論文 / 著書情報
Article / Book Information

題目(和文)	三本鎖β-ヘリックス蛋白質の構造・性状解析と人工三本鎖β-ヘリックスの均一性向上の試み
Title(English)	Structure and properties of three-stranded β-helical proteins and approaches to obtain homogeneous artificial three-stranded β-helical polypeptides
著者(和文)	内田一也
Author(English)	Kazuya Uchida
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第9485号, 授与年月日:2014年3月26日, 学位の種別:課程博士, 審査員:有坂　文雄,中村　聡,三原　久和,和地　正明,田口　英樹
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第9485号, Conferred date:2014/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	要約
Type(English)	Outline

Structure and properties of three-stranded β-helical proteins and approaches to obtain homogeneous artificial three-stranded β-helical polypeptides

A Doctoral Thesis

Submitted by

Kazuya Uchida

to the

Graduate school of Bioscience and Biotechnology

Tokyo Institute of Technology

March, 2014

Contents

[a] Chapter 2 includes the following article.
Kazuya Uchida, Petr G Leiman, Fumio Arisaka and Shuji Kanamaru
Structure and properties of the C-terminal β-helical domain of VgrG protein from *Escherichia coli* O157.
J Biochem first published online December 3, 2013 doi:10.1093/jb/mvt109
The definitive publisher-authenticated version is available online at: http://jb.oxfordjournals.org/content/early/2014/01/09/jb.mvt109.abstract.

Abbreviations

$(VXG)_6$	Val474-Leu521 of T4 gp5
$(VXG)_4$	Val490-Leu521 of T4 gp5
Å	Angstrom
AUC	analytical ultracentrifugation
b.p.	base pair
c(M)	molecular weight distribution function
c(s)	sedimentation coefficient distribution function
CD	circular dichroism
Cgp5C, C	Val522-Gly575 of T4 gp5
CHES	*N*-Cyclohexyl-2-aminoethanesulphonic acid
C_α	alpha-carbon
C_β	beta-carbon
Da	dalton
DNA	deoxyribonucleic acid
ds	double-stranded
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EV	sedimentation equilibrium
f	foldon
f/f_0	frictional ratio
far-UV	far-ultraviolet
gp	gene product
gp5C, C	C-terminal part of T4 gp5
Hcp	hemolysin-coregulated protein
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
his-tag	hexahistidine tag
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	Luria-Bertani
MES	2-(*N*-morpholino)ethanesulfonic acid
ml	milliliter
MME	monomethyl ether
MPD	2-methyl-2,4-pentanediol
mRNA	messenger RNA
Mw	molecular weight
nm	nanometer
OB fold	oligonucleotide/oligosaccharide-binding fold
OD_{660}	optical density at 660 nm
OM	outer membrane
PAAR	proline-alanine-alanine-arginine

PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
PG	peptidoglycan
PMSF	phenyl-methylsulfonyl fluoride
RHS	rearrangement hot spot
RMSD	root mean square deviation
S	sedimentation coefficient
SDS	sodium dodecyl sulfate
SE	sedimentation equilibrium
SEC	size-exclusion chromatography
SeMet	Seleno methionine
SlyD	sensitive to lysis D (prolyl cis-trans isomerase)
SV	sedimentation velocity
T6SS	Type 6 secretion system
TEV	tobacco etch virus

List of Figures

List of Tables

Abstract

The three-stranded β-helix is a highly regular and repeated structure and is extremely stable against heat and denaturing agents. It is thus suitable for the template in designing rod-shaped nanomaterial. In our laboratory, the artificial three-stranded β-helices have been developed by stacking a certain region of the three-stranded β-helix of gp5 from bacteriophage T4, which consists of 12 repeats of octapeptide VXGXXXXX (X: any amino acid). However, the engineered three-stranded β-helices tend to aggregate. To reduce the aggregation of the artificial three-stranded β-helices, deeper understanding of structural foundation of the three-stranded β-helix is required. In the past 10 years, a number of T4 gp5 homologs were found not only in bacteriophages, but also in the Type 6 secretion systems (T6SS) of Gram-negative bacteria. The putative three-stranded β-helical region of these proteins have distinctive features in amino acid sequence, *e.g.* another conserved motif (gp5 of KVP40 phage), different repeat sequence such as VGXXXXXX or no apparent repetitive sequence (VgrG proteins of T6SS). In the present study, in order to elucidate the relationship between folding and amino acid sequences of three-stranded β-helix, I examined the structure and properties of six VgrG proteins from pathogenic *E. coli* O157 (VgrG1, VgrG2 and VgrG3) and CFT073 (c3393, c1888 and c1883) and gp5 of bacteriophage KVP40. I also tried to obtain artificial three-stranded β-helix in trimeric form by the fusion of SlyD, and finally I was able to eliminate the partially unfolded artificial three-stranded β-helix by limited proteolysis.

VgrG1^{his}, VgrG2^{his}, VgrG3^{his} from *E. coli* O157 and c3393^{his} c1888^{his} and c1883^{his} from *E. coli* CFT073 were over-expressed in *E. coli* cells. They tend to become insoluble, but *E. coli* O157 VgrG1^{his} and *E. coli* CFT073 c3393 were partially soluble, which were enough to purify for further experiments. The sedimentation velocity analysis and far-UV CD analysis indicated that *E. coli* O157 VgrG1^{his} and *E. coli* CFT c3393^{his} are trimer in solution and rich in β-structure, which were in agreement with the previously proposed structure of VgrG proteins, namely T4 gp27-gp5-analogous structure without lysozyme domain.

The crystal structure of the C-terminal trypsin-resistant fragment of *E. coli* O157 VgrG1 (VgrG1C^{G561}) was determined at 1.95Å resolution by X-ray

crystallography. It forms a prism-shaped antiparallel three-stranded β-helix. The topology and structure of VgrG1C^{G561} were similar to the three-stranded β-helix of gp138 of phi92 phage, which indicates a possible evolutional relationship between them. The β-helix of VgrG1C^{G561} was stabilized by the hydrogen bonds and a salt bridge at the sharp kinks and the hydrophobic interaction at the center of β-helix. The intermolecular hydrogen bonds and hydrophobic interaction has been also observed in gp138 from phi92 phage thus they are considered to be conserved stabilization mechanism of intertwined three-stranded β-helix. VgrG1C^{G561} has a long loop of 17 amino acid residues between Asp570 and Ala586. I surmised that some VgrG proteins would have an enzymatic domain in this loop or a certain Type 6 secretion system protein would bind to the loop.

I also tried to determine the crystal structure of *E. coli* CFT c3393. The ^{his}c3393 was successfully crystallized and the X-ray diffraction data was collected at 3.35Å resolution. The phase was determined by molecular replacement but the search model needs to be improved for complete structure.

In order to identify the determinant of the topology of the three-stranded β-helix whether it takes intertwined β-helix or antiparallel three-stranded β-helices, structure-based sequence alignment of these structures were performed. It revealed the preferred and forbidden amino acid sequence for either antiparallel or intertwined three-stranded β-helix.

Gp5 of bacteriophage KVP40 is expected to have the longest three-stranded β-helix among the proteins with VXGXXXXX sequence. To examine its atomic structure by X-ray crystallography, I performed crystallization of KVP40 gp5. It was crystallized in two different forms, namely hexagonal-shape and rhombohedral-shape. The full X-ray diffraction data was collected from rhombohedral-shaped crystals at 2.59Å resolution, but the phase remains to be determined.

To obtain homogeneous artificial three-stranded β-helical polypeptides, I have tried to prevent unfavorable associations of β-helix by fusing a SlyD protein and a foldon domain of fibritin protein. It was found that the fused proteins has reduced amount of large aggregation, but complete prevention of oligomerization of the β-helix was not attained. Native PAGE indicated that the purified artificial three-stranded β-helical polypeptides contain partially unfolded species. It was also found that the

partially unfolded species could be eliminated by the limited trypsin digestion. Finally, I have succeeded in the purification of artificial three-stranded β-helix to near homogeneity, which was shown to be a dimer of β-helix by AUC. The purification method established in this study is expected to be applicable to longer β-helix constructs.

Chapter 1

General Introduction

1-1 Background

Proteins are synthesized at the ribosome through the process of so-called translation as a polypeptide chain based on the nucleotide sequence of the mRNA. However, the synthesized polypeptide chain acquires its function only after it is folded into a unique three-dimensional structure. The physical principle by which it is folded into a unique structure has been a central key question in protein science. In 1963, Anfinsen discovered that the protein folding is a spontaneous process which does not require any external factors, but only depends on the amino acid sequence. This meant that there must be a physical principle that governs the formulation of the three-dimensional structure based on the sequence. As a result of efforts of many investigations in this field, methods of predicting higher order structure based on the sequence have been developed including theoretical approaches based on the first principle and empirical methods based on the relationship between the determined three-dimensional structure in Protein Data Bank (PDB) and the amino acid sequence, or the combination of the two. Since the first structure determination of myoglobin by Kendrew in 1958 (*1*) and hemoglobin by Perutz in 1960 (*2*), a great number of protein structures have been determined and deposited in PDB. Today, the number of protein structure registered in PDB exceeds 88,000 and it is still growing exponentially. Although most of the newly determined structures have been attributed to known structure family, completely new folds are still being discovered.

1-2 The hierarchical structure of proteins

In order to understand the protein structure, it is convenient to dissect the structure into hierarchical order namely, primary (amino acid sequence), secondary (local regular structure formed by hydrogen bond between main chain carbonyl oxygen and amide nitrogen), tertiary (assembly of secondary structures) and quaternary (association of subunit proteins) structure. Primary structure is a sequence of 20 kinds of amino acids, and the number of the possible kinds of the primary structure of a polypeptides chain is huge; the possible number for a polypeptide chain consisting of N amino acids would be 20^N, assuming equal appearance of the 20 amino acids, which is indeed astronomical. However, the number of secondary structure is limited due to mainly the steric hindrance of the side chain β-carbon and the carbonyl oxygen in the peptide bonds; representative secondary structures are α-helix, β-sheet, turn and the remaining "random" structure.

α-helix is an abundant structure and can be found in many proteins either as a single helix, or two-, three- or four-stranded coiled coil. There is some variation of α-helix, namely, 3_{10} helix and the π-helix, but they are only found at the terminus or at some internal positions of α-helix and the appearance is limited. Another abundant helical structure is a three-stranded collagen helix.

β-structure is basically constituted of either parallel or anti-parallel β-sheet, but has been found to form a variety of higher order structure as described more in detail in the next section. I only point out here that β-sheet itself has variation in the number of β-strands, the length and the curvature of the sheet.

1-3 The variation in β-structure

As already mentioned, β-structure is either parallel or anti-parallel when dissected into a small local region. However, it constitutes a whole a variety of higher order structures, namely β-barrel, β-sandwich, β-spiral, β-propeller and β-helix.

β-barrel is a well-known β-structure, where a β-sheet twists and makes a closed structure (Figure 1-1 (A)). It is frequently found in the membrane proteins of mitochondria and the outer membrane of Gram-negative bacteria. It is also found in the structure of bacteriophage proteins as an OB-fold (oligonucleotide/oligosaccharide-binding fold) (Figure 1-4).

β-sandwich is formed by two stacked β-sheets (Figure 1-1 (B)). This fold is found in the Fab region of immunoglobulin thus it is also called as immunoglobulin fold. In addition to these classical β-structures, a number of new β-folds have been discovered in the last two decades.

β-spiral was first found in the fiber protein of adenovirus in 1999 (*3*). It is one of the most complex β-structure, which consists of repeated β-sheet structures consisting of six β-strands (two antiparallel β-strands constituted from three polypeptide strands) and the loops connecting between them (Figure 1-1 (C)).

β-propeller is found in the cytochrome cd_1-nitrate reductase, whose structure was determined in 1995 (*4*). β-propeller is assembled from 4 to 8 of slightly twisted intertwined β-sheets. Each β-sheet is formed by four β-strands. This fold is often found in the enzyme proteins and regulatory proteins (Figure 1-1 (D)).

The single stranded β-helix was first found in the pectate lyase from *Erwinia chrysanthemi* (*5*) and the alkaline protease from the *Pseudomonas aeruginosa* (*6*) almost at the same time in 1993. Single-stranded β-helices are formed by 2 to 4 parallel

β-sheets, which are formed by a single polypeptide chain folded into right-handed or left-handed helical structure (Figure 1-1 (E)).

The three-stranded β-helix was first found in the short tail fiber gp12 of T4 phage in 2001 (*7*) (Figure 1-1 (F)). It was subsequently found in the tail lysozyme, gp5, of T4 phage. The three-stranded β-helices have a triangular prism shape, which is formed by three antiparallel β-sheets (each β-sheet is formed by each polypeptide strand) or intertwined three parallel β-sheets (each β-sheet is formed by alternatively arranged three polypeptides) or the mixture of them. In the next section, I will focus on the three-stranded β-helix of T4 gp5.

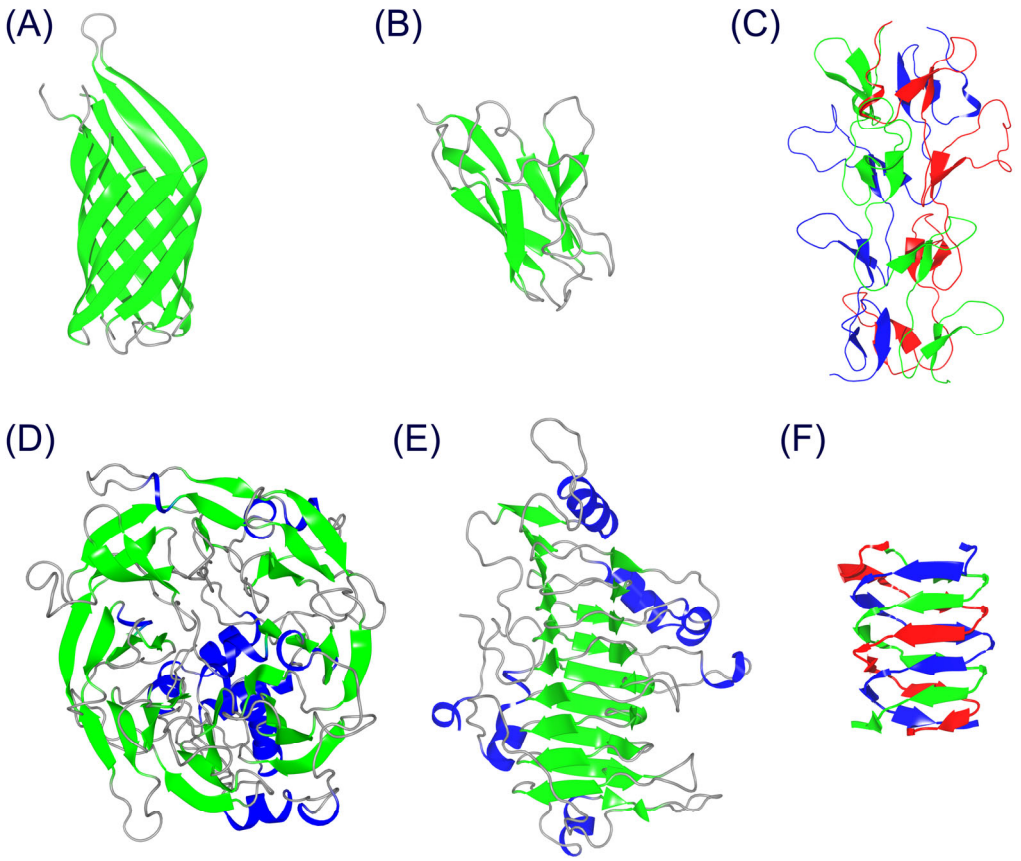

Figure 1-1. The variation of β-structure. (**A**) The β-barrel (OmpA from *Escherichia coli*: PDB 1QJP). (**B**) β-sandwich (Fab region of human antibody; residues 1-115 PDB 1DFB). (**C**) β-spiral (adenovirus fiber protein: residues 319-393, PDB 1QIU). (**D**) β-propeller (cytochrome *cd1*-nitrate reductase from *Paracoccus pantotrophus*: PDB 1AOF). (**E**) single-stranded β-helix (PelC from *Erwinia chrysanthemi*: PDB 1O8F). (**F**) three-stranded β-helix structure (gp12, short tail fiber of T4 phage: residues 290-327; PDB 1H6W).

1-4 The three-stranded β-helix of T4 phage gp5

Bacteriophage T4 is a virus which infects *Escherichia coli*. It injects the genome DNA into the host cell by the supramolecular structure called "tail". The tail of bacteriophage T4 is composed of two parts, namely, the co-cylinder structure and the baseplate at the distal end of it. The tail lysozyme gp5 (gene product 5) locates at the central tip of the baseplate in trimer (Figure 1-2). During the infection, it penetrates the outer membrane and locally degrades the peptidoglycan of the host cell (Figure 1-3).

The crystal structure of gp5 has been solved in complex with gp27 as a hetero hexamer ($(gp27)_3$-$(gp5)_3$) by Kanamaru et al. in 2002. T4 gp5 contains three domains, namely, the N-terminal OB-fold domain, the middle lysozyme domain and the C-terminal β-helix domain (gp5C) (Figure 1-4). The N-terminal OB-fold domain has a β-barrel fold consisting of five β-strands. The exact function of this domain is yet to be elucidated. The lysozyme domain is made of only α-helix. This domain contributes to the local degradation of the peptidoglycan. The C-terminal β-helix domain forms the three-stranded β-helix. This domain is thought to function as a needle to puncture the outer membrane.

The three-stranded β-helix of T4 gp5 can be divided into two parts, namely the β-prism made of three antiparallel β-sheets each constituted by five β-strands and an intertwined three-stranded β-helix which has a slightly twisted triangular prism-like shape with the dimensions of 30Å × 90Å. The latter part consists of intertwined three-stranded β-helix forming totally 18 steps. The intertwined region contains the 12 repeats of conserved eight amino acids VXGXXXXX where X indicates any amino acids. This peptide with eight amino acid residues forms each step of the triangular prism. The polypeptide sharply kinks at the conserved glycine residues and the side

chain of the other conserved valine residue fills the inner space at the kink. This repeated structure makes the three-stranded β-helix of T4 gp5 highly regular. This three-stranded β-helix is extraordinary stable and shows resistance to 10% SDS and 2M guanidine hydrochloride (*8*). This stability is considered to be needed for its penetration to the outer membrane of the host *E. coli* upon infection.

