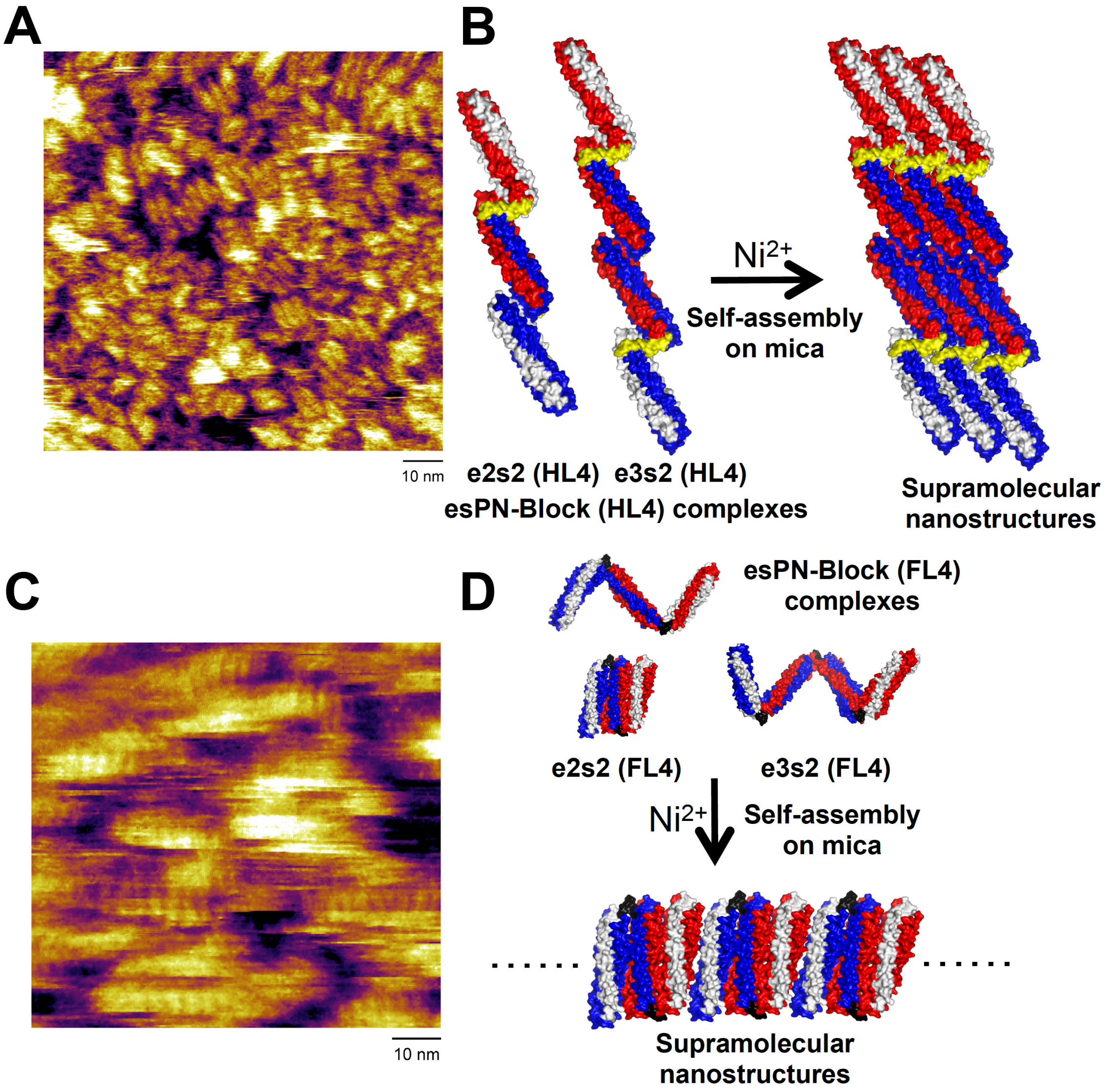


Figure 3-18. Atomic force microscopy (AFM) imaging of the self-assembling supramolecular nanostructures of the esPN-Block complexes with nickel ion (Ni^{2+}) on mica surface in liquid. (A) AFM image of the esPN-Block complex (HL4) fraction 20 (Figure 3-6) comprising mainly the e2s2 and e3s2 (HL4) complexes. (B) Schematics of self-assembling supramolecular nanostructures of the esPN-Block complexes (HL4) with Ni^{2+} on mica surface, based on the rigid-body models (e2s2 and e3s2) shown in Figure 3-14. (C) AFM image of the esPN-Block complex (FL4) fraction 19 (Figure 3-8) comprising the e2s2, e3s2, and e4s2 (FL4) complexes. (D) Schematics of self-assembling supramolecular nanostructures of the esPN-Block complexes (FL4) with Ni^{2+} on mica surface, based on the rigid-body models (e2s2 and e3s2) shown in Figure 3-16.