In the next section, I will introduce the application of the three-stranded β-helix of T4 phage to nanomaterials.

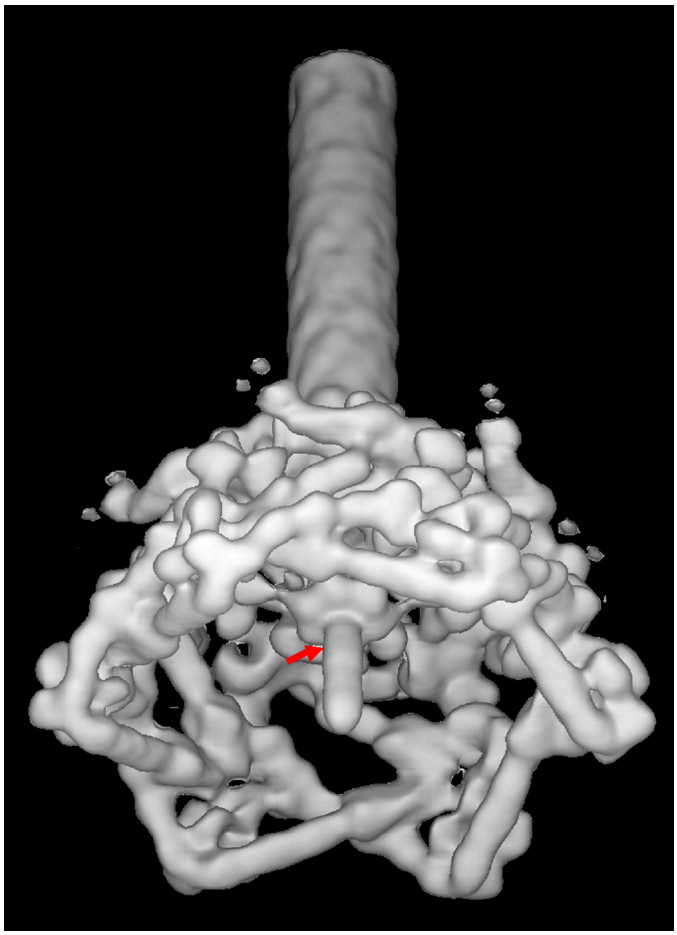

Figure 1-2. The location of gp5 in the tail of T4 phage. Red arrow indicates the position of gp5. Courtesy of Professor Petr Leiman.

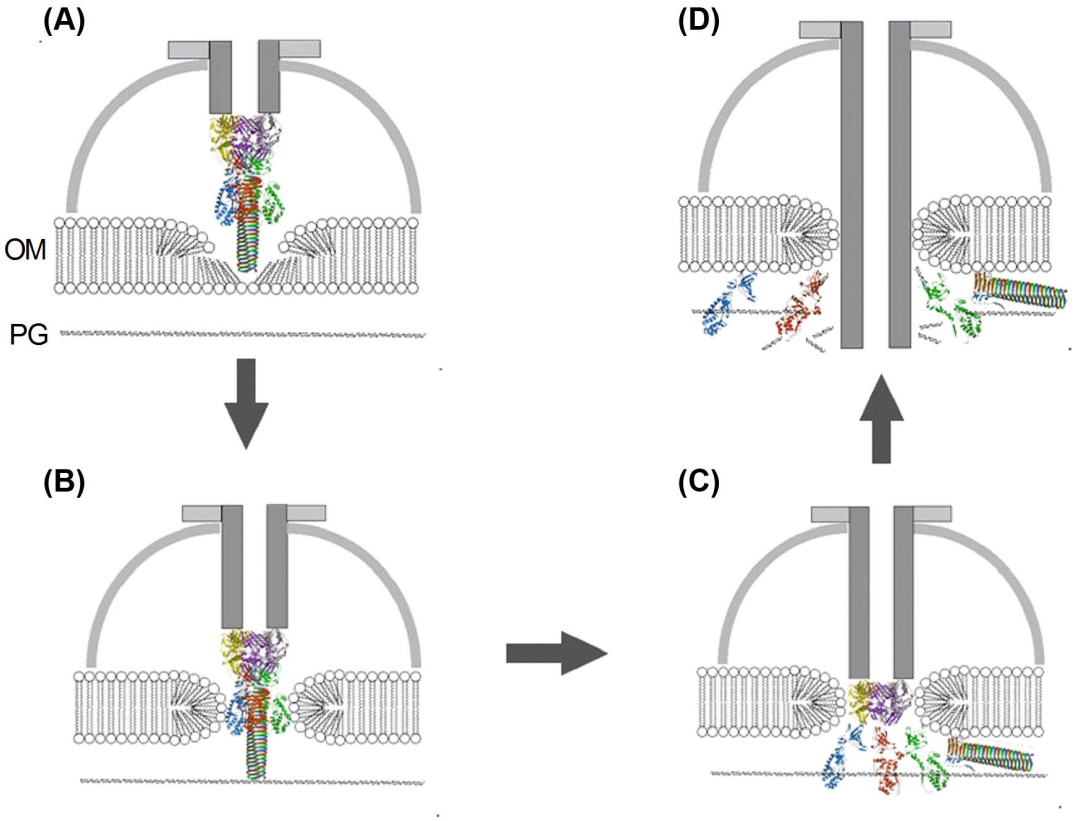

Figure 1-3. The function of T4 $(gp27)_3$-$(gp5)_3$ complex during the infection. Three gp27 are drawn in yellow purple and gray and three gp5 are drawn in blue, red and green. OM and PG indicate outer membrane and peptidoglycan, respectively.

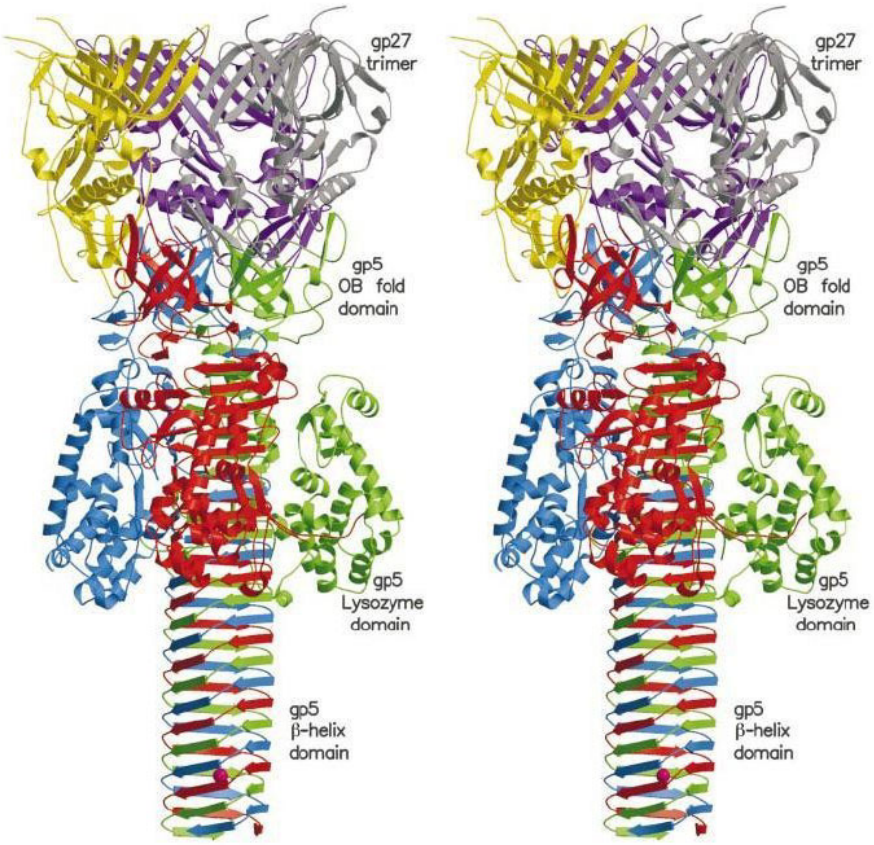

Figure 1-4. The crystal structure of T4 gp27-gp5 complex (*8*). The same colors are used as Figure 1-3.

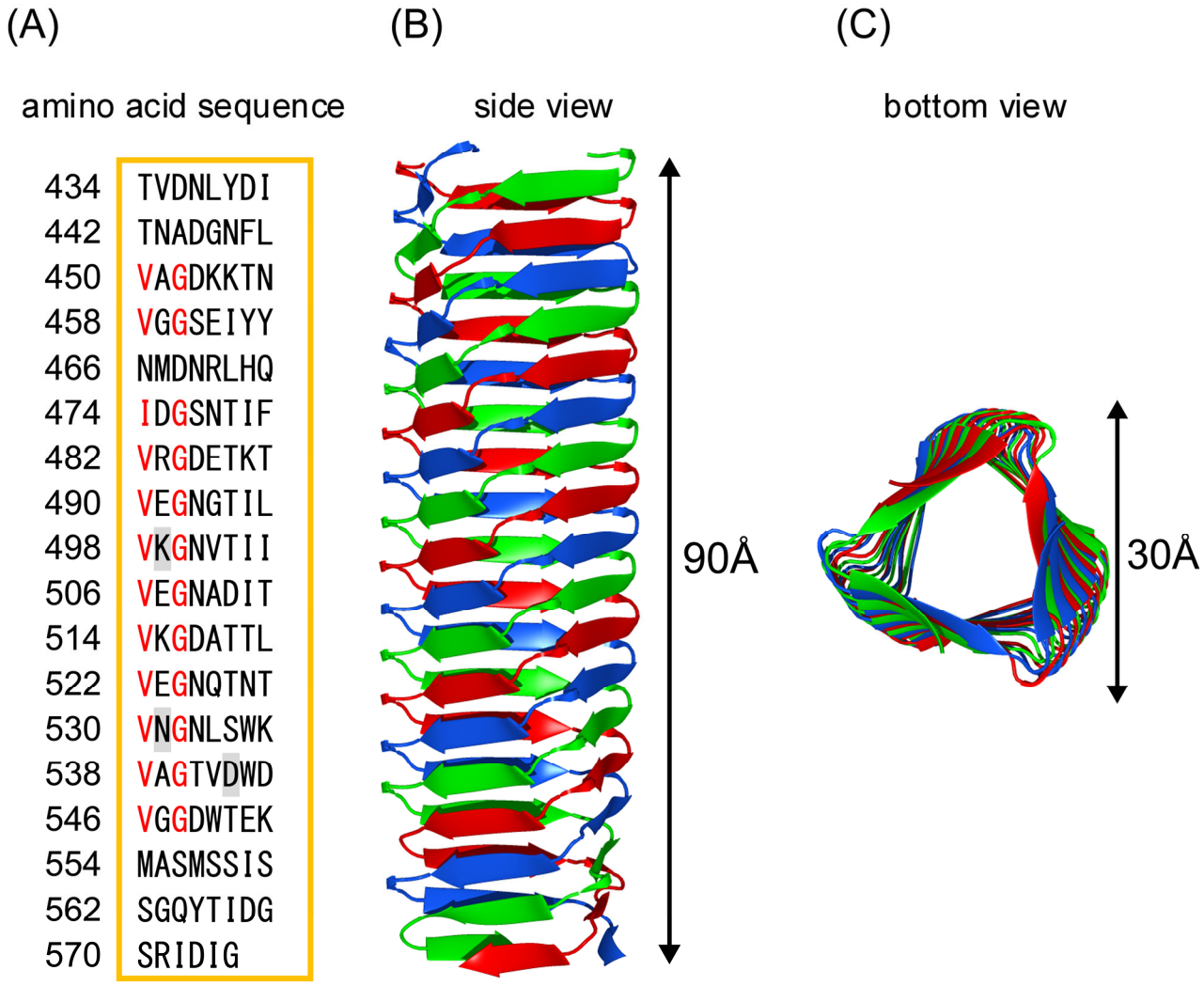

Figure 1-5. The octapeptide (VXGXXXXX) repeat of T4 gp5. (**A**) The amino acid sequence of the intertwined three-stranded β-helix of T4 gp5. (**B-C**) The side view and the bottom view of the intertwined three-stranded β-helix of T4 gp5.

1-5 The three-stranded β-helix as a nanomaterial

The three-stranded β-helix is extremely stable against heat and denaturing agent thus it is suitable for nanomaterial. As I mentioned in the previous section, T4 gp5 has a particularly regular and repetitive three-stranded β-helix thus some nanomaterials have been designed using it. For example, ball-and-spike structure was constructed by fusing the three-stranded β-helix of T4 gp5 to ferritin-like protein LisDps (Figure 1-6) (*9*) and tetrapod structure was constructed by assembling four T4 gp5 molecules with C-terminal hexa-histidine tag via a gold particle (*10*) (Figure 1-7). In 2010, the scaffold of a chemical reaction was also constructed by chemically conjugating the derivative of flavin to the three-dimensionally ordered lysine residues of the three-stranded β-helix of T4 gp5 (Figure 1-8) (*11*).

In our laboratory, we have been trying to produce protein nanotube using the three-stranded β-helix of T4 phage gp5 as a template structure. The results in this direction will be described in the introduction of Chapter 4.

In the following sections, I will review the variations of the C-terminal region of T4 gp5 in bacteriophages and recently discovered VgrG proteins of bacterial type 6 secretion systems.

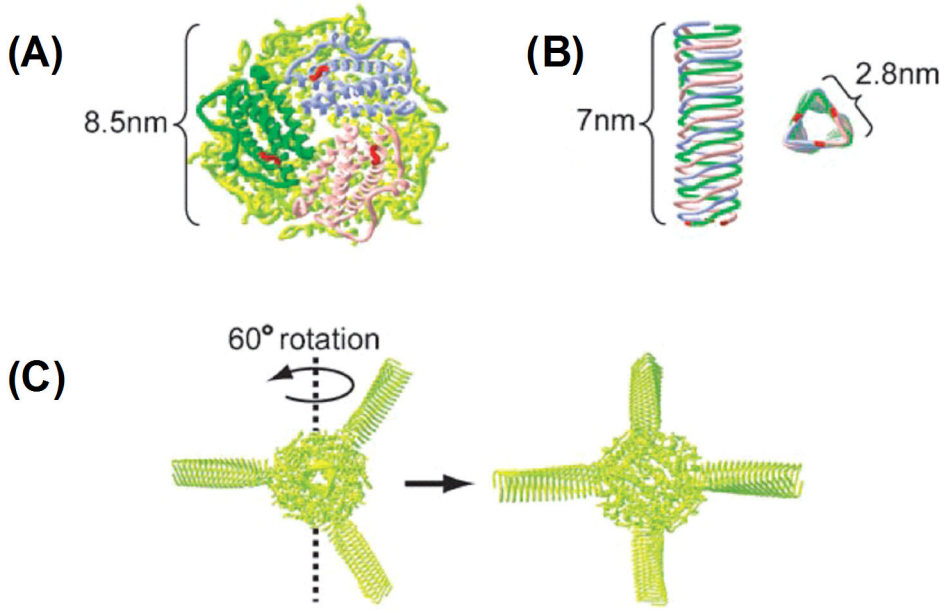

Figure 1-6. The application of T4 gp5 (ball and spike structure) (Sugimoto et al., 2006, partly modified) (*9*). (**A**) The structure of LisDps. (**B**) The structure of three-stranded intertwined β-helix of T4 gp5. (**C**) The ball and spike structure.

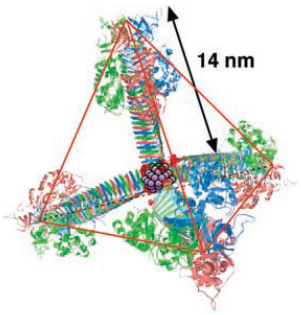

Figure 1-7. The application of T4 gp5 (tetrapod structure) (Ueno et al., 2006) (*10*). Four T4 gp5 molecules with C-terminal hexa-histidine tag are associated via a gold particle.

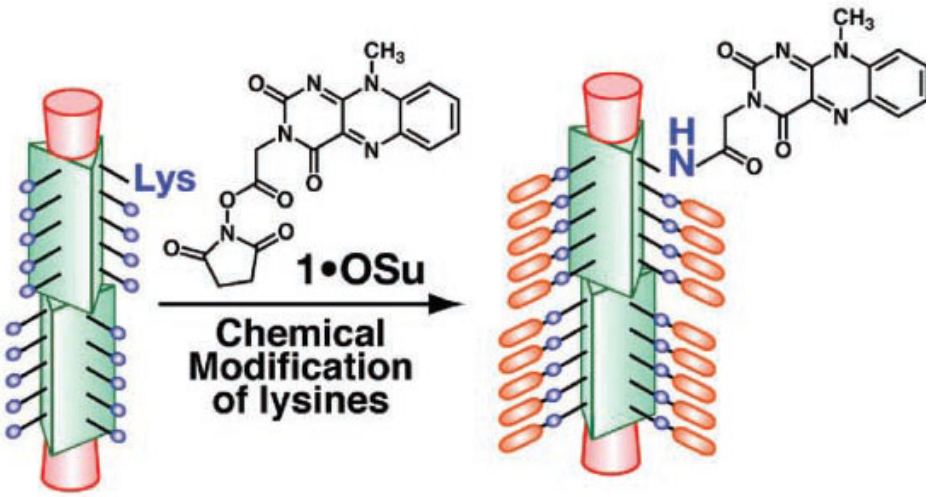

Figure 1-8. The application of T4 gp5 (scaffold for a chemical reaction) (Yokoi et al, 2010 (*11*)). The green prism indicates a part of T4 gp5C and the red cylinder indicates the foldon of T4 wac. The orange rod indicates the derivative of flavin.