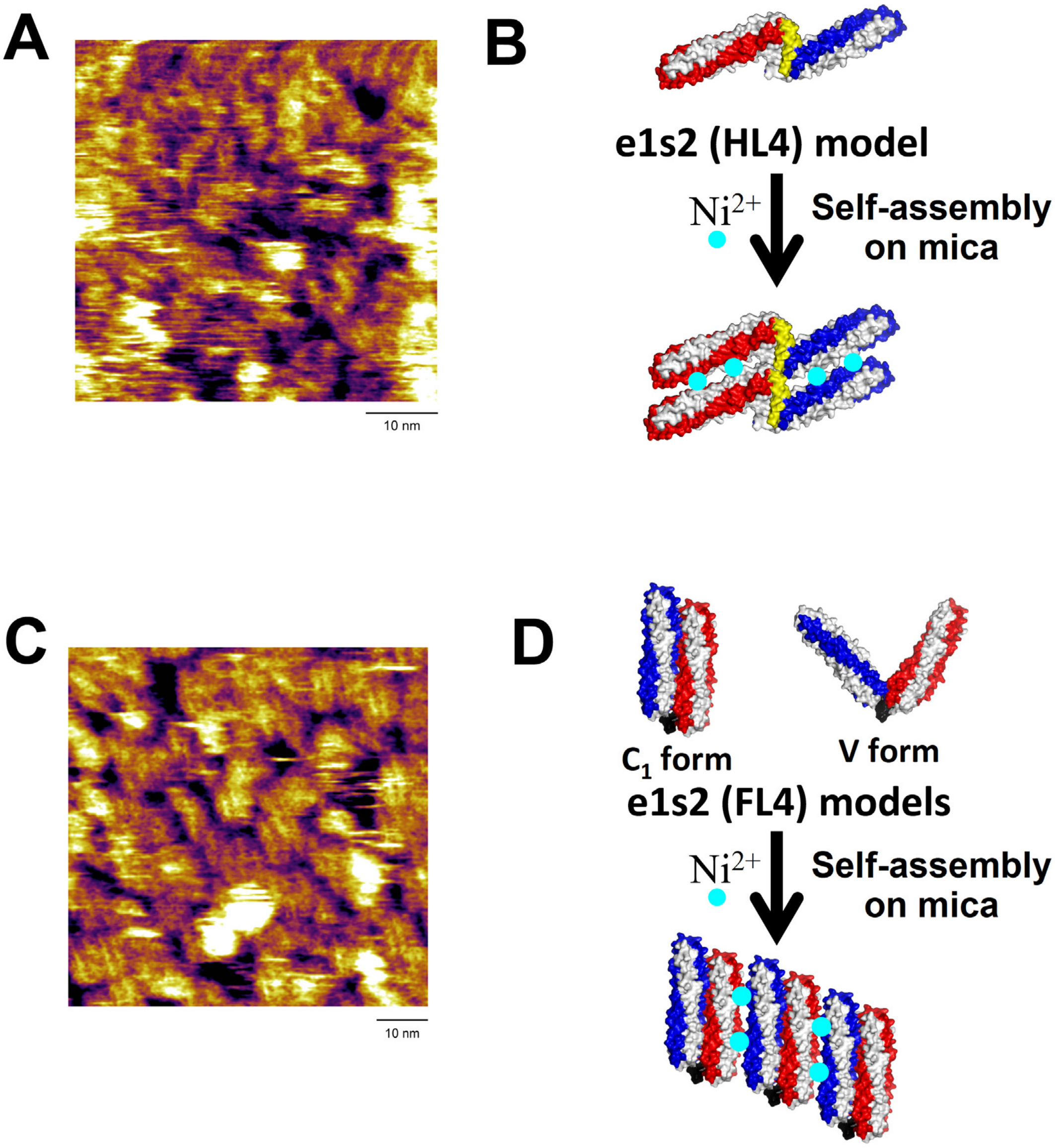


Figure 3-19. Atomic force microscopy (AFM) imaging of the self-assembling supramolecular nanostructures of the esPN-Block complexes with nickel ion (Ni^{2+}) on mica surface in liquid. (A) AFM image of the esPN-Block complex (HL4) sample V comprising mainly the e1s2 (HL4) complex. (B) Schematics of self-assembling supramolecular nanostructures of the esPN-Block complex (HL4) with Ni^{2+} on mica surface, based on the rigid-body model (e1s2) shown in Figure 3-14A. (C) AFM image of the esPN-Block complex (FL4) fractions 33 and 34 (Figure 3-6) comprising mainly the e1s2 (FL4) complex. (D) Schematics of self-assembling supramolecular nanostructures of the esPN-Block complex (FL4) with Ni^{2+} on mica surface, based on the rigid-body models (e1s2) shown in Figure 3-16A.

Hierarchical Structure Design of *De Novo* Proteins by the Binary Code Strategy and PN-Block approach.

Using the second series of PN-Blocks, I have demonstrated that not only self-assembling homooligomeric quaternary structures were constructed but also quaternary structures of heterooligomeric supramolecular complexes were efficiently reconstructed from different types of PN-Blocks, ePN-Blocks and sPN-Blocks, by denaturation and refolding. The findings significantly increase the possibilities of the PN-block approach as artificial building-block molecules. Moreover, I demonstrate that the complexes can further self-assemble into supramolecular nanostructures on mica surface as a “supra-quaternary structures,” expanding further possibilities of the PN-block approach in the field of nanotechnology. These results suggest that the PN-Blocks are versatile building blocks to create novel self-assembling supramolecular nanostructures of protein complexes on the hierarchical levels of tertiary, quaternary, and supra-quaternary structures. It is noteworthy that the second series of the PN-Blocks are fully *de novo* proteins, which have no sequences derived from any natural proteins. The *de novo* ePN-Block proteins are designed by tandemly linking two WA20 domains of the intermolecularly folded dimeric *de novo* protein (Arai et al., 2012), created by the binary code strategy (Hecht et al., 2004; Kamtekar et al., 1993), with the artificial linker sequences (Arai et al., 2001). On the hierarchical levels of primary, secondary, and tertiary structures for *de novo* proteins, the binary code strategy has been successfully developed to design and create *de novo* artificial superfamily proteins including variously functional *de novo* proteins *in vitro* (Cherny et al., 2012; Patel et al., 2009; Patel and Hecht, 2012) (e.g. cofactor and drug binding, and enzyme-like functions) and *in vivo* (e.g. life-sustaining functions (Digianantonio and Hecht, 2016; Fisher et al., 2011; Hoegler and Hecht, 2016; Smith et al., 2015)). Therefore, the hybrid approach of the binary code and PN-Block strategies can be a smart way of semirational design to create nanostructural and functional *de novo* protein complexes through all hierarchical levels from primary to quaternary structures, and additionally supra-quaternary structures.

Chapter 4 Conclusions and Future Directions

In chapter 1, I described that the *de novo* protein WA20 forms a stable domain-swapped four-helix dimer. These results demonstrate that our *de novo* library of proteins contains not only simple monomeric proteins but also self-assembling, stable, and functional multimeric proteins. This structure also suggests potential binding sites for heme cofactor binding and coordination, as well as sites for hydrolase-substrate binding. These data suggest that the binary patterning strategy can be used to design libraries of more complicated multimeric structures, which may pave the way for the discovery of new functions with applications in synthetic biology and biotechnology.

In chapter 2, I described the design and development of the self-assembling WA20-foldon fusion protein as a novel protein nanobuilding block (PN-Block) using the intermolecularly folded dimeric *de novo* protein WA20. The study revealed that the WA20-foldon, as one PN-Block, simultaneously formed several distinctive types of self-assembling homooligomers in multiples of 6-mer because of the combination of the WA20 dimer and foldon trimer. The SAXS analyses suggest that the S and M forms of the WA20-foldon exist in nanostructures of a barrel-like-shaped hexamer and tetrahedron-like-shaped dodecamer, respectively. These results demonstrate that the PN-Block strategy using the intertwined dimeric *de novo* protein is a powerful strategy to create self-assembling supramolecular nanoarchitectures.

In chapter 3, I described the design and construction of *de novo* extender protein nanobuilding blocks (ePN-Blocks) by fusing tandemly two *de novo* WA20 proteins with various linkers, as the new series of PN-Blocks. The purified ePN-Blocks migrated as ladder bands in native PAGE, suggesting that the ePN-Blocks form several homooligomeric states, probably chain-like structures. Then, I reconstructed heterooligomeric esPN-Block complexes from ePN-Blocks and sPN-Blocks by denaturation and refolding. SEC-MALS and SAXS analyses show that esPN-Block complexes formed different types of extended chain-like structures depending on their linker types. These results demonstrate that heterooligomeric complexes can be reconstructed from different PN-Blocks by denaturation and refolding, enhancing the potential of the PN-block strategy as artificial building-block molecules. Moreover, observation by AFM in liquid revealed that the esPN-Block complexes further self-assembled into supramolecular nanostructures on mica surface. These results suggest that the PN-Block strategy is a powerful strategy to create novel self-assembling supramolecular nanostructures of *de novo* protein complexes.