1-6 The variation of T4 gp5 homologs in bacteriophages

From the viewpoint of the evolutionary tree, T4 and T4-related phages constitute a subfamily of *Myoviridae* (the bacteriophage with contractile tail). T4-related phages are classified into four groups based on the amino acid sequence identity of the head forming protein to that of T4 phage, namely, T-even (>90% identity), pseudo T-even (80-90% identity), schizo T-even (50-80% identity) and exo T-even (<50% identity) (Figure 1-9) (*12, 13*). The homolog of T4 gp5 is found in the bacteriophages of all these four groups but their amino acids sequence identities against T4 gp5 are 40-50% in pseudo T-even, 30-50% in schizo T-even and <15% in exo-T-even. The lysozyme domain of T4 gp5 is not found in KVP40 gp5 (*14*). Comparing the amino acid sequence of the C-terminal domain of T4 gp5 with gp5 homologs, the T-even gp5 is almost the same as that of T4 gp5, but that of pseudo T-even and shizo T-even has less than 40 % of identity with that of T4 gp5 and some insertions of eight residues or so are observed. Among them, gp5 of KVP40 phage has a lower identity of about 35 % and four insertions of 7-8 residues are present (Figure 1-10). Thus gp5 of KVP40 was chosen as a target of this study as described later.

In addition to T4-related phages, the bacteriophages belonging to other subfamilies of *Myoviridae* have been shown to contain the homolog of gp5 in the past two years. In 2012, the crystal structure of gp138 from phi92 phage (rv5-like group) and the gpV from P2 phage (P2-like group) were determined (*15, 16*) and, in 2013, the crystal structure of gp45 from Mu phage (Mu-like group) was determined (*17*) (Figure 1-11). The detailed comparison will be described in Chapter 2. In the next section, I will focus on the recently discovered VgrG protein of bacterial type 6 secretion system.

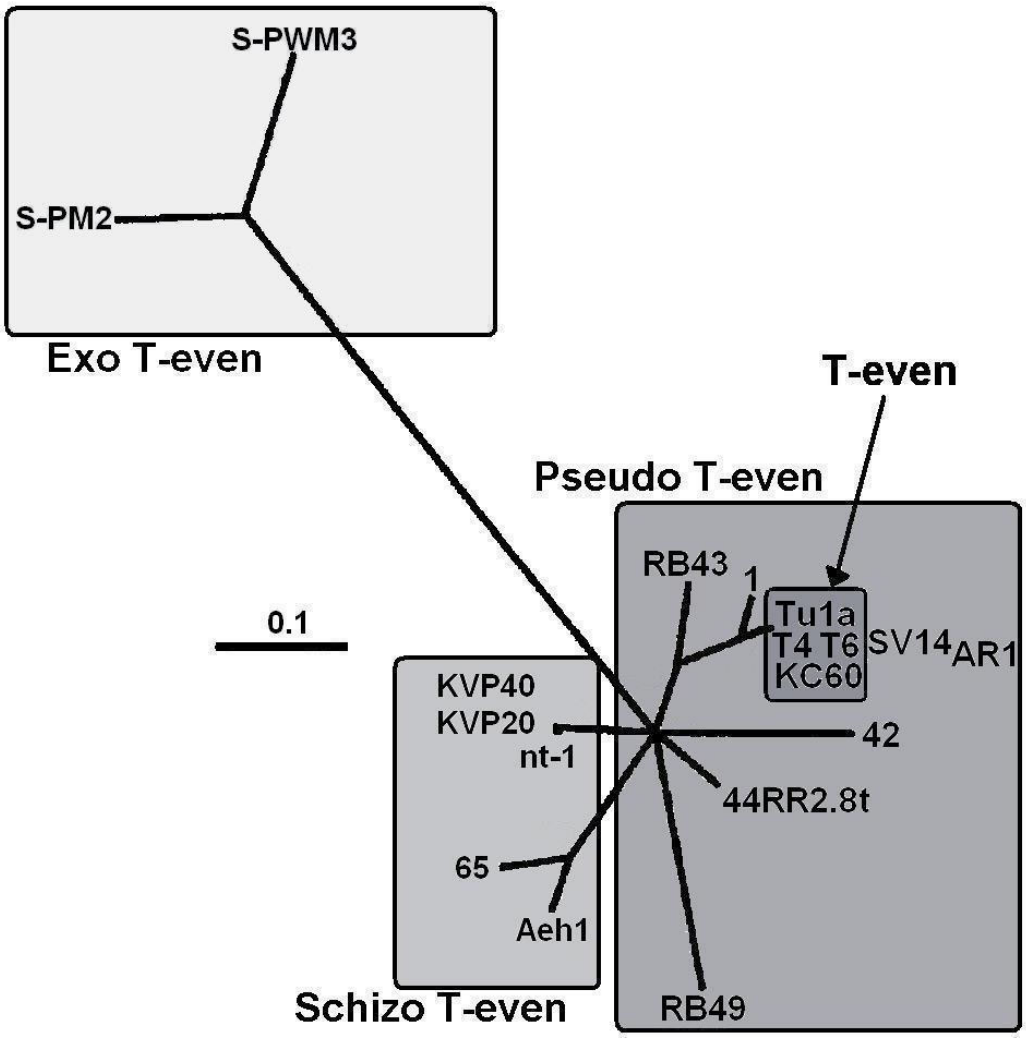

Figure 1-9. The four subgroups of T4-related phages (Emma Hambly et al., 2001, partly modified) (*13*).

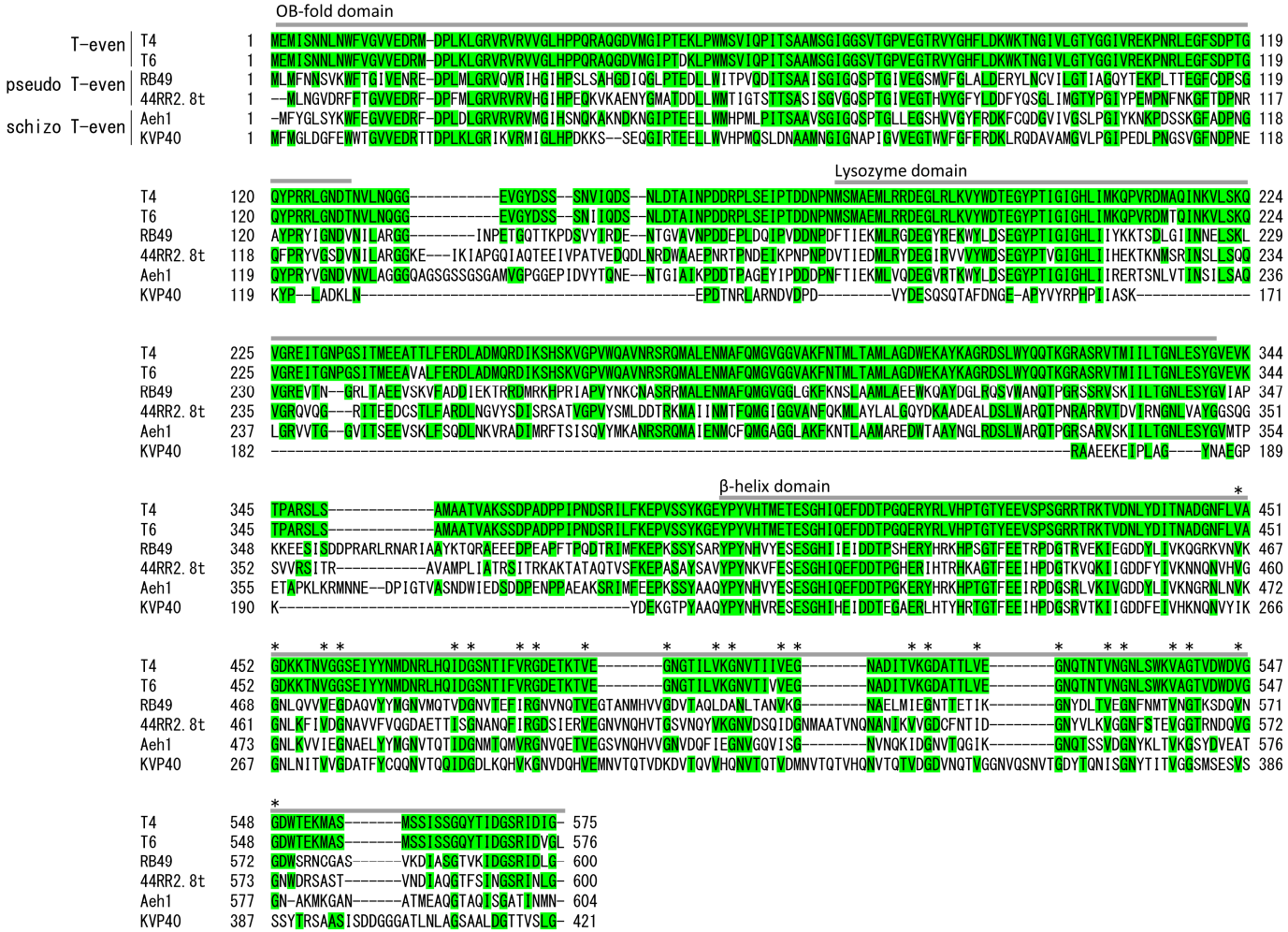

Figure 1-10. The multiple sequence alignment of gp5 homologs of T4-related phages. The sequences alignment was carried out by clustalW2 (*18*). The amino acids identical to T4 gp5 are colored in green. Asterisk indicates the conserved valine and glycine residues of T4 gp5 β-helix domain.

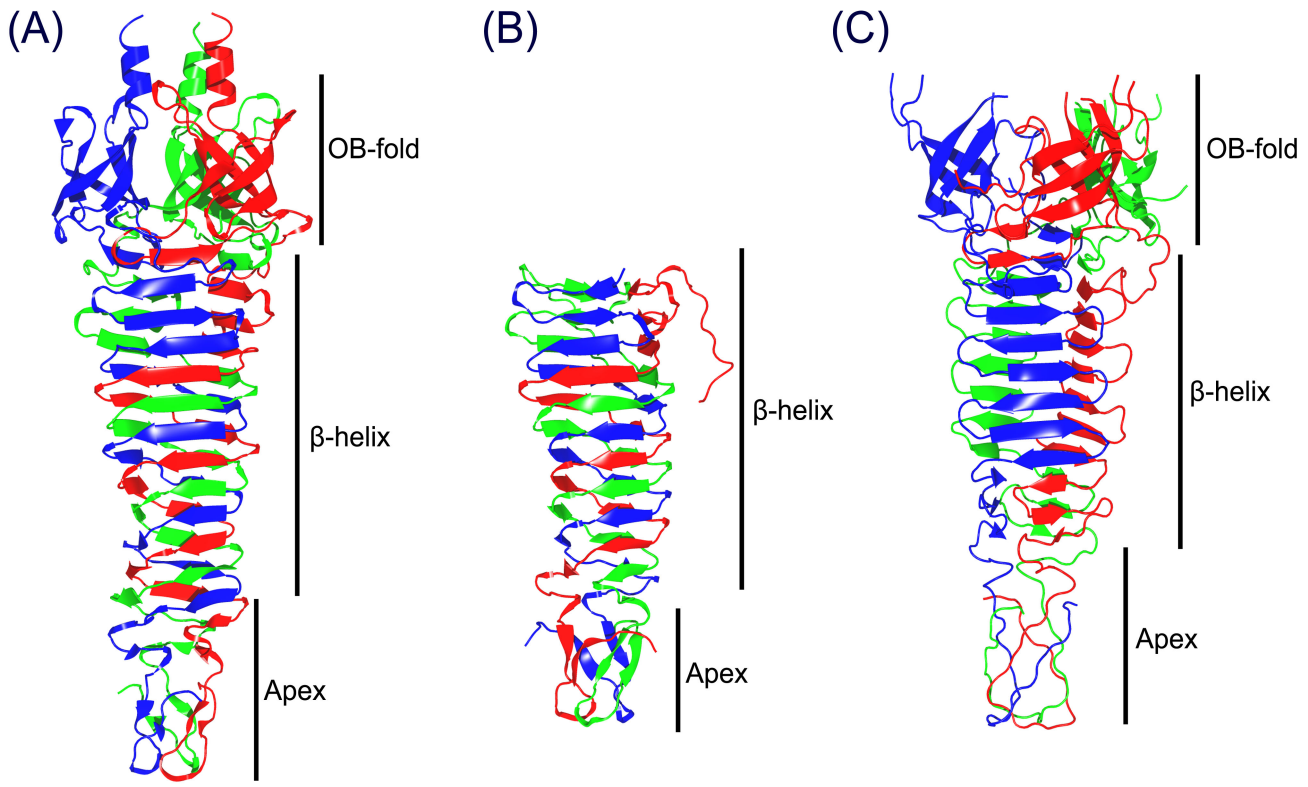

Figure 1-11. The crystal structures of T4 gp5 homologs of distantly related *Myophages*. (**A**) gpV of P2 phage (P2 group) (PDB code 3QR8 and 3QR7). (**B**) gp45 of Mu phage (Mu group) (PDB code 3VTO). (**C**) gp138 of phi92 phage (rv5-like) (PDB code 3PQI). In T4 phage, the apex domain is present as a separated protein (gp5.4 (*19*)).

1-7 VgrG proteins, the homolog of T4 gp27-gp5 complex in bacterial Type 6 secretion systems

Bacterial type 6 secretion system was first discovered in *Pseudomonas aeruginosa (20)* and *Vibrio cholerae* (*21*) at the almost same time in 2006. The following studies have revealed that the type 6 secretion system is widely distributed in Gram-negative bacteria; it is found in more than 25% of Gram-negative bacteria with known genome sequence (*22*). The overall structure of type 6 secretion systems and their various biological functions are described in the Introduction of Chapter 2.

The T4 gp5 homolog in Type 6 secretion system is called VgrG protein (valine glycine repeat protein G). VgrG proteins commonly contain a homologous region of T4 gp27, which sits on gp5 trimer (Figure 1-4) in T4 phage, as a domain of the polypeptide in addition to the OB-fold domain and the three-stranded β-helix domain of T4 gp5. However, it does not possess the lysozyme domain, which is present in T4 gp5. The number of predicted β-strands and their amino acids sequences, which presumably form three-stranded β-helix, are highly diverse among VgrG proteins even in the same strain of bacterium. I focus on the VgrG proteins of *E. coli*, but even limited to VgrG protein in *E. coli*, there are three types of VgrG proteins, namely, one type containing VXGXXXXX repeat like T4 gp5, another type containing VGXXXXXX and the last type containing no apparent repeat sequence.

The homology of the phage protein, gp27-gp5 and VgrG protein suggests some evolutional and structural relationship of this highly specified protein. I am interested in the relationship between the variation of the amino acid sequence and the three-dimensional structure.

1-8 The purpose of this study

The three-stranded β-helix of T4 gp5 has a highly regular and repeat structure and is extremely stable against heat and denaturing agent. It is thus suitable for the template in designing rod-shaped nanomaterial.

In the last 10 years, a huge number of T4 gp5 homologs were found in the bacteriophages and the bacterial type 6 secretion systems (in the latter, they are called VgrG proteins). Some of them contain an octapeptide repeat of VXGXXXXX like T4 gp5, but others contain the repeat of VGXXXXXX and the remaining have no apparent repeat sequence.

In the present study, I over-expressed, purified and characterized the VgrG proteins from pathogenic *E. coli* O157 and CFT073 strains and gp5 from KVP40 phage. The crystal structure of the C-terminal region of *E. coli* O157 VgrG1, which has no apparent repeat, was determined at 1.95Å resolution and the relationship between the amino acid sequence and the three-stranded β-helix fold were analyzed.

Gp5 from T4 and schizo T-even phage KVP40 have the same VXGXXXXX repeats, but have some distinct feature as will be described in Chapter 3. This system gives us a unique and intriguing opportunity to investigate how this distinct feature of the octa amino acids repeat is reflected in the three-dimensional structure of the β-helix.

I have also constructed and over-expressed the artificial three-stranded β-helical polypeptides by duplicating the 48 amino acids (six β-strands) of T4 phage gp5 and fusing the SlyD and foldon at N-terminus and C-terminus, respectively. It was characterized in detail, and, finally, the artificial three-stranded β-helical polypeptide was purified to near homogeneity in dimer of trimer state by the use of limited proteolysis.

Chapter 2

Structure and properties of VgrG proteins from *Escherichia coli* O157 and CFT073

Bacterial Type 6 secretion system (T6SS) is a highly complex organelle localized in the cell membrane (*23*). The T6SS genes often cluster together, and this cluster consists of highly conserved 13 core genes and several strain-specific genes (*24*). The structural proteins of T6SS assemble into a contractile supramolecular structure with the length of ~0.7 μm (*23*). Upon contraction of the sheath structure, it shortens down to ~0.4 μm and drives the secretion of protein toxins or effector molecules into target cells (*23*). The physiological functions played by T6SS significantly vary; *e.g.* cytotoxicity against eukaryotic cells by actin crosslinking (*Vibrio cholerae* (*25, 26*)) or actin ribosylation (*Aeromonas hydrophila* (*27*)), interbacterial competition (*Pseudomonas aeruginosa* (*28, 29*) and *V. cholera* (*30*)), biofilm formation (enteroaggregative *Escherichia coli* (EAEC) (*31*)) and virulence (*Burkholderia* mallei (*32, 33*) and *P. aeruginosa* (*20, 34*)).

The Type 6 secretion apparatus consists of two parts, namely the membrane-associated protein complex and a syringe-like structure extending from it to the cytosol. The syringe-like structure is structurally analogous to the tail of contractile bacteriophage though it is much longer than the contractile tail (the tail of T4 phage is 0.1 μm). The syringe-like structure is assumed to consist of the following proteins – VgrG, Hcp, and VipA/VipB – which are the homologs of the gp27-gp5 complex, gp19 (tail tube protein) and gp18 (tail sheath protein) of T4 phage, respectively (*23*). On the other hand, most of the membrane-associated proteins are not homologous to phage proteins. Interestingly, a homolog of a small protein from the T4 tail baseplate called gp25, is found in T6SS. Therefore, membrane-associated proteins may function as the baseplate of Type 6 secretion system (*35*).