In chapter 2 and chapter 3, I designed and created PN-Blocks for different types of nanostructures using *de novo* protein WA20 based on the concept of the "PN-Block strategy": various self-assembling nanostructures are created from a few types of simple and fundamental PN-Blocks. PN-Blocks using the intermolecularly folded dimeric *de novo* protein (e.g., WA20) as a PN-Block's component have some advantages: (1) the simple, stable, and intertwined rod-like structure of the *de novo* protein makes it easy to use PN-Blocks to design and construct simple and stable frameworks of nano-architectures (Kobayashi et al., 2015), and (2) the PN-Blocks using the *de novo* protein, based on the simple binary patterning, have great potential for redesigning functional protein nanostructures using binary-patterned *de novo* protein libraries functionable *in vitro* (Cherny et al., 2012; Patel et al., 2009; Patel and Hecht, 2012) and *in vivo* (Digianantonio and Hecht, 2016; Fisher et al., 2011; Hoegler and Hecht, 2016; Smith et al., 2015). Further designing and creating new types of PN-Blocks and reconstructing various PN-Blocks are essential steps to expanding the PN-Block strategy.

In this thesis, I demonstrated the design and construction of self-assembling nanostructures created from protein nanobuilding blocks using the intermolecular folding structure of the dimeric *de novo* protein WA20. Here we propose the "PN-Block strategy," a systematic design and construction strategy to create novel self-assembling supramolecular nanostructures on the hierarchical levels of tertiary, quaternary, and supra-quaternary structures for artificial protein complexes. As shown in Figure 4-1, (1) the PN-Block strategy begins with choice of PN-Block components. We can use several types of PN-Block components: the intermolecularly folded dimeric *de novo* proteins, oligomeric proteins and domains, and various linkers. (2) Combination and fusion of PN-Block components create various PN-Blocks such as vertex–frame PN-Blocks (vPN-Blocks), extender PN-Blocks (ePN-Block), and stopper PN-Blocks (sPN-Blocks). (3) Combination and self-assembly of PN-Blocks make various PN-Block complexes, not only homooligomers but also heterooligomeric complexes. (4) Coordination and further self-assembly of PN-Block complexes produce PN-Block supra-quaternary structures as higher-order supramolecular nanoarchitectures, which reflect structural properties of PN-Blocks and PN-Block complexes. This general and systematic strategy for hierarchical design has highly compatible modularity based on the combination of PN-Blocks and the redesignability of the artificial protein components, expanding the possibilities of PN-Blocks as artificial building-block molecules in the field of nanotechnology and synthetic biology.

The PN-Block strategy can be further enhanced by adding cofactors and/or synthetic ligands such as metal-directed protein self-assemblies (Bailey et al., 2016; Brodin et al., 2012; Sontz et al., 2015), and using computational methods for protein design such as the Rosetta software suite for macromolecular modeling (Baker, 2014; Huang et al., 2016; Leaver-Fay et al., 2013). In addition, the PN-Block strategy coupled with evolutionary molecular engineering will open the door to developing self-assembling multivalent functional nanobiomolecules. Furthermore, the design of supra-quaternary structures of artificial protein complexes and protein crystals is a growing area of nanobiomaterial science and nanobiotechnology. Inter-complex interactions and three-dimensional ordering of protein complexes can be induced by precipitants and/or cofactors, leading to supra-quaternary structures of protein crystals (Lai et al., 2014) and protein‑metal–organic crystalline frameworks (Sontz et al., 2015) as fascinating new porous nanomaterials that allow precise arrangements of exogenous compounds using metal coordination and chemical conjugation in solvent channels of crystals (Abe et al., 2016; Abe and Ueno, 2015).

As a potential application, protein nanostructures produced by the PN-Block strategy are expected to be useful for drug delivery systems because the surface of nanostructures can be redesigned to have high affinity for a specific drug target by screening from combinatorial libraries of a *de novo* protein component. In addition, the highly ordered symmetric structure of the protein nanostructures constructed from the PN-Block can be applied to the development of a non-viral vaccine symmetrically presenting antigens. Moreover, utilizing internal space of nanostructures and lattice structures as an enzyme-like catalytic site for chemical reactions can lead to the development of molecular flask and supramolecular container technologies.

As one of designers and engineers of artificial protein complexes, I believe a bright research future of the PN-Block strategy in the field of nanobiotechnology, nanobiomaterials science and synthetic biology.

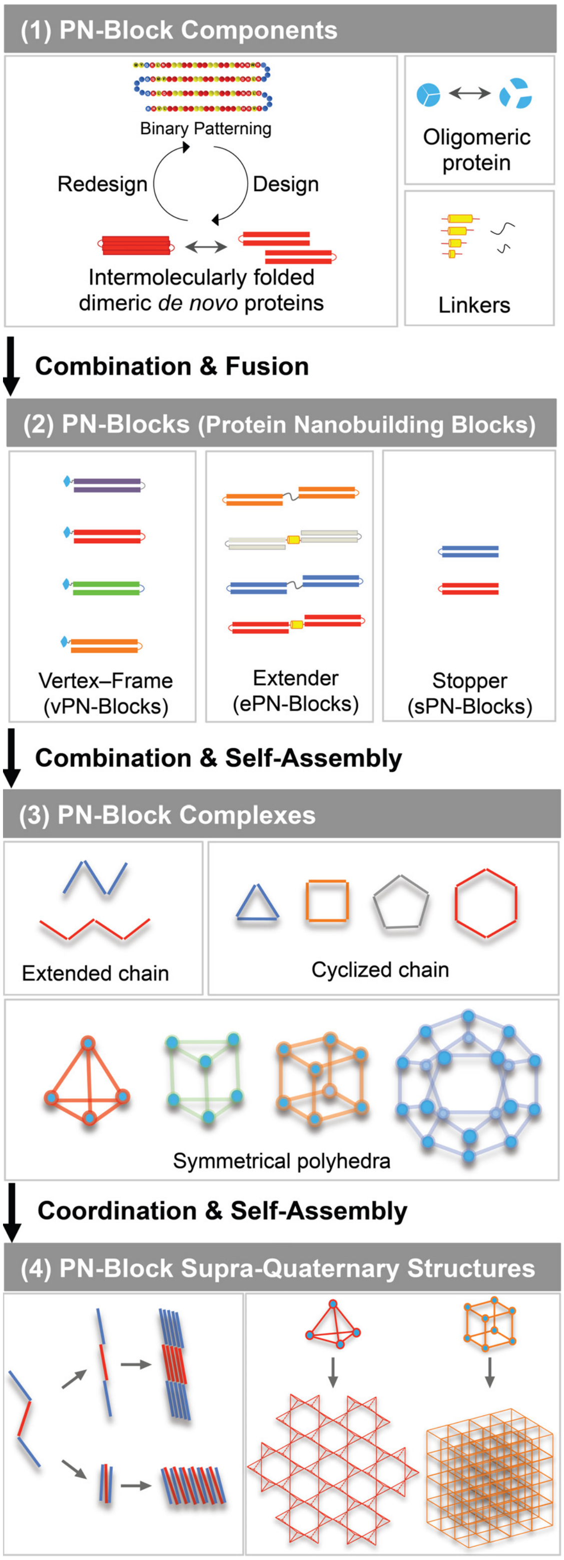

(1) PN-Block Components
Binary Patterning
Redesign
Design
Intermolecularly folded dimeric *de novo* proteins
Oligomeric protein
Linkers
Combination & Fusion
(2) PN-Blocks (Protein Nanobuilding Blocks)
Vertex–Frame (vPN-Blocks)
Extender (ePN-Blocks)
Stopper (sPN-Blocks)
Combination & Self-Assembly
(3) PN-Block Complexes
Extended chain
Cyclized chain
Symmetrical polyhedra
Coordination & Self-Assembly
(4) PN-Block Supra-Quaternary Structures

Figure 4-1. Hierarchical Design of Supramolecular Nanostructures by the PN-Block Strategy.