In T6SS, VgrG protein is thought to function similar to the gp27-gp5 complex in T4 tail; it may penetrate the outer membrane of the host cell and the membrane of target cell from inside and outside, respectively. VgrG is an essential component of T6SS and found in the culture supernatants of the T6SS expressing cells (*25, 30, 36*).

Multiple *vgrG* genes are often found in many bacterial genomes. In the case of pathogenic *E. coli* O157 and CFT073 strains, three *vgrG* genes are identified in each strains, namely *vgrG1*, *vgrG2* and *vgrG3* of *E. coli* O157 and *c3393*, *c1888* and *c1883* of *E. coli* CFT073 (Figure 2-2). The *vgrG* genes of some bacteria contain a variable 3' part, which encodes various enzymatic domains called "effectors" (*e.g.* actin-crosslinking domain and peptidoglycan-binding domain found in *V. cholerae* V52) (*25*), but *vgrG* genes of *E. coli* O157 and CFT073 strains does not contain them.

VgrG protein is predicted to contain triple-stranded β-helix as in gp5 of T4 phage (*25*), but the number of predicted β-strands and their amino acid sequences are highly diverse among VgrG proteins. The VgrG proteins of *E. coli* are classified to three groups, namely the first group containing a VXGXXXXX repeat (X, any amino acid), the second group containing VGXXXXXX repeat and the third group which does not contain apparent repeat sequence. The c1888 and c1883 of *E. coli* CFT073 belong to the first group (VXGXXXXX type) and The VgrG2 and VgrG3 of *E. coli* O157 belong to the second group (VGXXXXXX type) while the VgrG1 and c3393 belongs to the third group (Figure 2-3). How these sequence variations are reflected in the three dimensional structure is of great interest.

In this chapter, the characterization of six VgrG proteins from *E. coli* O157 and CFT073 strains and the structure determination of the C-terminal fragment of *E. coli* O157 VgrG1 at 1.95 Å resolution are reported. It is also reported the crystallization and the preliminary X-ray crystallographic analysis of *E. coli* CFT073 c3393. Finally, based on the correlation between amino acid sequences and three-dimensional structures of three-stranded β-helical proteins, a method for the prediction of β-helix topology is proposed.

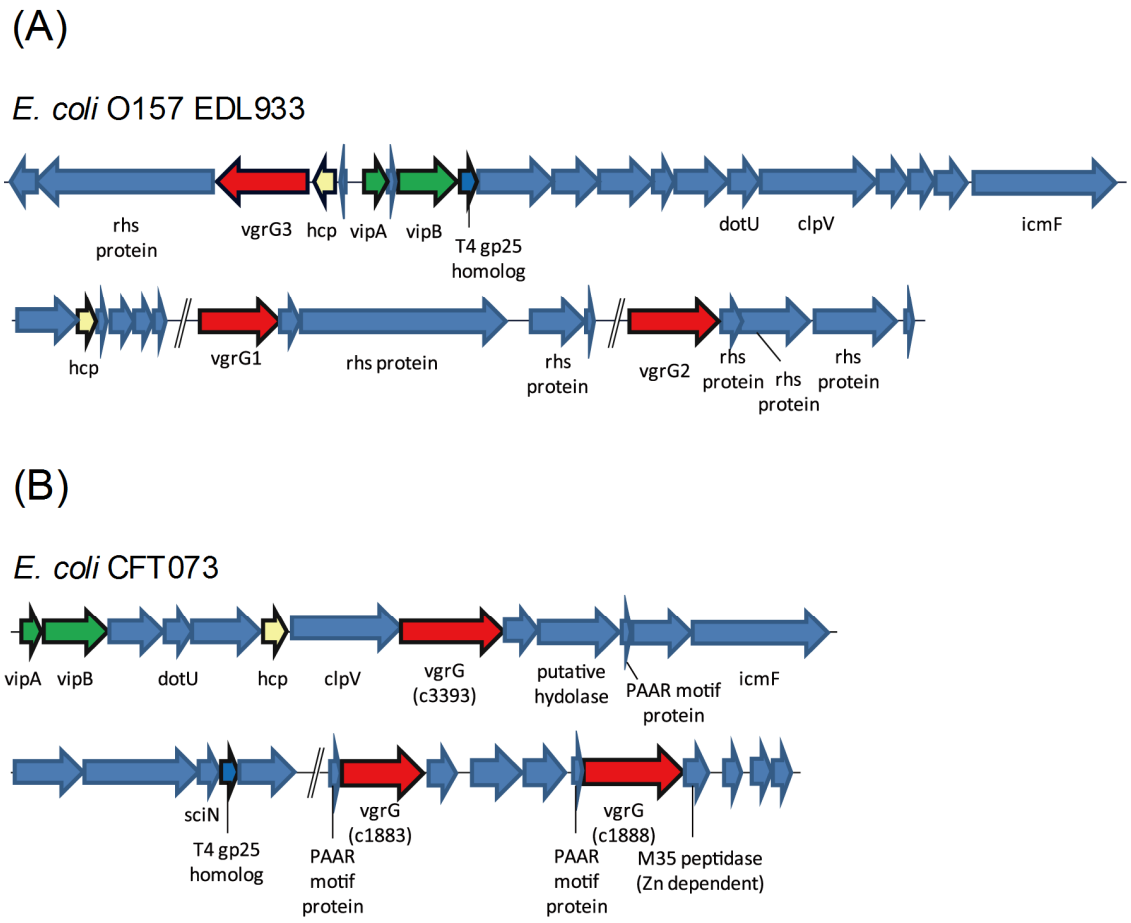

Figure 2-2. Gene origination of the T6SS loci of *E. coli* O157 (**A**) and CFT073 strains (**B**). The homologs of bacteriophage proteins are drawn in red (VgrG protein), yellow (hcp protein), green (vipA/vipB) and blue (T4 gp25 homolog).

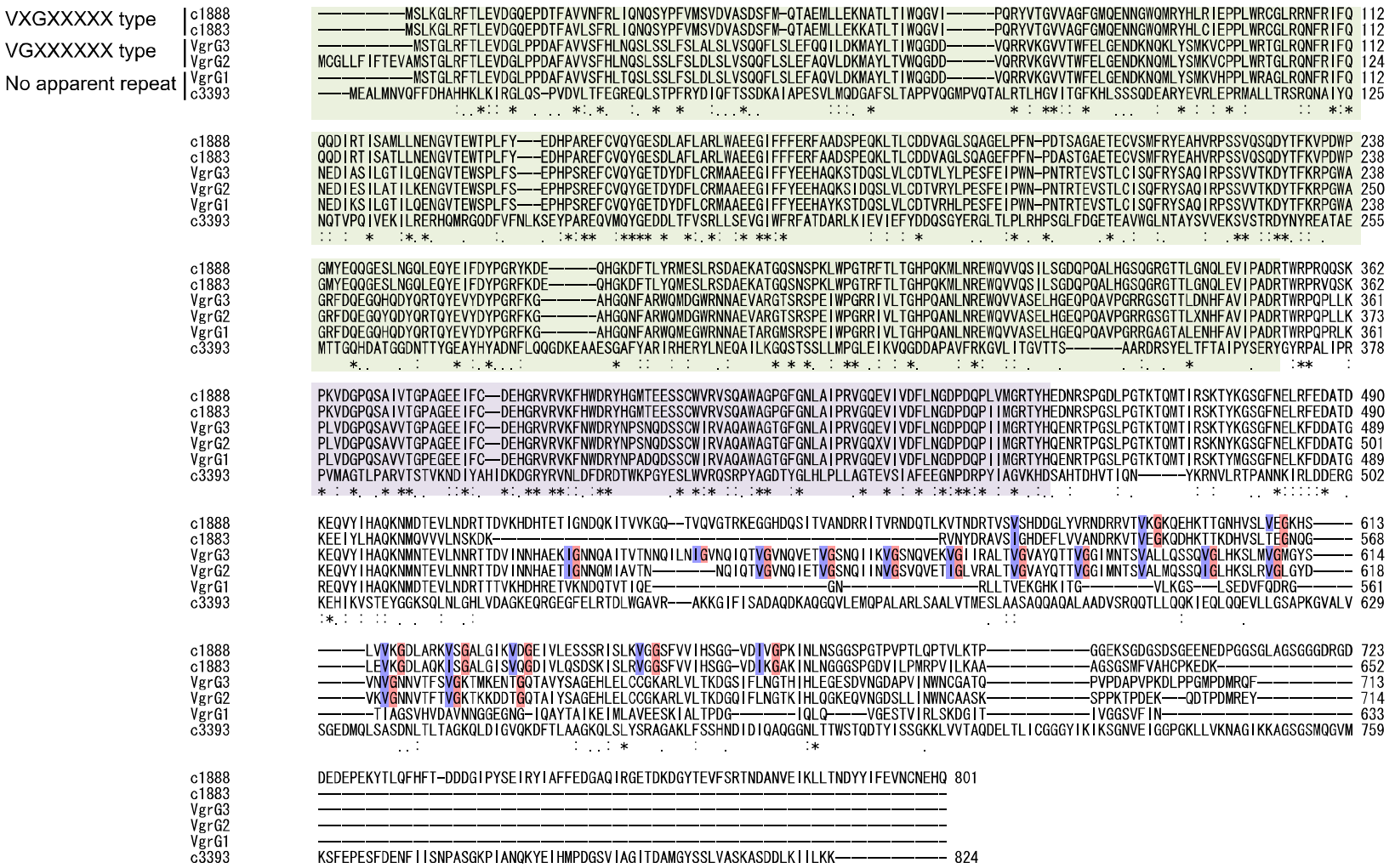

Figure 2-3. Multiple sequence alignment of the VgrG proteins from *E. coli* O157 and CFT073 strains. The yellow region is homologous to T4 gp27 and purple region is OB-fold (based on c3393 (*35*)). The conserved valine and glycine residues are colored in blue and red, respectively. The alignment was performed by clustalW2 (*18*).

Chapter 3

Crystallization and X-ray diffraction analysis of gp5 from bacteriophage KVP40

3-1 Introduction

Bacteriophage KVP40 was isolated from the sea water in Japan using *Vibrio parahaemolyticus* as a host cell (*59*). KVP40 belongs to the *Myoviridae* family like T4 phage and is classified to shizo T-even group for its more elongated head than T4 (*12*). Many structural proteins of KVP40 phage are homologous to that of T4 phage including gp5 (*60*). When the amino acids sequence of KVP40 gp5 is aligned with that of T4, the N-terminal OB fold domain and C-terminal putative β-helix domain show high sequence similarities (46% and 35% respectively) but the lysozyme domain is absent in KVP40 gp5 (*14, 61*). The C-terminal domain of KVP40 gp5 contains conserved asparagine/aspartic acid, valine and glycine residues like T4 gp5, but it also contains conserved valine, glutamine residues which are not seen in T4 gp5 and other β-helical proteins with known structure (Figure 3-1). The putative β-helix domain of KVP40 gp5 is considered to contain 22 β-strands (aside from antiparallel β-helix region), which is the longest among the T4 gp5 homologs that contain VXGXXXXX repeats (blastp search, identity below 85% and E-value cut off of 1×10^{-10}). In the previous analysis in our laboratory, KVP40 gp5 was shown to be trimer in solution and rich in β-structure (*61*), however its atomic structure remains to be determined. In this chapter, I report the crystallization of KVP40 gp5 and the collection of the full X-ray diffraction dataset.

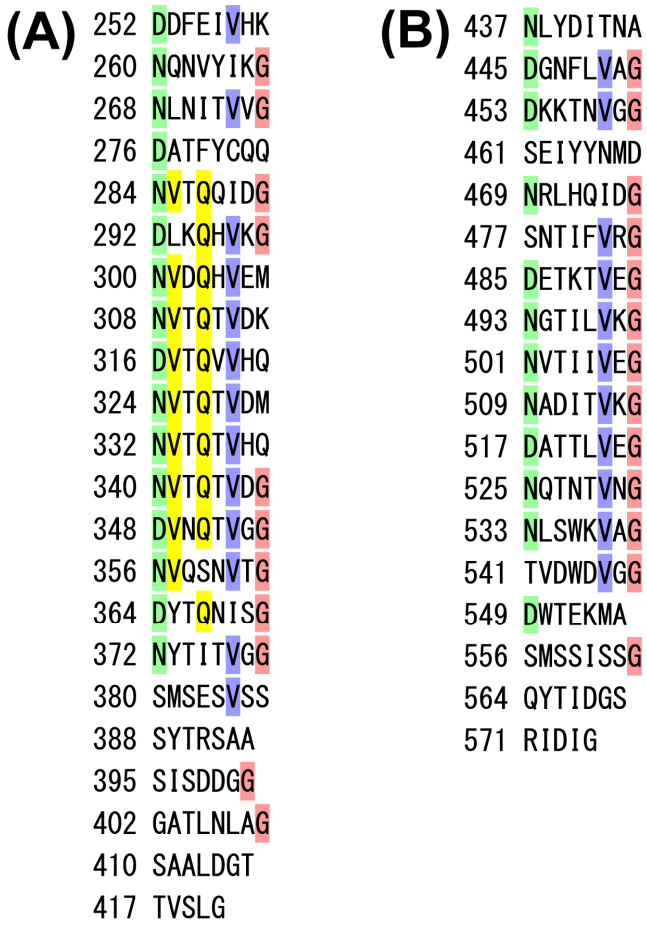

Figure 3-1. The VXGXXXXX repeats of KVP40 gp5 (**A**) and T4 gp5 (**B**). The conserved asparagine/aspartic acid, valine and glycine residues are colored in blue, red and green. The conserved valine-x-glutamine sequences in KVP40 gp5 are colored in yellow.

3-2 Materials and Methods

Media

LB media was used for all *E. coli* cultivations. The composition is described in Materials and Methods section of Chapter 2.

Expression and purification

The over-expression vector of KVP40 $gp5^{his}$ was constructed by cutting KVP40 *gene 5* from pMNK (*61*) using BamHI and SalI restriction enzymes and inserting it to pET29a vector (Novagen), which has been digested by the same restriction enzymes (Nemoto, unpublished data). The resulting plasmid was designated as pMNK3. The expression and purification procedures of KVP40 gp5 with C-terminal hexa-histidine tag ($gp5^{his}$) were referred to the method by Nemoto et al. (*61*) with slight modification. For the expression of KVP40 $gp5^{his}$, BL21(DE3) was transformed by pMNK3, and cultivated in LB medium supplemented with 60 μg/ml kanamycin at 37°C. When OD_{660} was about 0.5-0.6, the expression was induced by IPTG at a final concentration of 1 mM and further cultivated at 30°C for 6 h. Cells were pelleted by the centrifugation at 2,500 ×g for 10 min using Himac CR22G centrifuge with R9AF rotor (Hitachi Koki). The molecular weight of KVP40 $gp5^{his}$ is 47,900 Da in monomer including the purification tag. The purification procedure is same as full-length *E. coli* O157 VgrG1 (see Materials and Methods section of Chapter 2).

Crystallization Screening and the optimization of the crystallization condition

For the crystallization screening, purified KVP40 $gp5^{his}$ was dialyzed against 10 mM Tris pH 8.0 and concentrated to 9.8 mg/ml or 18.0 mg/ml by Amicon Ultra 50K (Millipore). Crystallization screening was performed by hanging drop vapor diffusion method at 20 °C using the 96-well PS microplate (Greiner Bio-one) and the 96 holed hanging drop seal (Asone) with the crystallization screening kits of JBScreen Pentaerythritol (Jena Bioscience), JBScreen PEG/Salt (Jena Bioscience), JBScreen JCSG++ (Jena Bioscience), JBScreen Nuc-Pro (Jena Bioscience), JBScreen PACT++ (Jena Bioscience), JBScreen Basic (Jena Bioscience), JBScreen Classic I (Jena Bioscience), JBScreen Classic II (Jena Bioscience), JBScreen Cryo (Jena Bioscience),

Pro Plex (Molecular Dimensions), Morpheus (Molecular Dimensions), MIDAS (Molecular Dimensions), Structure Screen 1 (Molecular Dimensions), Structure Screen 2 (Molecular Dimensions), MembFac (HAMPTON), Crystal Screen Cryo (HAMPTON) and Cryo (Emerald BioSystems). The drops were prepared by mixing 350 nl or 500 nl protein solution and equal volume of well solution and equilibrated against 30 ul of the same well solution.

Initial crystal was obtained in F5 of JBScreen Pentaerythritol which contains 35 % w/v Pentaerythritol ethoxylate (average M.W. ~270), 100 mM MES pH6.5 and 200 mM $(NH_4)_2SO_4$. The crystallization condition was subsequently optimized in 24-well VDX plate (HAMPTON) by changing the concentration of the protein, pH, the buffers, the type and the concentration of precipitants, and the concentration of $(NH_4)_2SO_4$. The supplementation of the additives was also examined. After the optimization of the crystallization condition, the crystals of KVP40 $gp5^{his}$ were obtained in two different forms namely, rhombohedral shape crystal and hexagonal shaped crystals. The best hexagonal shaped crystals were obtained by mixing 1 ul of 20.0 mg/ml protein and 1ul of the solution containing 300 mM $(NH_4)_2 SO_4$, 100 mM Tris pH7.9 and 10% PEG2000MME and equilibrated against the 1ml of the same well solution. The best rhombohedral shaped crystals were obtained by hanging drop vapor diffusion method using the well solution containing 300 mM $(NH_4)_2SO_4$, 100 mM Tris pH7.9 and 100 mM NH_4Cl or 100 mM KCl. Occasionally, the hexagonal shaped crystals and rhombohedral shaped crystals were obtained in the same drop.