References

Abe, S., Maity, B., and Ueno, T. (2016). Design of a confined environment using protein cages and crystals for the development of biohybrid materials. Chem Commun *52*, 6496-6512.

Abe, S., and Ueno, T. (2015). Design of protein crystals in the development of solid biomaterials. RSC advances *5*, 21366-21375.

Ahnert, S.E., Marsh, J.A., Hernandez, H., Robinson, C.V., and Teichmann, S.A. (2015). Principles of assembly reveal a periodic table of protein complexes. Science *350*, aaa2245-aaa2245.

Arai, R., Kobayashi, N., Kimura, A., Sato, T., Matsuo, K., Wang, A.F., Platt, J.M., Bradley, L.H., and Hecht, M.H. (2012). Domain-swapped dimeric structure of a stable and functional *de novo* four-helix bundle protein, WA20. J Phys Chem B *116*, 6789-6797.

Arai, R., Ueda, H., Kitayama, A., Kamiya, N., and Nagamune, T. (2001). Design of the linkers which effectively separate domains of a bifunctional fusion protein. Protein Eng, Des Sel *14*, 529-532.

Arai, R., Wriggers, W., Nishikawa, Y., Nagamune, T., and Fujisawa, T. (2004). Conformations of variably linked chimeric proteins evaluated by synchrotron X-ray small-angle scattering. Proteins *57*, 829-838.

Armstrong, C.T., Boyle, A.L., Bromley, E.H.C., Mahmoud, Z.N., Smith, L., Thomson, A.R., and Woolfson, D.N. (2009). Rational design of peptide-based building blocks for nanoscience and synthetic biology. Faraday Discuss *143*, 305-317.

Bai, Y., Luo, Q., Zhang, W., Miao, L., Xu, J., Li, H., and Liu, J. (2013). Highly Ordered Protein Nanorings Designed by Accurate Control of Glutathione S-Transferase Self-Assembly. J Am Chem Soc *135*, 10966-10969.

Bailey, J.B., Subramanian, R.H., Churchfield, L.A., and Tezcan, F.A. (2016). Metal-Directed Design of Supramolecular Protein Assemblies. Methods Enzymol *580*, 223-250.

Baker, D. (2014). Centenary Award and Sir Frederick Gowland Hopkins Memorial Lecture. Protein folding, structure prediction and design. Biochem Soc Trans *42*, 225-229.

Bale, J.B., Gonen, S., Liu, Y., Sheffler, W., Ellis, D., Thomas, C., Cascio, D., Yeates, T.O., Gonen, T., King, N.P.*, et al.* (2016). Accurate design of megadalton-scale two-component icosahedral protein complexes. Science *353*, 389-394.

Beneš, P., Chovancová, E., Kozlíková, B., Pavelka, A., Strnad, O., Brezovský, J., Šustr, V., Klvaňa, M., Szabó, T., Gora, A.*, et al.* (2010). CAVER 2.1. In Software.

Bennett, M.J., Schlunegger, M.P., and Eisenberg, D. (1995). 3D domain swapping: a mechanism for oligomer assembly. Protein Sci *4*, 2455-2468.

Bennett, M.J.C., S.; Eisenberg, D. (1994). Domain swapping- entangling alliances between proteins. Proc Natl Acad Sci USA *91*, 3127–3131.

Boyken, S.E., Chen, Z., Groves, B., Langan, R.A., Oberdorfer, G., Ford, A., Gilmore, J.M., Xu, C., DiMaio, F., Pereira, J.H.*, et al.* (2016). *De novo* design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. Science *352*, 680-687.

Boyle, A.L., Bromley, E.H., Bartlett, G.J., Sessions, R.B., Sharp, T.H., Williams, C.L., Curmi, P.M., Forde, N.R., Linke, H., and Woolfson, D.N. (2012). Squaring the circle in peptide assembly: from fibers to discrete nanostructures by de novo design. J Am Chem Soc *134*, 15457-15467.

Bozic, S., Doles, T., Gradisar, H., and Jerala, R. (2013). New designed protein assemblies. Curr Opin Chem Biol *17*, 940-945.

Bradley, L.H., Kleiner, R.E., Wang, A.F., Hecht, M.H., and Wood, D.W. (2005). An intein-based genetic selection allows the construction of a high-quality library of binary patterned *de novo* protein sequences. Protein Eng, Des Sel *18*, 201-207.

Brodin, J.D., Ambroggio, X.I., Tang, C., Parent, K.N., Baker, T.S., and Tezcan, F.A. (2012). Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. Nat Chem *4*, 375-382.

Brodin, J.D., Smith, S.J., Carr, J.R., and Tezcan, F.A. (2015). Designed, Helical Protein Nanotubes with Variable Diameters from a Single Building Block. J Am Chem Soc *137*, 10468-10471.

Brunette, T.J., Parmeggiani, F., Huang, P.-S., Bhabha, G., Ekiert, D.C., Tsutakawa, S.E., Hura, G.L., Tainer, J.A., and Baker, D. (2015). Exploring the repeat protein universe through computational protein design. Nature *528*, 580-584.

Brunner-Popela, J., and Glatter, O. (1997). Small-Angle Scattering of Interacting Particles. I. Basic Principles of a Global Evaluation Technique. J Appl Crystallogr *30*, 431-442.

Burgess, N.C., Sharp, T.H., Thomas, F., Wood, C.W., Thomson, A.R., Zaccai, N.R., Brady, R.L., Serpell, L.C., and Woolfson, D.N. (2015). Modular Design of Self-Assembling Peptide-Based Nanotubes. J Am Chem Soc *137*, 10554-10562.

Burton, A.J., Thomson, A.R., Dawson, W.M., Brady, R.L., and Woolfson, D.N. (2016). Installing hydrolytic activity into a completely *de novo* protein framework. Nat Chem *8*, 837-844.

Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr, Sect D *66*, 12-21.

Cherny, I., Korolev, M., Koehler, A.N., and Hecht, M.H. (2012). Proteins from an Unevolved Library of *de novo* Designed Sequences Bind a Range of Small Molecules. ACS synthetic biology *1*, 130-138.

Churchfield, L.A., Medina-Morales, A., Brodin, J.D., Perez, A., and Tezcan, F.A. (2016). *De Novo* Design of an Allosteric Metalloprotein Assembly with Strained Disulfide Bonds. J Am Chem Soc *138*, 13163-13166.

Collaborative Computational Project (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr, Sect D *50*, 760-763.

Crick, F.H.C. (1953). The packing of α-helices: simple coiled-coils. Acta Crystallographica *6*, 689-697.

Dahiyat, B.I., and Mayo, S.L. (1997). De novo protein design: fully automated sequence selection. Science *278*, 82-87.

Das, A., Wei, Y., Pelczer, I., and Hecht, M.H. (2011). Binding of small molecules to cavity forming mutants of a *de novo* designed protein. Protein Sci *20*, 702-711.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S*., et al.* (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res *35*, W375-383.

Digianantonio, K.M., and Hecht, M.H. (2016). A protein constructed *de novo* enables cell growth by altering gene regulation. Proc Natl Acad Sci USA *113*, 2400-2405.

DiMaio, F., Leaver-Fay, A., Bradley, P., Baker, D., and André, I. (2011). Modeling Symmetric Macromolecular Structures in Rosetta3. PloS one *6*, e20450.

Doyle, L., Hallinan, J., Bolduc, J., Parmeggiani, F., Baker, D., Stoddard, B.L., and

Bradley, P. (2015). Rational design of α-helical tandem repeat proteins with closed architectures. Nature *528*, 585-588.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr, Sect D *66*, 486-501.

Endo, M., Yang, Y., and Sugiyama, H. (2013). DNA origami technology for biomaterials applications. Biomater Sci *1*, 347-360.

Finzel, B.C., Weber, P.C., Hardman, K.D., and Salemme, F.R. (1985). Structure of ferricytochrome *c′* from Rhodospirillum molischianum at 1.67 Å resolution. J Mol Biol *186*, 627-643.

Fisher, M.A., McKinley, K.L., Bradley, L.H., Viola, S.R., and Hecht, M.H. (2011). *De novo* designed proteins from a library of artificial sequences function in *Escherichia coli* and enable cell growth. PloS one *6*, e15364.

Fletcher, J.M., Harniman, R.L., Barnes, F.R., Boyle, A.L., Collins, A., Mantell, J., Sharp, T.H., Antognozzi, M., Booth, P.J., Linden, N.*, et al.* (2013). Self-assembling cages from coiled-coil peptide modules. Science *340*, 595-599.

Fujinaga, M., Cherney, M.M., Oyama, H., Oda, K., and James, M.N. (2004). The molecular structure and catalytic mechanism of a novel carboxyl peptidase from *Scytalidium lignicolum*. Proc Natl Acad Sci USA *101*, 3364-3369.

Fujita, D., Suzuki, K., Sato, S., Yagi-Utsumi, M., Yamaguchi, Y., Mizuno, N., Kumasaka, T., Takata, M., Noda, M., Uchiyama, S.*, et al.* (2012). Protein encapsulation within synthetic molecular hosts. Nat commun *3*, 1093.

Fukuma, T., Kimura, M., Kobayashi, K., Matsushige, K., and Yamada, H. (2005). Development of low noise cantilever deflection sensor for multienvironment frequency-modulation atomic force microscopy. Rev Sci Instrum *76*, 053704.

Glatter, O. (1980a). Computation of distance distribution functions and scattering functions of models for small angle scattering experiments. Acta Phys Austrica *52*, 243-256.

Glatter, O. (1980b). Evaluation of small-angle scattering data from lamellar and cylindrical particles by the indirect Fourier transformation method. J Appl Cryst *13*, 577-584.

Glatter, O., and Kratky, O. (1982). Small-Angle X-Ray Scattering (New York: Academic Press).

Glykos, N.M., Cesareni, G., and Kokkinidis, M. (1999). Protein plasticity to the extreme: changing the topology of a 4-alpha-helical bundle with a single amino acid substitution. Structure *7*, 597-603.

Go, A., Kim, S., Baum, J., and Hecht, M.H. (2008). Structure and dynamics of de novo proteins from a designed superfamily of 4-helix bundles. Protein Sci *17*, 821-832.

Gonen, S., DiMaio, F., Gonen, T., and Baker, D. (2015). Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. Science *348*, 1365-1368.

Gradisar, H., Bozic, S., Doles, T., Vengust, D., Hafner-Bratkovic, I., Mertelj, A., Webb, B., Sali, A., Klavzar, S., and Jerala, R. (2013). Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. Nat Chem Biol *9*, 362-366.

Grigoryan, G., Kim, Y.H., Acharya, R., Axelrod, K., Jain, R.M., Willis, L., Drndic, M., Kikkawa, J.M., and DeGrado, W.F. (2011). Computational design of virus-like protein assemblies on carbon nanotube surfaces. Science *332*, 1071-1076.

Guthe, S., Kapinos, L., Moglich, A., Meier, S., Grzesiek, S., and Kiefhaber, T. (2004). Very fast folding and association of a trimerization domain from bacteriophage T4 fibritin. J Mol Biol *337*, 905-915.

Harding, M.M. (2002). Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium. Acta Crystallogr, Sect D *58*, 872-874.

Hecht, M.H., Das, A., Go, A., Bradley, L.H., and Wei, Y. (2004). De novo proteins from designed combinatorial libraries. Protein Sci *13*, 1711-1723.

Hendlich, M., Rippmann, F., and Barnickel, G. (1997). LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. J Mol Graph Model *15*, 359-363.

Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science *254*, 51-58.

Hirota, S., Hattori, Y., Nagao, S., Taketa, M., Komori, H., Kamikubo, H., Wang, Z., Takahashi, I., Negi, S., Sugiura, Y.*, et al.* (2010). Cytochrome c polymerization by successive domain swapping at the C-terminal helix. Proc Natl Acad Sci USA *107*, 12854-12859.

Hoegler, K.J., and Hecht, M.H. (2016). A *De Novo* Protein Confers Copper Resistance in *Escherichia Coli*. Protein Sci *25*, 1249-1259.

Hsia, Y., Bale, J.B., Gonen, S., Shi, D., Sheffler, W., Fong, K.K., Nattermann, U., Xu, C., Huang, P.S., Ravichandran, R.*, et al.* (2016). Design of a hyperstable 60-subunit protein icosahedron. Nature *535*, 136-139.

Huang, P.-S., Feldmeier, K., Parmeggiani, F., Fernandez Velasco, D.A., Höcker, B., and Baker, D. (2015). *De novo* design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. Nat Chem Biol *12*, 29-34.

Huang, P.S., Boyken, S.E., and Baker, D. (2016). The coming of age of *de novo* protein design. Nature *537*, 320-327.

Huang, P.S., Oberdorfer, G., Xu, C., Pei, X.Y., Nannenga, B.L., Rogers, J.M., DiMaio, F., Gonen, T., Luisi, B., and Baker, D. (2014). High thermodynamic stability of parametrically designed helical bundles. Science *346*, 481-485.