Screening of cryoprotectant

The cryosolutions were prepared by supplementing glycerol, 2-Methyl-2,4-pentanediol (MPD), PEG400 or Li_2SO_4 to the well solution till it becomes cryogenic and increasing the concentration of $(NH_4)_2SO_4$ till the crystal dose not dissolve. The crystals were dipped in the cryo solution directly or in two steps, *e.g.* 5% glycerol for the first and 25% glycerol for final. The crystals were flash frozen in the liquid nitrogen or under cryo stream and their X-ray diffraction patterns were analyzed at BL-1A or BL-17A beamline of Photon Factory at KEK (Tsukuba, Japan).

Collection of X-ray diffraction data and data processing

The rhombohedral shaped crystal was cryoprotected by soaking in the cryosolution containing 400 mM $(NH_4)_2SO_4$, 100 mM Tris pH7.9, 100 mM KCl and 1.6 M Li_2SO_4 for a few seconds and flash frozen by the N_2 cryostream (100K). The X-ray diffraction experiment was performed at BL-1A beamline of Photon Factory at KEK (Tsukuba, Japan). The X-ray diffraction data was collected by fine phi slice strategy (*62, 63*) with the oscillation angle of 0.1° and the exposure time of 0.1 sec using Pilatus 2M-F single photon-counting detector at 100K. Helical mode, *i.e.* changing the diffracting position from one edge to another edge in many steps, was used to minimize the damage of the crystal by the X-ray exposure.

The indexing of the diffraction data was performed with XDS (*47*). The conversion of HKL file to MTZ file and the determination of the space group were carried out by Pointless (*48*). The combination of the separately indexed data were carried out using Pointless and scaling was performed by SCALA (*49*). The statistics of X-ray diffraction data collection are summarized in Table 3-1.

3-3 Results

Crystallization

KVP40 gp5^{his} was over-expressed and purified by the methods reported by M. Nemoto (*61*) with slight modification (see Materials and Methods 3-2). The crystallization screening of ~1050 conditions gave a micro crystal under the condition F5 of JBScreen Pentaerythritol (Figure 3-2 (A)). To obtain larger crystals, the crystallization condition was optimized as described in Materials and Methods. After the optimization of the crystallization condition, the hexagonal shape crystals and rhombohedral shape crystals were obtained under the similar and identical conditions (Figure 3-2 (B-C)). The hexagonal shaped crystals grew to 0.2-0.4 mm within 2-3 weeks. On the other hand, the rhombohedral shaped crystals grew very slowly and it took at least 3 months to obtain big crystals enough for X-ray diffraction experiments.

Screening of the cryogenic condition

The crystallization condition was not cryogenic thus the screening of cryoprotectant was carried out for both hexagonal shaped crystals and rhombohedral shaped crystals. For cryoprotection of the hexagonal shaped crystals, 25% w/v glycerol and 30% w/v MPD were tested. However, no X–ray diffraction was observed in either cryoprotectant though the crystal under only one condition diffracted to ~44Å when it was cryoprotected by glycerol in two steps (5% for the first and 25% for the second) (data not shown). To examine if the cryoprotection process damaged the crystal, I analyzed the X-ray diffraction of the crystal using capillary at room temperature. The result indicated that the crystal itself is unable to diffract the X-ray. For the cryoprotection of rhombohedral shaped crystals, 1.6M Li_2SO_4, 25% w/v glycerol, 30% w/v PEG400 and 30% w/v MPD were tested. When the crystals were cryoprotected by 1.6M Li_2SO_4, the crystals diffracted to 4Å reproducibly. On the other hand, the crystals, which were cryoprotected by 25% glycerol or 25% PEG400, diffracted X-ray to ~3.6Å and ~3.0Å at the best, respectively but if the crystal diffracts the X-ray depended on the batch (data not shown). The crystals, which were cryoprotected by 30% w/v MPD, did not diffract X-ray at all (data not shown).

1.6 M Li_2SO_4 cryoprotected the rhombohedral shaped crystals most

reproducibly but the salt tends to precipitate during the freezing process in liquid nitrogen. To avoid the precipitation of the Li_2SO_4, I examined the freezing of the crystal under the nitrogen cryostream (100K), which enables more rapid freezing than liquid nitrogen. The crystals which were flash frozen in cryostream showed no deposition of the salt and diffracted to 2.5-2.6Å reproducibly.

X-ray crystallographic analysis

The rhombohedral shaped crystal was cryoprotected by 1.6 M Li_2SO_4 and subjected to X-ray diffraction data collection. The full dataset was collected at a resolution of 2.6Å by helical mode. The crystal showed strong smear diffraction around 4.7Å (Figure 3-3). The spots were very clear and the mosaicity was estimated to be 0.091 by XDS, which indicated that the KVP40 $gp5^{his}$ is ordered very well in the crystal. Phenix.xtriage (*50*) indicated that the crystal is twin, but the indexing in space group P1 could identify that the crystal switches at the center of the crystal (two crystals associated in head to head manner). The indexing by XDS and the Laue group determination by Pointless indicated that the crystal belonged to space group H3, with cell unit size a=b=72.16Å and c=817.72Å. The statistics of the X-ray diffraction data collection are summarized in Table 3-1.

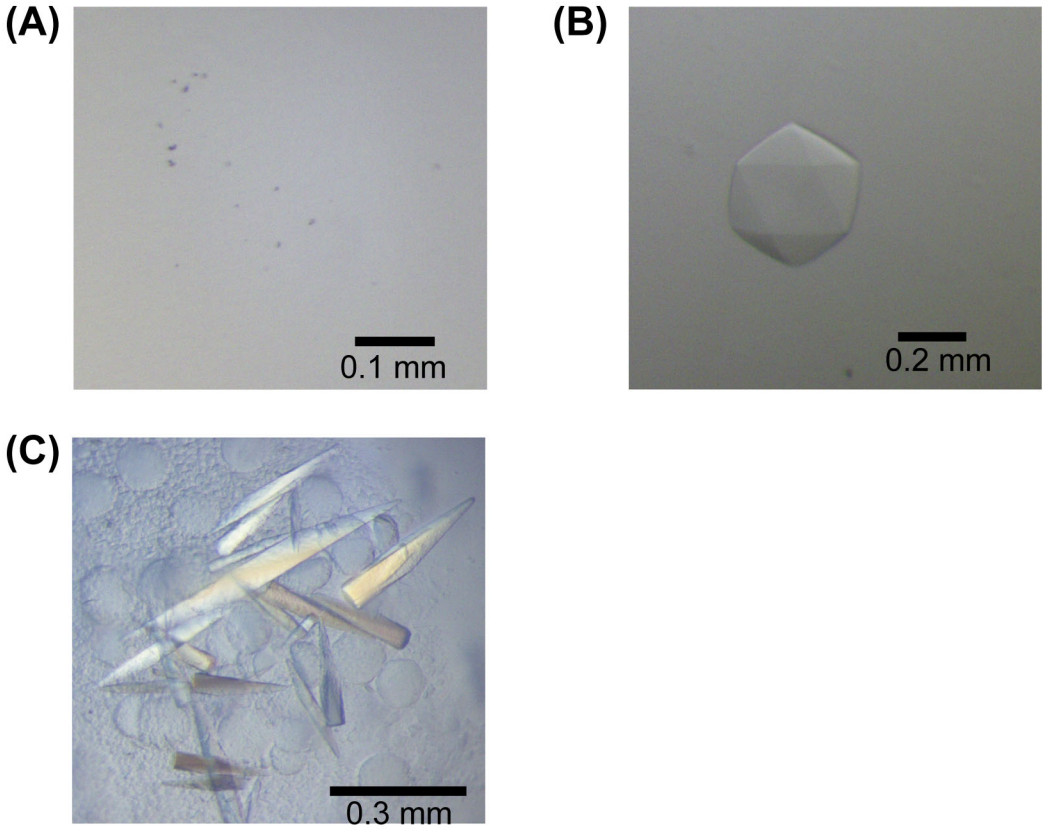

Figure 3-2. The crystals of KVP40 gp5^{his}. (**A**) The initial crystals of KVP40 gp5^{his}. The bar indicates 0.1 mm. (**B**) The hexagonal shaped crystals of KVP40 gp5^{his}. The bar indicates 0.2 mm. The crystal did not polarize. (**C**) The rhombohedral shaped crystal of KVP40 gp5^{his} (polarized crystals). The bar indicates 0.3 mm. The rhombohedral shaped crystals and the hexagonal shaped crystals were occasionally obtained in the same drop.

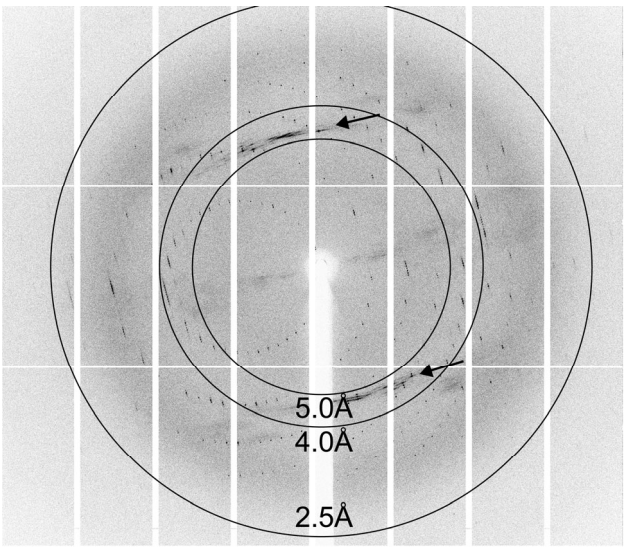

Figure 3-3. The typical X-ray diffraction pattern of KVP40 gp5^{his}. The arrows indicate the smear diffractions at ~4.7Å, which are characteristic in β-helical proteins. The black circles indicate the resolution at 5.0Å, 4.0Å and 2.5Å.

Table 3-1 The statistics of the X-ray diffraction data collection.

	Native
Data Collection	
Beam line	BL-1A, Photon Factory
Wavelength (Å)	1.100000
Space group	H3
Unit cell (Å)	72.16, 72.16, 817.72 90°, 90°, 120°
Resolution range (Å)	47.84-2.59 (2.74-2.59)
No. of observations [a]	411507 (34875)
No. of unique observations [a]	49040 (7058)
Multiplicity [a]	8.4
Completeness (%) [a]	99.7 (98.5)
Mean [(I)/sd(I)] [a]	15.2 (3.4)
R_{merge} (%) a	11.2 (47.6)

[a] The number in the parenthesis shows the statistics for the highest resolution bin (2.74-2.59Å)

3-4 Discussion

KVP40 phage belongs to the schizo T-even group while T4 phage belongs to the T-even group (*12*). The VXGXXXXX repeat, which forms intertwined three-stranded β-helix in T4 gp5, is also found in KVP40 gp5, whereas KVP40 gp5 contains four more repeats than T4 gp5 (Figure 3-1). In addition, KVP40 gp5 contains the conserved valine-x-glutamine motif, which is not found in T4 gp5. How these differences of the amino acid sequence affect the structure of KVP40 gp5, namely whether it is different from that of T4 gp5, is of great interest.

The crystal of KVP40 $gp5^{his}$ was successfully produced, but it was twinned; two crystals associated at the center of the crystal. The analysis of the crystal orientations indicated that two crystals have the opposite directions of c-axis and the a- and b- axes of one crystal are parallel with the b- and a- axes of the other crystal. This crystal orientation suggests that two crystals have grown from the same position to the opposite directions.

When the crystal of KVP40 $gp5^{his}$ was subjected to X-ray diffraction analysis, two strong smear diffractions were observed at ~4.7Å. They are often observed with the β-helical proteins such as $VgrG1C^{G561}$ (Figure 2-21) and T4 gp27-gp5 complex and thought to be the diffraction from the three-dimensionally ordered β-structure. Hence it is most probable that KVP40 $gp5^{his}$ contains the three-stranded β-helix. In Chapter 2, I described a prediction method of the topology of three-stranded β-helix (Figure 2-26). Based on the prediction method, I predict that the putative three-stranded β-helix of KVP40 gp5 would be the intertwined type (Figure 3-4).

To determine the phase, molecular replacement was tried by Phaser (*58*) using the T4 gp5 without lysozyme domain and loops as a search model. The molecular replacement found two molecules with the reasonable packing; each molecule seems to interact side by side and tail to tail if hypothesizing four more β-strands at the end of putative β-helix. However the final R-factor of 0.596 in molecular replacement was inconclusive whether the solution is true or not. I am currently improving the initial model and also trying to crystallize the SeMet derivative to determine the phase unambiguously.

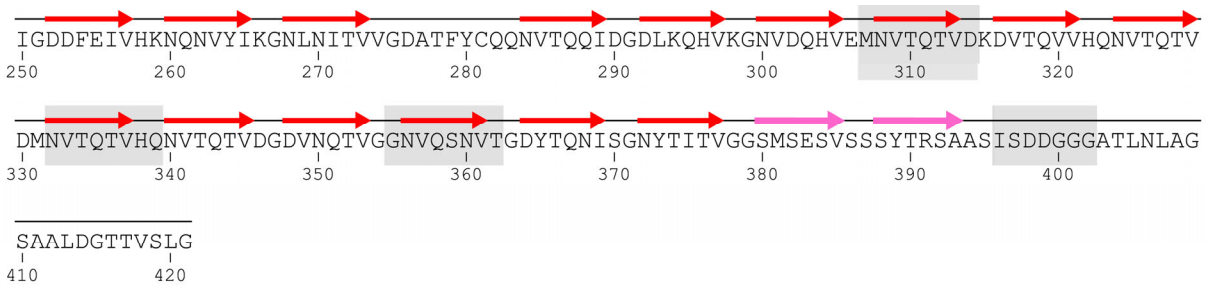

Figure 3-4. The prediction of the β-helix topology of KVP40 gp5. The topology of the putative β-helix region of KVP40 gp5 was predicted by the method described in Chapter 2. The red arrows indicate the intertwined β-helix and the pink arrows indicate the possible intertwined β-helix (weakly match intertwined β-helix, but not match antiparallel β-helix). The gray regions are the insertion sequences (compared with T4 gp5).

Chapter 4

Approaches to obtain homogeneous artificial thee-stranded β-helical polypeptides

4-1 Introduction

The C-terminal domain of T4 gp5 has a highly regular three-stranded β-helix, which contains 12 repeats of octapeptide VXGXXXXX. It shows extraordinary stability against heat and denaturing agents; the melting temperature is around 105°C (*11*) and the structure is resistant to 10% SDS or 2 M guanidine HCl (*8*). In addition, it is stable against chemical and pH stress; the secondary structure is maintained in the presence of *t*-butyl alcohol up to 50%, methanol up to 70% or acetonitrile up to 60% or in the pH range of 2.0 - 10.0 (*11*). Thus the three-stranded β-helix of T4 gp5 is suitable for the template of nanomaterial design. In fact, several nanostructures have been developed using wild type T4 gp5 as described in Chapter 1.

In our laboratory, it has been challenged to produce protein nanotube with desired lengths by stacking the three-stranded β-helix of T4 gp5 (Figure 4-1). The control of the length of three-stranded β-helix is important, because it provides not only new, homogeneous nanomaterial, but also gives rise to the variation of established nanostructures. In the system which was developed in our laboratory, the length of the three-stranded β-helix is controlled by the tandem connection of a DNA fragment coding a certain region of T4 gp5, the "building block" (Figure 4-1 (C) left), which theoretically enables to regulate the length of three-stranded β-helix by 4.7Å unit (corresponding to one β-strand).

The first artificial three stranded β-helix was expressed in *E. coli* using the building block of four VXGXXXXX repeats ($(VXG)_4$) (Figure 4-2 (A) pink box and Figure 4-3 (A) left). The expressed polypeptide chain, however, did not form trimer (Ryuto, unpublished data). Since then, a number of artificial three-stranded β-helices have been designed using 12 repeats of VXGXXXXX sequence as a model system in order to obtain trimeric artificial three-stranded β-helix. A partial success has been achieved by the fusion of the hydrophobic core of T4 gp5C (Figure 4-2(A) blue box,

hereafter called “Cgp5C” or “C”) and foldon (Figure 4-2 (B) red box, hereafter called foldon or “f”) to the C-terminus (Figure 4-3 middle); Cgp5C is important for the folding of T4 gp5 (Boudko, unpublished data) and the foldon is known to promote the trimerization of fused proteins (*64*). The yield of trimeric $(VXG)_{4\times3}Cf$ has been increased up to 10-15%, however, the remaining polypeptides formed a dimer of trimer and large aggregates (bigger than dimer of trimer) (*65*). The decrease of large aggregates has been achieved by changing the building block from four repeat ($(VXG)_4$) to six repeat ($(VXG)_6$) (Figure 4-2 (A) red box) (*66*). $(VXG)_{6\times2}C$ has formed large aggregates for ~26% of total polypeptides (Figure 4-3), which was dramatically smaller than for that of $(VXG)_{4\times3}Cf$ (~50%). However, the main product has been dimer of trimer and the formation of large aggregates could not be completely abolished. In the following two years, the crystal structures of $(VXG)_{4\times1}$ (Figure 4-4) as well as several natural three-stranded β-helical proteins have been determined and the natural β-helical proteins are shown to contain a “capping structure” at the N terminal end of β-helix (*16*). The artificial β-helical polypeptides do not contain the “capping structure”, hence they are considered to dimerize via the N-terminus like $(VXG)_{4\times1}$. Thus, a certain “capping structure”, which prevents the β-helix - β-helix interaction, seems to be required for further decrease of associations.

In the present study, I constructed the over-expression system of N-terminally SlyD fused artificial three-stranded β-helix (Figure 4-5) to examine if it could inhibit the association of the β-helix via the N-terminus. I also succeeded in detecting partially unfolded species by native PAGE and discriminating the fully folded homogeneous three-stranded β-helix from partially unfolded species by trypsin digestion.

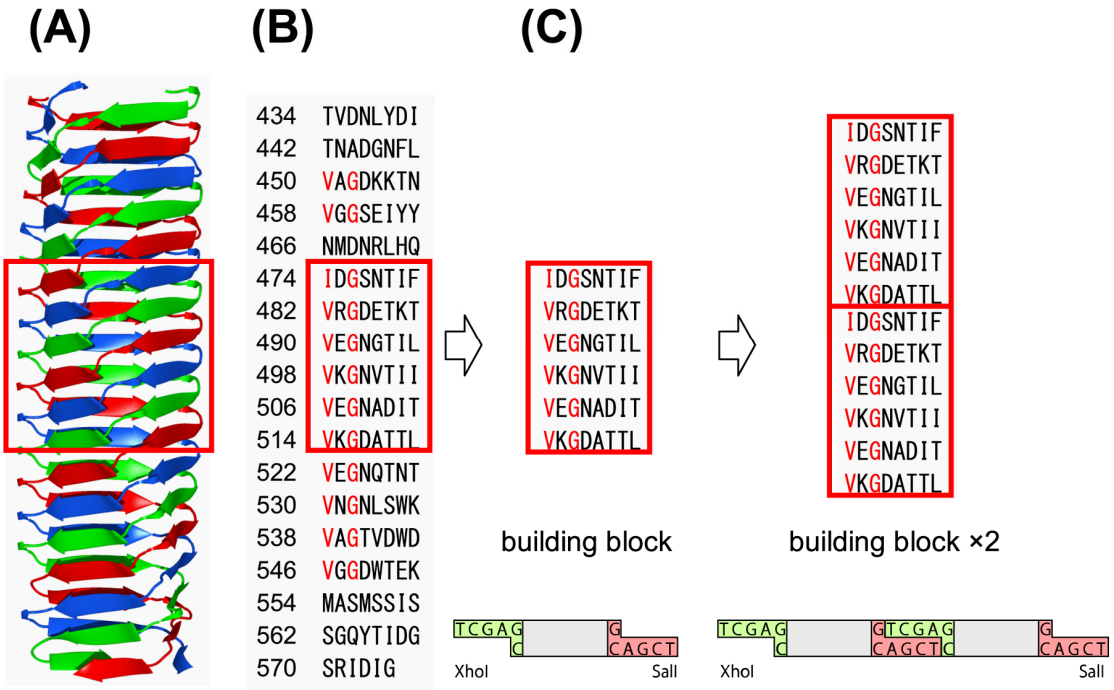

Figure 4-1. The strategy of the construction of artificial three-stranded β-helix. (**A**) The crystal structure of the three-stranded β-helix of T4 gp5 (PDB 1k28). (**B**) The amino acid sequence of the three-stranded β-helix of T4 gp5. (**C**) The concept of the length control of three-stranded β-helix. The DNA fragment with the sticky end of XhoI and SalI can be ligated so that the resulting DNA (5'-GTCGAG-3') is not digested by neither XhoI nor SalI, which enables the increment of the building block one by one.

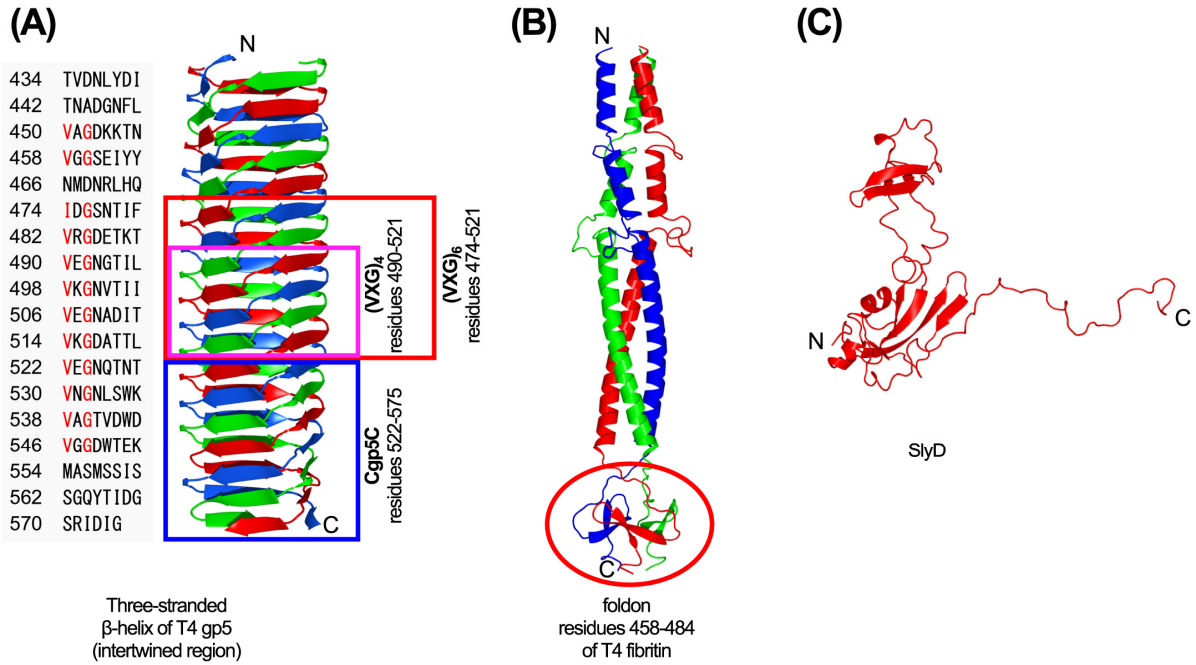

Figure 4-2. The building blocks and the protein fragments used for the construction of artificial three-stranded β-helical polypeptides. (**A**) The crystal structure of three-stranded β-helix of T4 gp5 (PDB 1k28) and its amino acid sequence. The building blocks of six repeats ($(VXG)_6$) and four repeats ($(VXG)_4$) are indicated with the red box and the pink box, respectively. The C-terminal hydrophobic core (Cgp5C or "C") is indicated with the blue box. (**B**) The crystal structure of the C-terminal region of T4 fibritin (PDB 1AA0). The red circle indicates the C-terminal foldon domain (27 amino acids). (**C**) The solution structure *E. coli* SlyD protein (PDB 2KFW, model 1). The C-terminus and N-terminus are indicated with C and N, respectively (A-C).

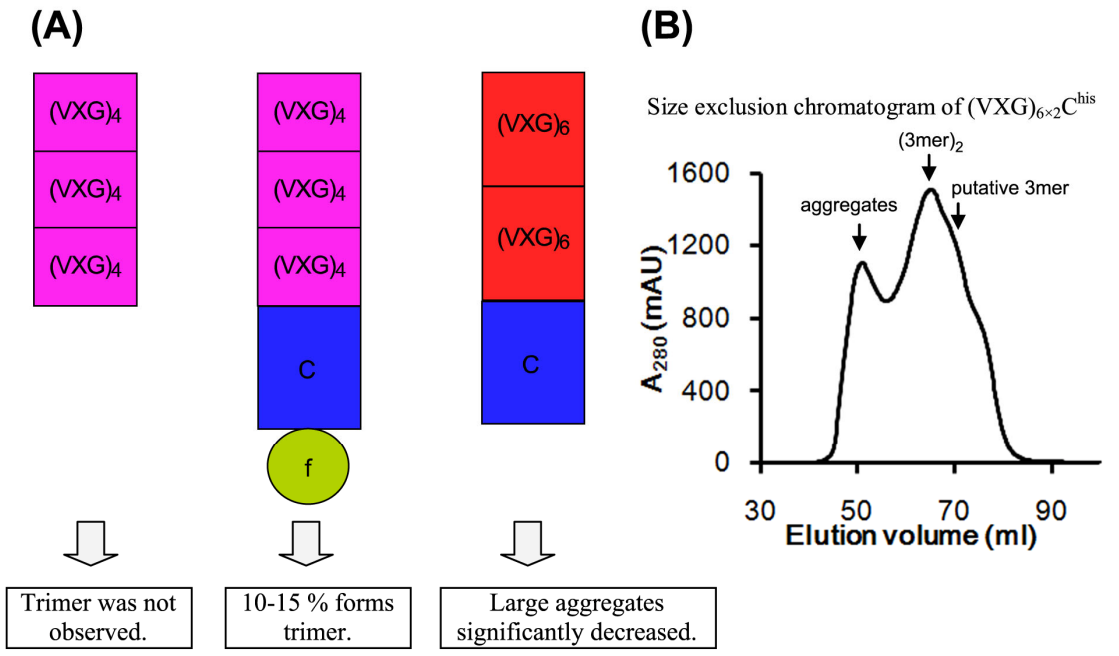

Figure 4-3. The previously synthesized artificial three-stranded β-helices. (VXG) in the figure represents (VXGXXXXX). (**A**) The cartoon of the artificial three-stranded β-helices synthesized in our laboratory. The specifications of each construct are indicated in the box below the figure. $(VXG)_4$, $(VXG)_6$, C and f stand for the building block of four repeat, the building block of six repeat, the C-terminal hydrophobic core of T4 gp5 and foldon, respectively (see Figure 4-2). (**B**) The size exclusion chromatogram of $(VXG)_{6\times2}C^{his}$ (*66*). The peak of aggregates, dimer of trimer and putative trimer are indicated.

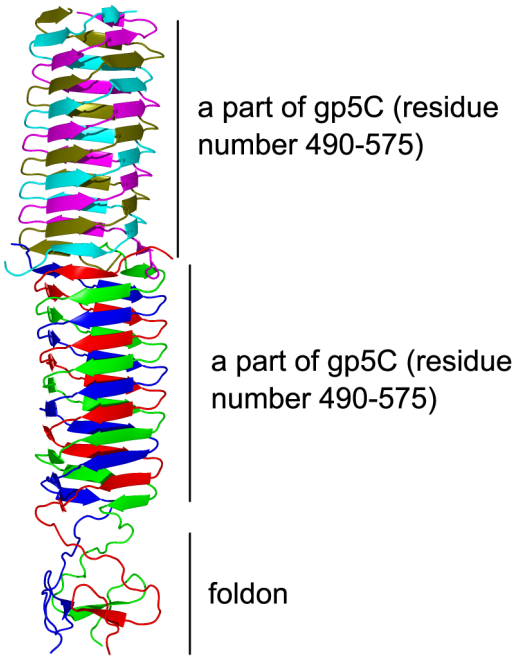

Figure 4-4. Crystal structure of a part of T4 gp5C fused with foldon in dimer of trimer state (PDB 3A1M) (*11*). The two trimers of a part of gp5C–foldon are associating in head to head manner via the N-terminus. One trimeric gp5C–foldon molecule is drawn in red, green, blue and the other gp5C–foldon molecule is drawn in light blue, purple and dark yellow. The foldon of one molecule is not displayed because it was disordered in the crystal.

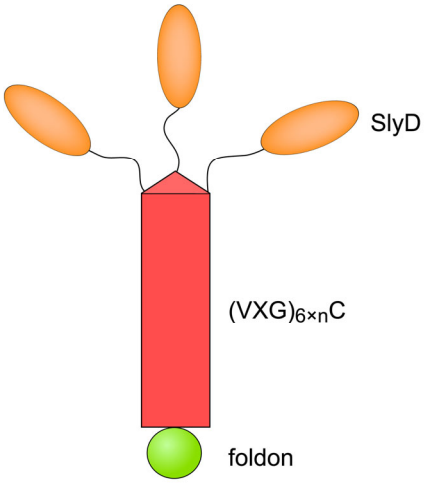

Figure 4-5. The schematic representation of the newly constructed ^{his}SlyD-$(VXG)_{6\times n}$Cf. SlyD (orange ellipse) was added to N-terminus in order to inhibit the association mediated via the N-terminus. The foldon was added to the C-terminus as it decreased the aggregation to some extent in $(VXG)_{4\times 3}$Cf presumably by promoting the trimerization and capping the β-helix (*65*).

4-2 Materials and Methods

Media

LB media was used for all *E. coli* cultivation. The composition is described in Materials and Methods section of Chapter 2.

Plasmid construction

The over-expression vector of ^{his}SlyD-$(VXG)_{6\times1}$Cf was constructed as follows. First, the foldon (residues 457-483 of T4 gp wac, Figure 4-2) with a stop codon was made by inverse PCR using PCR KOD -Plus- Mutagenesis Kit (YOYOBO CO, LTD) and the primer pairs of foldon-STP(+)iPCR and foldon-STP(-)iPCR (Table 4-1). pMTpf 3-5 which is the derivative of pUC19 containing foldon (*65*) was used as a template DNA. The DNA fragment of foldon with stop codon was cut out by EcoRI - SalI and inserted to the EcoRI and XhoI sites of pSL(b) vector. pSL(b) vector was made from pSL vector (see Chapter 2) by changing the reading frame just before the multiple cloning site. The resulting vector was digested by EcoRI - XhoI and the DNA fragment of T4 gp5 residues 522–575 (Cgp5C, Figure 4-2), which was cut out from pMTpC (*65*) by EcoRI - SalI, was inserted. Then the resulting vector was digested by EcoRI - XhoI and the fragment of T4 gp5 residues Val474-Leu521 ($(VXG)_{6\times1}$, Figure 4-2), which was cut out from pKA150 (*66*) by EcoRI - SalI, was inserted. The resulting over-expression vector of ^{his}SlyD-$(VXG)_{6\times1}$Cf was designated as pKN8-1. The over-expression vector of ^{his}SlyD-$(VXG)_{6\times2}$Cf was constructed by inserting one more $(VXG)_{6\times1}$ coding DNA fragment into pKN8-1 vector in the same way. The resulting vector was named pKN9-1.

The DNA sequence of foldon with stop codon was confirmed by DNA sequencing. Due to the block-by-block vector construction strategy, Ile474 and 475Asp of the first $(VXG)_6$ block are changed to Leu-Glu and those of second $(VXG)_6$ block are changed to Val-Glu.

Protein expression

For the expression of ^{his}SlyD-$(VXG)_{6\times1}$Cf and ^{his}SlyD-$(VXG)_{6\times2}$Cf, BL21(DE3) was transformed by pKN8-1 or pKN9-1 and cultivated at 37°C in LB medium supplemented with 60 ug/ml kanamycin. When OD_{660} was 0.5-0.6, the

expression was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) at a final concentration of 1 mM and further cultivated at 30°C for 10 h. Cells were harvested by centrifugation at 2,500 ×g for 10 min using Himac CR22G centrifuge with R9AF rotor (Hitachi Koki).

Protein Purification

The purification procedure of ^{his}SlyD-$(VXG)_{6\times1}$Cf and ^{his}SlyD-$(VXG)_{6\times2}$Cf was same as that of VgrG1 (see Materials and Methods section of Chapter 2) but the buffer D (50 mM Tris pH 8.0, 100 mM NaCl, 25 mM imidazole) was used instead of buffer A (50 mM Tris pH8.0, 100 mM NaCl, 20 mM imidazole).

*$(VXG)_{6\times1}$Cf and *$(VXG)_{6\times2}$Cf, which are the products of ^{his}SlyD-$(VXG)_{6\times1}$Cf and ^{his}SlyD-$(VXG)_{6\times2}$Cf after ^{his}SlyD removal, were purified as follows. The cell expressing ^{his}SlyD-$(VXG)_{6\times1}$Cf or ^{his}SlyD-$(VXG)_{6\times2}$Cf was suspended in 10× volume of buffer D and lysed by sonication in the presence of 1 mM phenylmethylsulfonyl fluoride (PMSF). The cell lysate was centrifuged at 20,000 ×g (R18A rotor with Himac CR22G centrifuge, Hitachi Koki) for 20 min and the supernatant was loaded to HisTrap FF column (GE Healthcare) which was equilibrated with buffer D. ^{his}SlyD-$(VXG)_{6\times1}$Cf and ^{his}SlyD-$(VXG)_{6\times2}$Cf were eluted by the imidazole gradient of 25 to 500 mM and EDTA was added to the eluted fractions to a final concentration of 5 mM. The fractions containing the desired protein were collected and applied to HiTrap Q HP column (GE Healthcare) which was equilibrated with buffer B. The proteins were eluted with NaCl gradient of 0 to 1 M. Then the fractions containing the desired protein were collected and TEV protease was added on the weight ratio of 10 to 1 (^{his}SlyD-$(VXG)_{6\times1}$Cf or ^{his}SlyD-$(VXG)_{6\times2}$Cf to TEV protease). After the incubation at 30°C for 4 h, the solution was applied to Histrap FF column in order to remove ^{his}SlyD and TEV protease which has a hexa-histidine tag. The flow through fractions which contain *$(VXG)_{6\times1}$Cf or *$(VXG)_{6\times2}$Cf were mixed and further purified by Hitrap Q HP column (GE Healthcare) and Hiload 16/60 Superdex200 prep grade column (GE Healthcare). The purification was performed either room temperature or 4°C.

*Analytical size exclusion chromatography (SEC) of ^{his}SlyD-cleaved *$(VXG)_{6\times2}$Cf*

For the analytical SEC of *$(VXG)_{6\times2}$Cf, 160 µl of the SEC fractions of ^{his}SlyD-$(VXG)_{6\times2}$Cf (ranging from 0.6 - 1.2 mg/ml) were mixed with TEV protease on the weight ratio of 10:1 (^{his}SlyD-$(VXG)_{6\times2}$Cf to TEV protease) and incubated at 30°C for 4h. The 150 ul of the reaction solutions were analyzed by Superdex200 HR 10/30 (GE Healthcare).

Limited proteolysis of partially unfolded species by trypsin

For the limited proteolysis of partially unfolded species of *$(VXG)_{6\times2}$Cf, 40 ul of purified *$(VXG)_{6\times2}$Cf (0.1 - 0.2 mg/ml) was mixed with trypsin on the weight ratio of 100:1 (*$(VXG)_{6\times2}$Cf to trypsin). After the incubation at 37°C for 6 h, the proteolysis was quenched by the addition of Tosyl-Lys-chloromethyl ketone (TLCK) at a final concentration of 0. 1 mM and the products were analyzed by 10% native PAGE.