Huard, D.J.E., Kane, K.M., and Tezcan, F.A. (2013). Re-engineering protein interfaces yields copper-inducible ferritin cage assembly. Nat Chem Biol *9*, 169-176.

Igarashi, N., Watanabe, Y., Shinohara, Y., Inoko, Y., Matsuba, G., Okuda, H., Mori, T., and Ito, K. (2011). Upgrade of the small angle X-ray scattering beamlines at the Photon Factory. J Phys Conf Ser *272*, 012026.

Joh, N.H., Wang, T., Bhate, M.P., Acharya, R., Wu, Y., Grabe, M., Hong, M., Grigoryan, G., and DeGrado, W.F. (2014). *De novo* design of a transmembrane Zn^{2+}-transporting four-helix bundle. Science *346*, 1520-1524.

Johnson, B.H., and Hecht, M.H. (1994). Recombinant proteins can be isolated from *E. coli* cells by repeated cycles of freezing and thawing. Biotechnology (NY) *12*, 1357-1360.

Kamtekar, S., Schiffer, J., Xiong, H., Babik, J., and Hecht, M. (1993). Protein design by binary patterning of polar and nonpolar amino acids. Science *262*, 1680-1685.

Ke, Y. (2014). Designer three-dimensional DNA architectures. Curr Opin Struct Biol *27*, 122-128.

Keefe, A.D., and Szostak, J.W. (2001). Functional proteins from a random-sequence library. Nature *410*, 715-718.

King, N.P., Bale, J.B., Sheffler, W., McNamara, D.E., Gonen, S., Gonen, T., Yeates, T.O., and Baker, D. (2014). Accurate design of co-assembling multi-component protein nanomaterials. Nature *510*, 103-108.

King, N.P., and Lai, Y.T. (2013). Practical approaches to designing novel protein assemblies. Curr Opin Struct Biol *23*, 632-638.

King, N.P., Sheffler, W., Sawaya, M.R., Vollmar, B.S., Sumida, J.P., Andre, I., Gonen, T., Yeates, T.O., and Baker, D. (2012). Computational design of self-assembling protein nanomaterials with atomic level accuracy. Science *336*, 1171-1174.

Kobayashi, N., Yanase, K., Sato, T., Unzai, S., Hecht, M.H., and Arai, R. (2015). Self-Assembling Nano-Architectures Created from a Protein Nano-Building Block Using an Intermolecularly Folded Dimeric *de Novo* Protein. J Am Chem Soc *137*, 11285-11293.

Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T.B., Montelione, G.T., and Baker, D. (2012). Principles for designing ideal protein structures. Nature *491*, 222-227.

Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L., and Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. Science *302*, 1364-1368.

Lai, Y.T., Cascio, D., and Yeates, T.O. (2012a). Structure of a 16-nm cage designed by using protein oligomers. Science *336*, 1129.

Lai, Y.T., King, N.P., and Yeates, T.O. (2012b). Principles for designing ordered protein assemblies. Trends Cell Biol *22*, 653-661.

Lai, Y.T., Reading, E., Hura, G.L., Tsai, K.L., Laganowsky, A., Asturias, F.J., Tainer, J.A., Robinson, C.V., and Yeates, T.O. (2014). Structure of a designed protein cage that self-assembles into a highly porous cube. Nat Chem *6*, 1065-1071.

Lai, Y.T., Tsai, K.L., Sawaya, M.R., Asturias, F.J., and Yeates, T.O. (2013). Structure and flexibility of nanoscale protein cages designed by symmetric self-assembly. J Am Chem Soc *135*, 7738-7743.

Lanci, C.J., MacDermaid, C.M., Kang, S.G., Acharya, R., North, B., Yang, X., Qiu, X.J., DeGrado, W.F., and Saven, J.G. (2012). Computational design of a protein crystal. Proc Natl Acad Sci USA *109*, 7304-7309.

Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993). Procheck - a Program to Check the Stereochemical Quality of Protein Structures. J Appl Crystallogr *26*, 283-291.

Laue, T.M., Shah, B.D., Ridgeway, T.M., and Pelletier, S.L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In Analytical Ultracentrifugation in Biochemistry and Polymer Science, S.E. Harding, A.J. Rowe, and J.C. Horton, eds. (Cambridge, UK: Royal Society of Chemistry), pp. 90-125.

Leaver-Fay, A., O'Meara, M.J., Tyka, M., Jacak, R., Song, Y., Kellogg, E.H., Thompson, J., Davis, I.W., Pache, R.A., Lyskov, S., *et al.* (2013). Scientific benchmarks for guiding macromolecular energy function improvement. Methods

Enzymol *523*, 109-143.

Lederer, F., Glatigny, A., Bethge, P.H., Bellamy, H.D., and Matthew, F.S. (1981). Improvement of the 2.5 Å resolution model of cytochrome b_{562} by redetermining the primary structure and using molecular graphics. J Mol Biol *148*, 427-448.

LeMaster, D.M., and Richards, F.M. (1985). ^{1}H-^{15}N heteronuclear NMR studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type. Biochemistry *24*, 7263-7268.

Lin, Y.-R., Koga, N., Tatsumi-Koga, R., Liu, G., Clouser, A.F., Montelione, G.T., and Baker, D. (2015a). Control over overall shape and size in *de novo* designed proteins. Proc Natl Acad Sci USA *112*, E5478-E5485.

Lin, Y.W., Nagao, S., Zhang, M., Shomura, Y., Higuchi, Y., and Hirota, S. (2015b). Rational design of heterodimeric protein using domain swapping for myoglobin. Angew Chem Int Ed Engl *54*, 511-515.

Liu, Y., and Eisenberg, D. (2002). 3D domain swapping: as domains continue to swap. Protein Sci *11*, 1285-1299.

Lovell, S.C., Davis, I.W., Arendall, W.B., 3rd, de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. (2003). Structure validation by C_{α} geometry: phi, psi and C_{β} deviation. Proteins *50*, 437-450.

Luo, Q., Hou, C., Bai, Y., Wang, R., and Liu, J. (2016). Protein Assembly: Versatile Approaches to Construct Highly Ordered Nanostructures. Chem Rev *116*, 13571–13632.

Marqusee, S., and Baldwin, R.L. (1987). Helix stabilization by Glu^{-}...Lys^{+} salt bridges in short peptides of *de novo* design. Proc Natl Acad Sci USA *84*, 8898-8902.

Mason, J.M., and Arndt, K.M. (2004). Coiled coil domains: stability, specificity, and biological implications. Chembiochem *5*, 170-176.

Medina-Morales, A., Perez, A., Brodin, J.D., and Tezcan, F.A. (2013). *In Vitro* and Cellular Self-Assembly of a Zn-Binding Protein Cryptand via Templated Disulfide Bonds. J Am Chem Soc *135*, 12013-12022.

Miyamoto, T., Kuribayashi, M., Nagao, S., Shomura, Y., Higuchi, Y., and Hirota, S. (2015). Domain-swapped cytochrome cb_{562} dimer and its nanocage encapsulating a Zn–SO_4cluster in the internal cavity. Chem Sci *6*, 7336-7342.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr, Sect D *67*, 355-367.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr, Sect D *53*, 240-255.

Niesen, F.H., Berglund, H., and Vedadi, M. (2007). The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc *2*, 2212-2221.

Ogihara, N.L., Ghirlanda, G., Bryson, J.W., Gingery, M., DeGrado, W.F., and Eisenberg, D. (2001). Design of three-dimensional domain-swapped dimers and fibrous oligomers. Proc Natl Acad Sci USA *98*, 1404-1409.

Ortega, A., Amoros, D., and Garcia de la Torre, J. (2011). Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. Biophys J *101*, 892-898.

Orthaber, D., Bergmann, A., and Glatter, O. (2000). SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. J Appl Crystallogr *33*, 218-225.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol *276*, 307-326.

Pace, C.N., and Scholtz, J.M. (1998). A helix propensity scale based on experimental studies of peptides and proteins. Biophys J *75*, 422-427.

Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995). How to Measure and Predict the Molar Absorption-Coefficient of a Protein. Protein Sci *4*, 2411-2423.

Padilla, J.E., Colovos, C., and Yeates, T.O. (2001). Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. Proc Natl Acad Sci USA *98*, 2217-2221.

Pandya, M.J., Spooner, G.M., Sunde, M., Thorpe, J.R., Rodger, A., and Woolfson, D.N. (2000). Sticky-end assembly of a designed peptide fiber provides insight into protein fibrillogenesis. Biochemistry *39*, 8728-8734.

Papapostolou, D., Smith, A.M., Atkins, E.D., Oliver, S.J., Ryadnov, M.G., Serpell, L.C., and Woolfson, D.N. (2007). Engineering nanoscale order into a designed protein fiber. Proc Natl Acad Sci USA *104*, 10853-10858.

Park, K., Shen, B.W., Parmeggiani, F., Huang, P.-S., Stoddard, B.L., and Baker, D. (2015). Control of repeat-protein curvature by computational protein design. Nat Struct

Mol Biol *22*, 167-174.

Patel, S.C., Bradley, L.H., Jinadasa, S.P., and Hecht, M.H. (2009). Cofactor binding and enzymatic activity in an unevolved superfamily of *de novo* designed 4-helix bundle proteins. Protein Sci *18*, 1388-1400.

Patel, S.C., and Hecht, M.H. (2012). Directed evolution of the peroxidase activity of a *de novo*-designed protein. Protein Eng, Des Sel *25*, 445-452.

Peters, K., Hinz, H.J., and Cesareni, G. (1997). Introduction of a proline residue into position 31 of the loop of the dimeric 4-alpha-helical protein ROP causes a drastic destabilization. Biol Chem *378*, 1141-1152.

Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., Gorba, C., Mertens, H.D.T., Konarev, P.V., and Svergun, D.I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. J Appl Crystallogr *45*, 342-350.

Pieters, B.J., van Eldijk, M.B., Nolte, R.J., and Mecinovic, J. (2016). Natural supramolecular protein assemblies. Chem Soc Rev *45*, 24-39.

Platt, J.M. (2007). Biophysical Characterization of *de novo* Four Helix Bundle Proteins. Senior Thesis, Princeton University, NJ,.

Radford, R.J., Brodin, J.D., Salgado, E.N., and Tezcan, F.A. (2011). Expanding the utility of proteins as platforms for coordination chemistry. Coordin Chem Rev *255*, 790-803.

Rämisch, S., Weininger, U., Martinsson, J., Akke, M., and André, I. (2014). Computational design of a leucine-rich repeat protein with a predefined geometry. Proc Natl Acad Sci USA *111*, 17875-17880.

Reedy, C.J., and Gibney, B.R. (2004). Heme protein assemblies. Chem Rev *104*, 617-649.

Rothemund, P.W. (2006). Folding DNA to create nanoscale shapes and patterns. Nature *440*, 297-302.

Roy, S., and Hecht, M.H. (2000). Cooperative thermal denaturation of proteins designed by binary patterning of polar and nonpolar amino acids. Biochemistry *39*, 4603-4607.

Salgado, E.N., Faraone-Mennella, J., and Tezcan, F.A. (2007). Controlling protein-protein interactions through metal coordination: assembly of a 16-helix bundle protein. J Am Chem Soc *129*, 13374-13375.

Salgado, E.N., Radford, R.J., and Tezcan, F.A. (2010). Metal-directed protein self-assembly. Acc Chem Res *43*, 661-672.

Sato, T., Shimozawa, T., Fukasawa, T., Ohtaki, M., Aramaki, K., Wakabayashi, K., and Ishiwata, S.i. (2010). Actin oligomers at the initial stage of polymerization induced by increasing temperature at low ionic strength: Study with small-angle X-ray scattering. Biophysics *6*, 1-11.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J *78*, 1606-1619.

Sciore, A., Su, M., Koldewey, P., Eschweiler, J.D., Diffley, K.A., Linhares, B.M., Ruotolo, B.T., Bardwell, J.C., Skiniotis, G., and Marsh, E.N. (2016). Flexible, symmetry-directed approach to assembling protein cages. Proc Natl Acad Sci USA *113*, 8681-8686.

Seeman, N.C. (2003). DNA in a material world. Nature *421*, 427-431.

Sharp, T.H., Bruning, M., Mantell, J., Sessions, R.B., Thomson, A.R., Zaccai, N.R., Brady, R.L., Verkade, P., and Woolfson, D.N. (2012). Cryo-transmission electron microscopy structure of a gigadalton peptide fiber of *de novo* design. Proc Natl Acad Sci USA *109*, 13266-13271.

Sinclair, J.C., Davies, K.M., Venien-Bryan, C., and Noble, M.E. (2011). Generation of protein lattices by fusing proteins with matching rotational symmetry. Nat Nanotechnol *6*, 558-562.

Smith, B.A., Mularz, A.E., and Hecht, M.H. (2015). Divergent evolution of a bifunctional *de novo* protein. Protein Sci *24*, 246-252.

Song, W.J., and Tezcan, F.A. (2014). A designed supramolecular protein assembly with in vivo enzymatic activity. Science *346*, 1525-1528.

Sontz, P.A., Bailey, J.B., Ahn, S., and Tezcan, F.A. (2015). A Metal Organic Framework with Spherical Protein Nodes: Rational Chemical Design of 3D Protein Crystals. J Am Chem Soc *137*, 11598-11601.

Sontz, P.A., Song, W.J., and Tezcan, F.A. (2014). Interfacial metal coordination in engineered protein and peptide assemblies. Curr Opin Chem Biol *19*, 42-49.

Suzuki, Y., Cardone, G., Restrepo, D., Zavattieri, P.D., Baker, T.S., and Tezcan, F.A. (2016). Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. Nature *533*, 369-373.

Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J *76*, 2879-2886.

Tanford, C. (1961). Physical chemistry of macromolecules (New York: Wiley).

Tao, Y., Strelkov, S.V., Mesyanzhinov, V.V., and Rossmann, M.G. (1997). Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. Structure *5*, 789–798.