*Purification of trypsin treated *$(VXG)_{6\times2}$Cf*

For the purification of trypsin treated *$(VXG)_{6\times2}$Cf, *$(VXG)_{6\times2}$Cf was purified as described above and the size exclusion chromatography fractions containing *$(VXG)_{6\times2}$Cf were collected and trypsin was added on the weight ratio of 100 to 1 (*$(VXG)_{6\times2}$Cf to trypsin). After the incubation at 37°C for 6 h, the proteolysis was stopped by the addition of Tosyl-Lys-chloromethyl ketone (TLCK) at a final concentration of 0.1 mM and the trypsin-resistant *$(VXG)_{6\times2}$Cf was further purified by Hitrap Q HP column (GE Healthcare) and Superdex200 HR 10/30 column (GE Healthcare).

SDS-PAGE

SDS-PAGE was carried out in the same protocol as Chapter 2 but the samples were not heat denatured.

Native PAGE

For the Native PAGE of *$(VXG)_{6\times2}$Cf, the samples and 10% polyacrylamide gel were prepared without SDS and DTT (*67*). The electrophoresis was carried out at 10 mM constant current at 4°C. The polyacrylamide gel was stained with 0.1% Coomassie

Brilliant Blue G-250 in 10% acetic acid.

Analytical Ultracentrifugation

Sedimentation velocity analysis was carried out in the same protocol as Chapter 2 except for following things. The sedimentation velocity data of ^{his}SlyD-$(VXG)_{6\times1}$Cf and ^{his}SlyD-$(VXG)_{6\times2}$Cf were collected with the rotor speed of 45,000 rpm, those of *$(VXG)_{6\times1}$Cf and *$(VXG)_{6\times2}$Cf were collected at 40,000 rpm and those of trypsin treated *$(VXG)_{6\times2}$Cf were collected at 45,000 rpm. Moving boundaries were recorded at the wave length of 280 nm for ^{his}SlyD-$(VXG)_{6\times1}$Cf, ^{his}SlyD-$(VXG)_{6\times2}$Cf, *$(VXG)_{6\times1}$Cf and *$(VXG)_{6\times2}$Cf and at the wavelength of 230 nm for trypsin treated *$(VXG)_{6\times2}$Cf.

Table 4-1 Primers for the construction of foldon with a stop codon.

Name	Nucleotide Sequence (5'-3') [a]	Description
foldon-STP(+)iPCR	TAAGTCGACCTGCAGGCATGCAAGC	Inverse primer for the construction of foldon with a stop codon.
foldon-STP(-)iPCR	TAAAAAGGTAGAAAGCAATACCC	Inverse primer for the construction of foldon with a stop codon.

[a] Stop codon is underlined.

4-3 Results

Plasmid construction

In order to prevent the interactions between exposed N terminus of the three-stranded β-helix and promote the trimerization, the over-expression vectors of SlyD and foldon fused artificial three-stranded β-helix, namely, ^{his}SlyD-$(VXG)_{6\times1}$Cf (pKN8-1) and ^{his}SlyD-$(VXG)_{6\times2}$Cf (pKN9-1), were constructed. The foldon is a polypeptide consisting of residues 457-483 of gp wac, which is essential for folding of T4 fibritin (*68*) and has also shown to promote the trimerization of artificial three-stranded β-helices (*65*). SlyD is known to facilitate solubilisation of fusion proteins as described in Chapter 2. The plasmid map of pKN8-1 and pKN9-1 is shown in Figure 4-6.

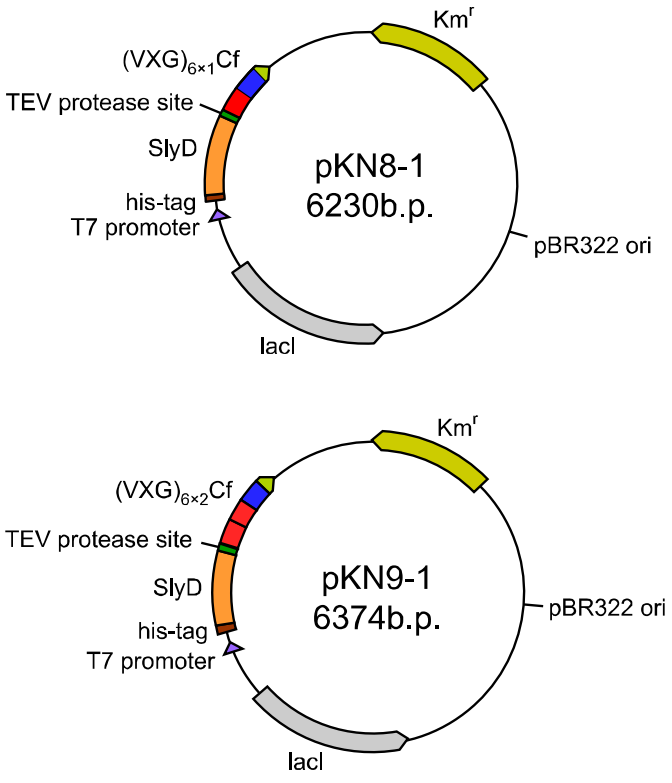

Figure 4-6. The plasmid map of the over-expression vector of ^{his}SlyD-$(VXG)_{6\times1}$Cf (pKN8-1) and ^{his}SlyD-$(VXG)_{6\times2}$Cf (pKN9-1).

Expression, purification and sedimentation velocity analysis of ^{his}SlyD-(VXG)$_{6\times1}$Cf

^{his}SlyD-(VXG)$_{6\times1}$Cf was expressed in BL21(DE3) and purified by Ni affinity chromatography, anion exchange chromatography and size exclusion chromatography (see Materials and Methods). ^{his}SlyD-(VXG)$_{6\times1}$Cf eluted from the size exclusion chromatography with a major peak about 52 ml and several minor peaks before and after it (Figure 4-7 (A)). The SDS-PAGE of the peak fraction of the size exclusion chromatography (SEC) showed multiple bands with higher molecular weight than monomers (monomer is 35.4 kDa) (Figure 4-7 (B)). The three-stranded β-helix of T4 is SDS-resistant, thus they are presumably SDS-resistant ^{his}SlyD-(VXG)$_{6\times1}$Cf. The sedimentation velocity analysis indicated that the major peak contains the major species with the sedimentation coefficient of 7.24 and minor species of the sedimentation coefficient of 5.03 (Figure 4-8). The molecular weights of 7.24 S species and 5.03 S species were estimated to be 190,000 and 110,000, respectively thus they are considered to be dimer of trimer and trimer (calculated molecular mass of monomer is 35,400).

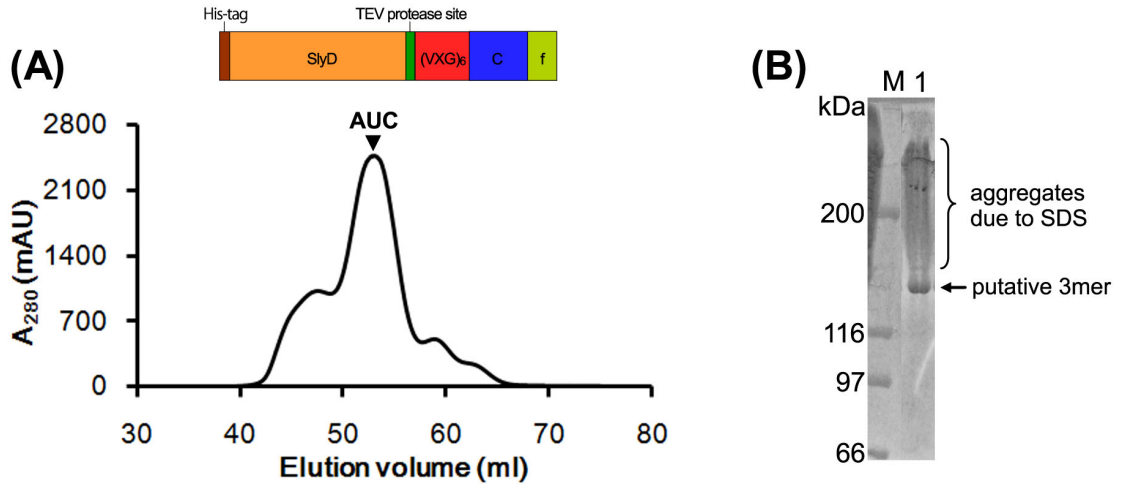

Figure 4-7. Purification of ^{his}SlyD-(VXG)$_{6\times1}$Cf. (**A**) Size exclusion chromatogram. The chromatography was performed with Hiload 16/60 Superdex200 prep grade column (GE Healthcare). (**B**) 6% SDS-PAGE of the peak fraction of size exclusion chromatography. Lane M, marker; lane 1, the peak fraction of size exclusion chromatography. The SDS-PAGE samples were prepared only by adding the sample loading buffer (the proteins were not denatured by heat). The polyacrylamide gels were stained with Coomassie brilliant blue.

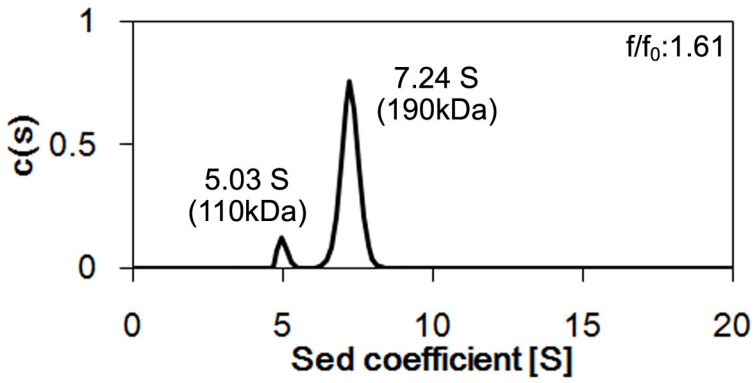

Figure 4-8. Sedimentation velocity analysis of ^{his}SlyD-$(VXG)_{6\times1}$Cf. The sedimentation velocity data was analyzed by SEDFIT. The moving boundaries were measured at 280 nm at 20°C. The rotor speed was 45,000 rpm. The molecular weights were obtained by converting *c(s)* to *c(M)* as implemented in SEDFIT.

*Purification and sedimentation velocity analysis of *(VXG)$_{6\times1}$Cf*

^{his}SlyD-(VXG)$_{6\times1}$Cf, which had been purified by Ni affinity chromatography and anion exchange chromatography, was digested by TEV protease (cleavage of ^{his}SlyD) then the *(VXG)$_{6\times1}$Cf was further purified by Ni affinity chromatography (removal of ^{his}SlyD and TEV protease), anion exchange chromatography and SEC. *(VXG)$_{6\times1}$Cf eluted from SEC column as an almost single peak (Figure 4-9 (A)). SDS-PAGE shows two bands with higher molecular weight than monomers but they are considered to be SDS-resistant *(VXG)$_{6\times1}$Cf (Figure 4-9 (B)).

The sedimentation velocity analysis of the peak fraction indicated that it contains only a single species with the sedimentation coefficient of 4.48 S (Figure 4-10). The molecular weight of 4.48 S was estimated to be 91,700 thus it is considered to be the dimer of trimer (calculated molecular mass of monomer is 15,100).

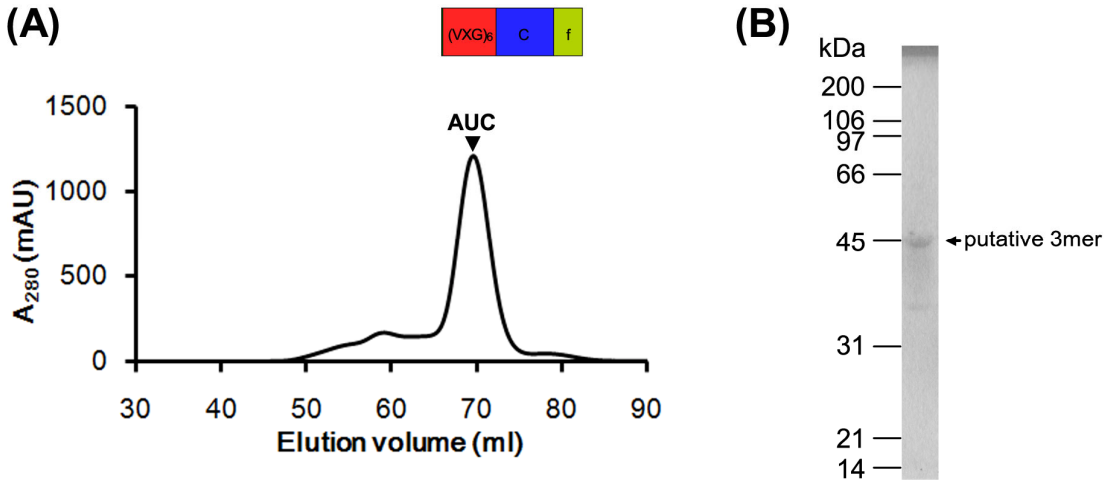

Figure 4-9. Purification of *(VXG)$_{6\times1}$Cf. (**A**) Size exclusion chromatogram. The chromatography was performed with Hiload 16/60 Superdex200 prep grade column (GE Healthcare). (**B**) 12% SDS-PAGE of the peak fraction of size exclusion chromatography. The SDS-PAGE samples were prepared only by adding the sample loading buffer (the proteins were not denatured by heat). The polyacrylamide gels were stained with Coomassie brilliant blue.

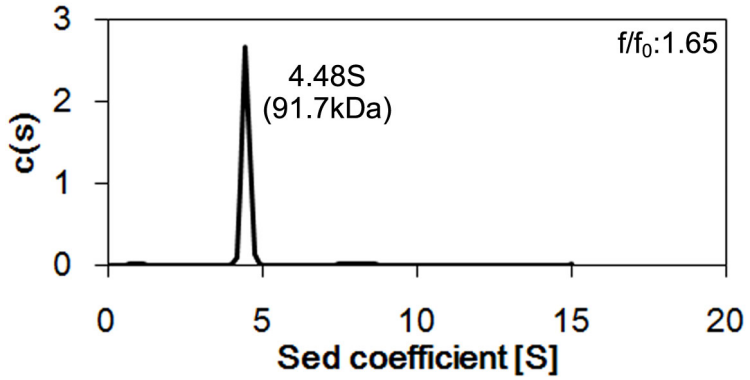

Figure 4-10. Sedimentation velocity analysis of *$(VXG)_{6\times1}$Cf. The sedimentation velocity data was analyzed by SEDFIT. The moving boundaries were measured at 280 nm at 20°C. The rotor speed was 40,000 rpm. The molecular weights were obtained by converting *c(s)* to *c(M)* as implemented in SEDFIT.

Expression and purification of ^{his}SlyD-(VXG)$_{6\times2}$Cf

^{his}SlyD-(VXG)$_{6\times2}$Cf was expressed in BL21(DE3) and purified by Ni affinity chromatography, anion exchange chromatography and size exclusion chromatography (see Materials and Methods) (Figure 4-11 A-D). ^{his}SlyD-(VXG)$_{6\times2}$Cf eluted from the size exclusion chromatography column as two peaks and a shoulder at ~52ml (Figure 4-11 (C)). The SDS-PAGE of the fractions from the SEC showed multiple bands with higher molecular weight than monomers (Figure 4-10 (D)). They are presumably SDS-resistant ^{his}SlyD-(VXG)$_{6\times2}$Cf as observed in ^{his}SlyD-(VXG)$_{6\times1}$Cf. The fractionated samples were analyzed by AUC as described in the next section.

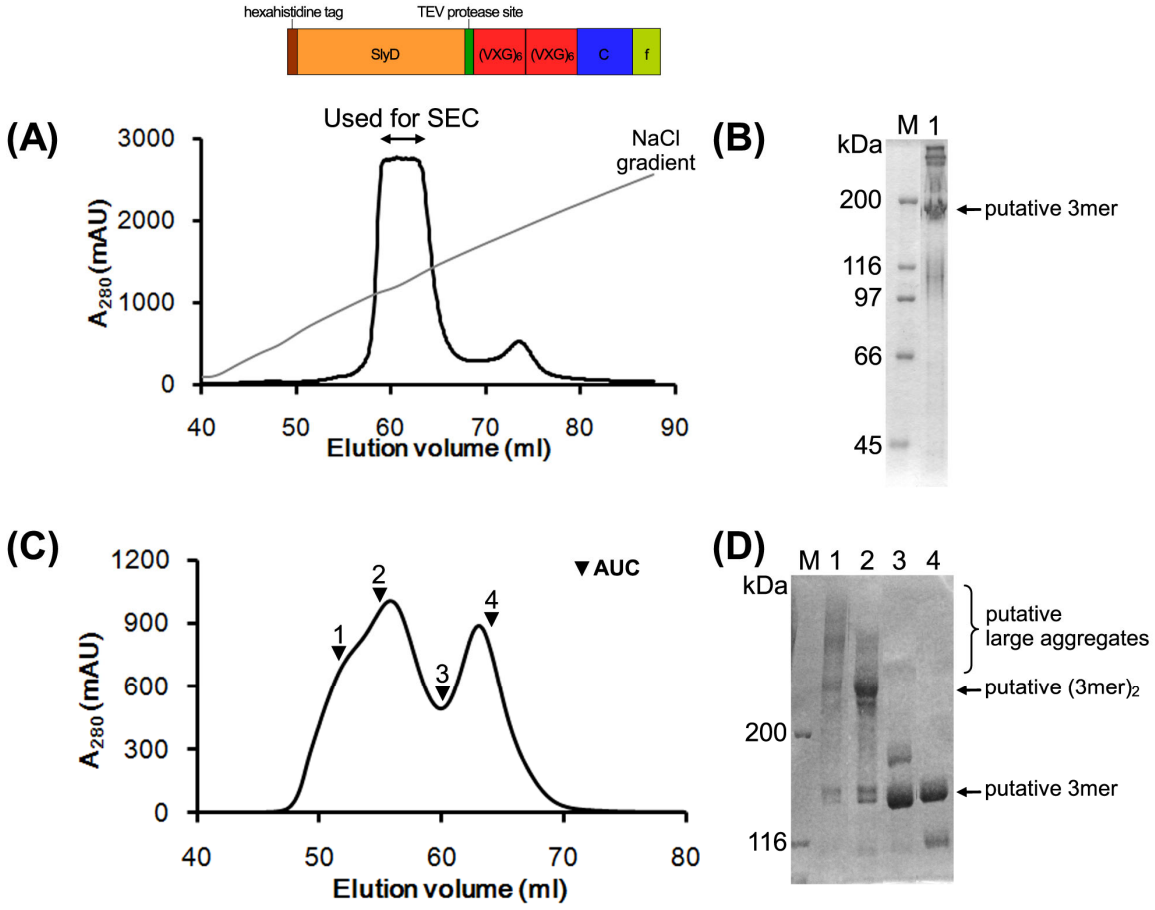

Figure 4-11. Purification of ^{his}SlyD-$(VXG)_{6\times2}$Cf. (**A**) Anion exchange chromatogram. The thick black line indicates the absorbance at 280 nm and the thin gray line indicates the concentration of NaCl (from 0 M to 1 M). (**B**) 8% SDS-PAGE of the peak top fraction of anion exchange chromatography. Lane M, Marker; lane1, the peak top fraction. (**C**) Size exclusion chromatogram. The chromatography was performed with Hiload 16/60 Superdex200 prep grade column (GE Healthcare). The fractions indicated 1-4 were subjected to AUC-SV analysis. (**D**) 5% SDS-PAGE of the fractions of size exclusion chromatography. Lane M, marker; lane 1-4, the fractions of size exclusion chromatography in (C). The SDS-PAGE samples were prepared only by adding the sample loading buffer (the proteins were not denatured by heat). The polyacrylamide gels were stained with Coomassie brilliant blue.