Terwilliger, T.C. (2002). Automated structure solution, density modification and model building. Acta Crystallogr, Sect D *58*, 1937-1940.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr, Sect D *55*, 849-861.

Thomas, F., Burgess, N.C., Thomson, A.R., and Woolfson, D.N. (2016). Controlling the Assembly of Coiled-Coil Peptide Nanotubes. Angew Chem Int Ed Engl *55*, 987-991.

Thomson, A.R., Wood, C.W., Burton, A.J., Bartlett, G.J., Sessions, R.B., Brady, R.L., and Woolfson, D.N. (2014). Computational design of water-soluble α-helical barrels. Science *346*, 485-488.

Urvoas, A., Valerio-Lepiniec, M., and Minard, P. (2012). Artificial proteins from combinatorial approaches. Trends Biotechnol *30*, 512-520.

Vedadi, M., Niesen, F.H., Allali-Hassani, A., Fedorov, O.Y., Finerty, P.J., Jr., Wasney, G.A., Yeung, R., Arrowsmith, C., Ball, L.J., Berglund, H., *et al.* (2006). Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. Proc Natl Acad Sci USA *103*, 15835-15840.

Vistica, J., Dam, J., Balbo, A., Yikilmaz, E., Mariuzza, R.A., Rouault, T.A., and Schuck, P. (2004). Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. Anal Biochem *326*, 234-256.

Voet, A.R.D., Noguchi, H., Addy, C., Simoncini, D., Terada, D., Unzai, S., Park, S.Y., Zhang, K.Y.J., and Tame, J.R.H. (2014). Computational design of a self-assembling symmetrical beta-propeller protein. Proc Natl Acad Sci USA *111*, 15102-15107.

Voet, A.R.D., Noguchi, H., Addy, C., Zhang, K.Y.J., and Tame, J.R.H. (2015). Biomineralization of a Cadmium Chloride Nanocrystal by a Designed Symmetrical Protein. Angew Chem Int Ed Engl *54*, 9857-9860.

Wallace, A.C., Laskowski, R.A., and Thornton, J.M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng, Des Sel *8*, 127-134.

Wang, A.F. (2006). Sequence, Expression, and Structural Characterization of a High-Quality de novo Combinatorial Four-Helix Bundle Library. Senior Thesis, Princeton University, NJ,.

Wei, Y., Kim, S., Fela, D., Baum, J., and Hecht, M.H. (2003a). Solution structure of a *de novo* protein from a designed combinatorial library. Proc Natl Acad Sci USA *100*, 13270-13273.

Wei, Y., Liu, T., Sazinsky, S.L., Moffet, D.A., Pelczer, I., and Hecht, M.H. (2003b). Stably folded *de novo* proteins from a designed combinatorial library. Protein Sci *12*, 92-102.

West, M.W., Wang, W., Patterson, J., Mancias, J.D., Beasley, J.R., and Hecht, M.H. (1999). *De novo* amyloid proteins from designed combinatorial libraries. Proc Natl Acad Sci USA *96*, 11211-11216.

Woolfson, D.N., Bartlett, G.J., Bruning, M., and Thomson, A.R. (2012). New currency for old rope: from coiled-coil assemblies to alpha-helical barrels. Curr Opin Struct Biol *22*, 432-441.

Woolfson, D.N., Bartlett, G.J., Burton, A.J., Heal, J.W., Niitsu, A., Thomson, A.R., and Wood, C.W. (2015). *De novo* protein design: how do we expand into the universe of possible protein structures? Curr Opin Struct Biol *33*, 16-26.

Wyatt, P.J. (1993). Light-Scattering and the Absolute Characterization of Macromolecules. Anal Chim Acta *272*, 1-40.

Yeates, T.O., Liu, Y., and Laniado, J. (2016). The design of symmetric protein nanomaterials comes of age in theory and practice. Curr Opin Struct Biol *39*, 134-143.

Yokoi, N., Inaba, H., Terauchi, M., Stieg, A.Z., Sanghamitra, N.J., Koshiyama, T., Yutani, K., Kanamaru, S., Arisaka, F., Hikage, T., *et al.* (2010). Construction of robust bio-nanotubes using the controlled self-assembly of component proteins of bacteriophage T4. Small *6*, 1873-1879.

List of Publications

Ryoichi Arai, **Naoya Kobayashi**, Akiho Kimura, Takaaki Sato, Kyoko Matsuo, Anna F. Wang, Josse M. Platt, Luke H. Bradley, Michael H. Hecht, Domain-Swapped Dimeric Structure of a Stable and Functional *De Novo* Four-Helix Bundle Protein, WA20. *J. Phys. Chem. B* **116**, 6789–6797 (2012).
(Reprinted with permission. Copyright (2012) American Chemical Society.)

Naoya Kobayashi, Keiichi Yanase, Takaaki Sato, Satoru Unzai, Michael H. Hecht, Ryoichi Arai. Self-Assembling Nano-Architectures Created from a Protein Nano-Building Block Using an Intermolecularly Folded Dimeric *De novo* protein. *J. Am. Chem. Soc.* **137**, 11285–11293 (2015).
(Reprinted with permission. Copyright (2015) American Chemical Society.)

Acknowledgements

I would like to express my sincere gratitude to Prof. Ryoichi Arai at Shinshu University for his detailed comments, constructive advice, and important supports during these works. Also I would like to express my sincere gratitude to Prof. Nobuaki Hayashida at Shinshu University for helpful advice and supports.

I also wish to express my gratitude to Prof. Takaaki Sato, Mr. Keiichi Yanase, Mr. Kouichi Inano, Mr. Takuya Sengoku, and Prof. Hiroshi Ogasawara at Shinshu University, Prof. Satoru Unzai at Hosei University, Prof. Takeshi Fukuma and Mr. Keisuke Miyazawa at Kanazawa University, Prof. Michael H. Hecht at Princeton University, Prof. Nobuyasu Koga and Dr. Rie Koga at Institute for Molecular Science (IMS), Prof. Teruyuki Nagamune at the University of Tokyo, and Dr. Masayuki Hashimoto at National Cheng Kung University for their helpful advice and technical supports.

The reviewers of the thesis, Prof. Toshio Shida, Prof. Yasushi Tamada, Prof. Hiroshi Moriwaki at Shinshu University, and Dr. Shinya Honda at Advanced Industrial Science and Technology provided valuable advice that helped make this thesis more substantial.

This study was supported by Japan Society for the Promotion of Science (JSPS) Research Fellowships (DC2), JSPS KAKENHI Grant Numbers JP14J10185 and JP16H06837, and Joint Studies Program (2014–2016) of IMS.

Synchrotron X-ray experiments were performed at Photon Factory (PF), KEK, under the approval of PF program advisory committee (Proposal Nos. 2008G511, 2008G512, 2014G111, 2016G153, and 2016G606).

Finally, I would like to express my deepest appreciation to my parents for their numerous supports, and thank the members of the Arai lab, both past and present, for their help, advice, insights, and friendships.