Sedimentation velocity analysis of ^{his}SlyD-(VXG)$_{6\times2}$Cf

In order to examine the oligomerization state of ^{his}SlyD-(VXG)$_{6\times2}$Cf, AUC-SV was carried out with the fractions from the SEC (Figure 4-11 (C)). The fractions of the shoulder of the first peak (fraction 1), the peak top fraction of the first peak (fraction 2), a fraction between the first and the second peak (fraction 3) and the peak top fraction of the second peak (fraction 4) were examined. The result showed that fraction 1 contained large aggregates with the sedimentation coefficient above 8.71 and fraction 2 contained the 7.13 S species and the small amount of 4.74 S species (Figure 4-12). The fraction 3 contained two molecular species with the sedimentation coefficient of 5.16 and 7.25 and fraction 4 contained the species with the sedimentation coefficient of 4.75 (Figure 4-12). The molecular weights of 7.13 S species in the first peak and the 4.75 S species in the second peak were estimated to be 206,000 and 111,000, respectively thus they are considered to be dimer of trimer and trimer (calculated molecular mass of monomer is 40,400). The molecular weight estimation from AUC-SV generally contains ~10% of error and the molecular weight of 206,000, which is 15% lower than the calculated molecular weight of a dimer of trimer (242,400), is likely within the error.

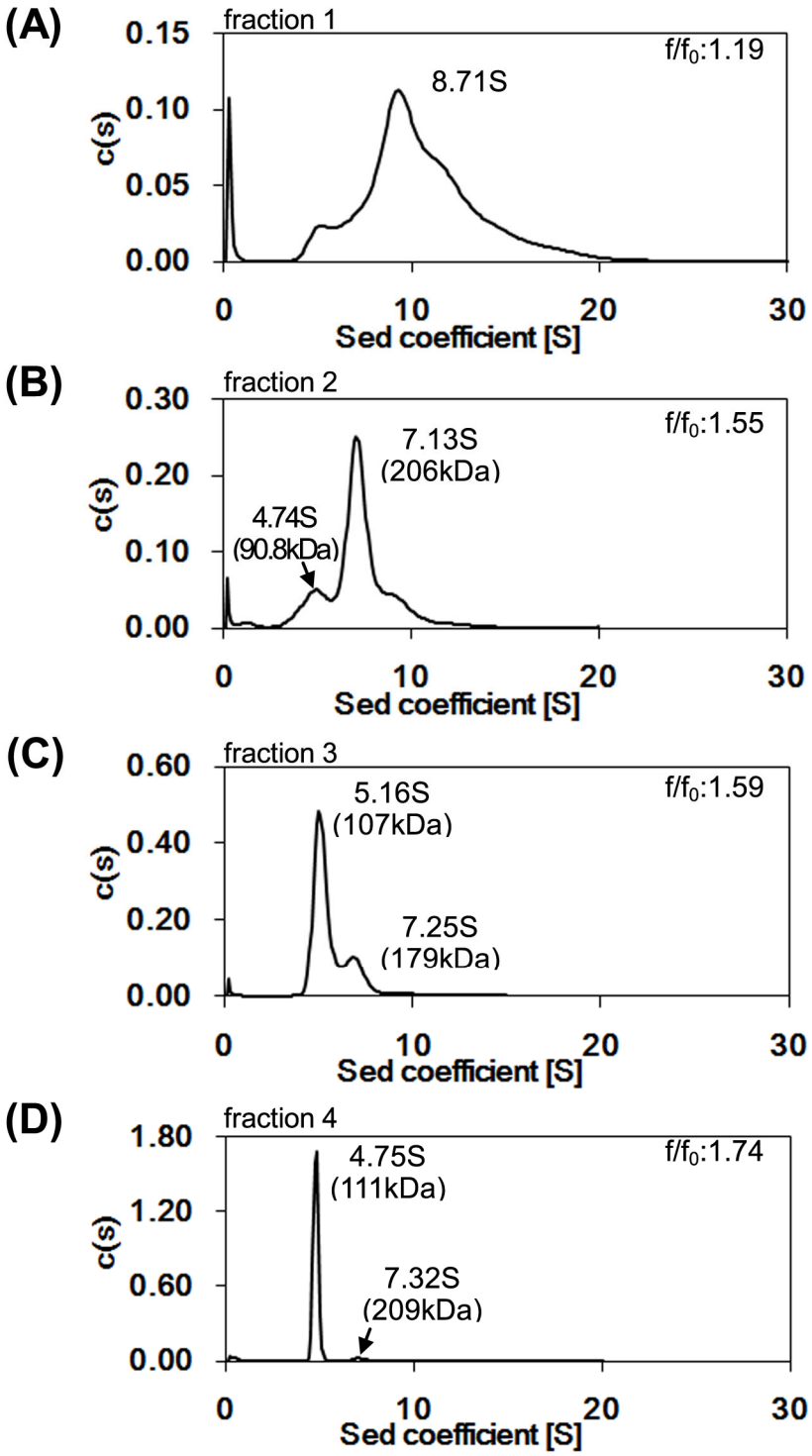

Figure 4-12. Sedimentation velocity analysis of ^{his}SlyD-$(VXG)_{6\times2}$Cf. The sedimentation velocity data of fractions 1-4 of SEC in Figure 4-11 (C) were analyzed by SEDFIT. The moving boundaries were measured at 280nm at 20°C. The rotor speed was 45,000 rpm. The molecular weights were obtained by converting *c(s)* to *c(M)* or analyzing in *c(s, f/f₀)* mode as implemented in SEDFIT.

^{his}SlyD-(VXG)$_{6\times2}$Cf contains partially unfolded species.

In order to examine the property of ^{his}SlyD-(VXG)$_{6\times2}$Cf in more detail, ^{his}SlyD was cleaved off by TEV protease (Figure 4-13 (A)) and the resulting *(VXG)$_{6\times2}$Cf was analyzed by analytical size exclusion chromatography (superdex200 HR 10/30 (GE Healthcare)) (Figure 4-13 (B)). The fractions containing large aggregates (fraction 1), dimer of trimer (fraction 2), mixture of dimer of trimer and trimer (fraction 3) and trimer (fraction 4) in Figure 4-11 (C) were used after the depletion of ^{his}SlyD. The result indicated that fraction 1 gave the molecular species which eluted at 10.3 ml and fraction 2 gave the molecular species which eluted at 11.4 ml (major) and 11.3 ml (minor) (Figure 4-13 (B)). Notably, fraction 3 gave the molecular species, which eluted at 11.4 ml and 13.0 ml, 0.3 ml earlier compared with fraction 2, while fraction 4 gave the molecular species, which eluted at 11.4 ml and 13.3 ml again (Figure 4-13 (B)). The molecular species eluted at 13.0 ml has a slightly bigger stokes radius than those of the species eluted at 13.3 ml. It is also notable that the fraction 4 gave only a single peak if it was not treated with TEV protease (Figure 4-13 (C)) thus the molecular species which eluted at 11.4 ml, came from the trimeric ^{his}SlyD-(VXG)$_{6\times2}$Cf.

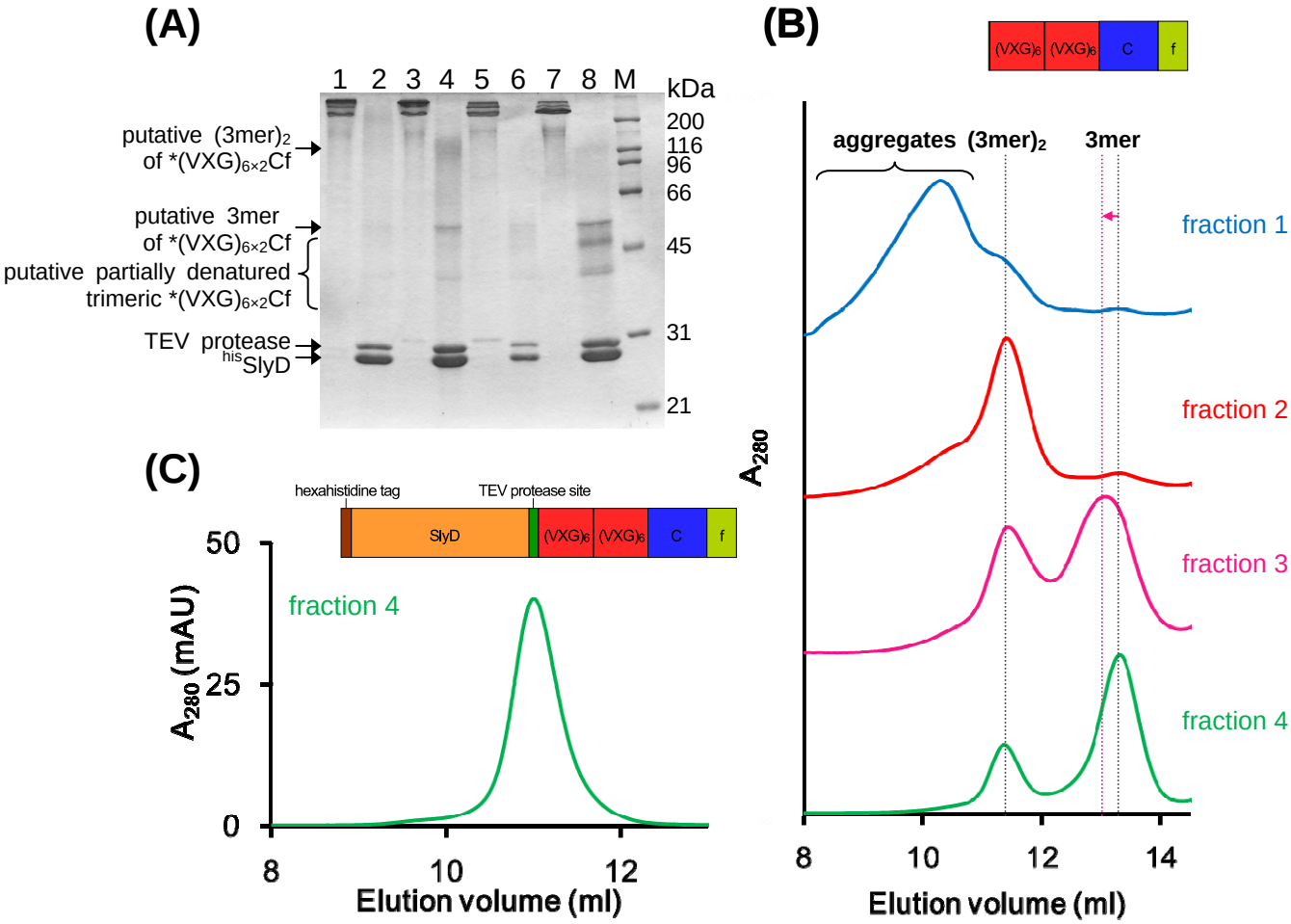

Figure 4-13. Analytical size exclusion chromatography of ^{his}SlyD-(VXG)$_{6\times2}$Cf after the cleavage of ^{his}SlyD. (**A**) Cleavage of ^{his}SlyD from ^{his}SlyD-(VXG)$_{6\times2}$Cf by TEV protease. Lane 1 and 2, before and after the removal of ^{his}SlyD of fraction 1; lane 3 and 4, before and after the removal of ^{his}SlyD of fraction 2; lane 5 and 6, before and after the removal of ^{his}SlyD of fraction 3; lane 7 and 8, before and after the removal of ^{his}SlyD of fraction 4; M, marker. The bands of TEV protease and ^{his}SlyD are indicated. (**B**) The analytical size exclusion chromatography of the TEV protease treated ^{his}SlyD-(VXG)$_{6\times2}$Cf. The oligomerization states of *(VXG)$_{6\times2}$Cf are indicated for each peak. ^{his}SlyD and TEV protease were eluted at ~16 ml (not displayed in the figure). Note that the second peak from fraction 3 (pink) is ~0.3 ml earlier than that of fraction 2 (red) and fraction 4 (green). (**C**) The analytical size exclusion chromatography of fraction 4 without ^{his}SlyD removal. Note that fraction 4 gave only single peak if ^{his}SlyD was present.

*Sedimentation velocity analysis and native PAGE of *(VXG)$_{6\times2}$Cf*

In order to examine the oligomerization state and structural homogeneity of *(VXG)$_{6\times2}$Cf, it was purified from anion exchange chromatography sample (Figure 4-11 (A-B)) by TEV protease digestion, Ni affinity chromatography (removal of ^{his}SlyD and TEV protease), anion exchange chromatography and size exclusion chromatography (Figure 4-14). The size exclusion chromatogram of *(VXG)$_{6\times2}$Cf was similar to that of ^{his}SlyD-(VXG)$_{6\times2}$Cf but the shoulder of the first peak of ^{his}SlyD-(VXG)$_{6\times2}$Cf was separated as an another peak in *(VXG)$_{6\times2}$Cf (Figure 4-14 (A)).

The sedimentation velocity analysis indicated that the first peak of the size exclusion chromatography contains the molecular species with the sedimentation coefficient above 7.1 (Figure 4-15 (A)) and the second peak of the size exclusion chromatography contains the major molecular species of 5.08 S and minor molecular species of 3.41 S and 7.77 S (Figure 4-15 (B)). The third peak of the size exclusion chromatography mainly contains the molecular species with the sedimentation coefficient with 3.50 (Figure 4-15 (C)). The conversion of *c(s)* to *c(M)* indicated that the 7.1 S species in the first peak, the 5.08 S species in the second peak and the 3.50 S species in the third peak are 199,000, 118,000 and 60,600, respectively thus they are considered to be the aggregates above the trimer of trimer, the dimer of trimer and trimer, respectively (calculated molecular mass of monomer is 20,000).

The native PAGE was carried out for the same fractions as AUC-SV. The *(VXG)$_{6\times1}$Cf in dimer of trimer was also analyzed in the same polyacrylamide gel for comparison. The result indicated that the fraction containing the dimer of trimer of *(VXG)$_{6\times2}$Cf gave both a clear band and smear bands but the fractions containing large aggregates or trimer gave only smear bands (Figure 4-15). The molecular species which appears as smear bands on native PAGE would have many conformations. The clear band of *(VXG)$_{6\times2}$ was slightly upper than that of (VXG)$_{6\times1}$ presumably due to the increased molar mass of one building block.

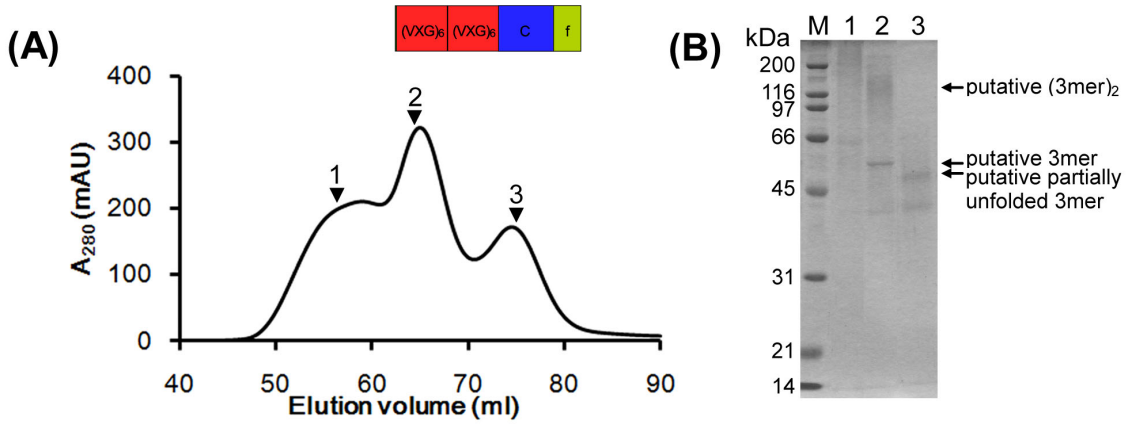

Figure 4-14. The purification of *$(VXG)_{6\times2}$Cf. (**A**) Size exclusion chromatogram. The chromatography was performed with Hiload 16/60 Superdex200 prep grade column (GE Healthcare). The fractions indicated 1-3 were subjected to AUC-SV analysis. (**B**) 12% SDS-PAGE of purified *$(VXG)_{6\times2}$Cf. Lane M, molecular weight marker; lane 1, the peak top of the first peak; lane 2, the peak top of the second peak; lane 3, the peak top of the third peak. The polyacrylamide gel was stained with Coomassie brilliant blue.

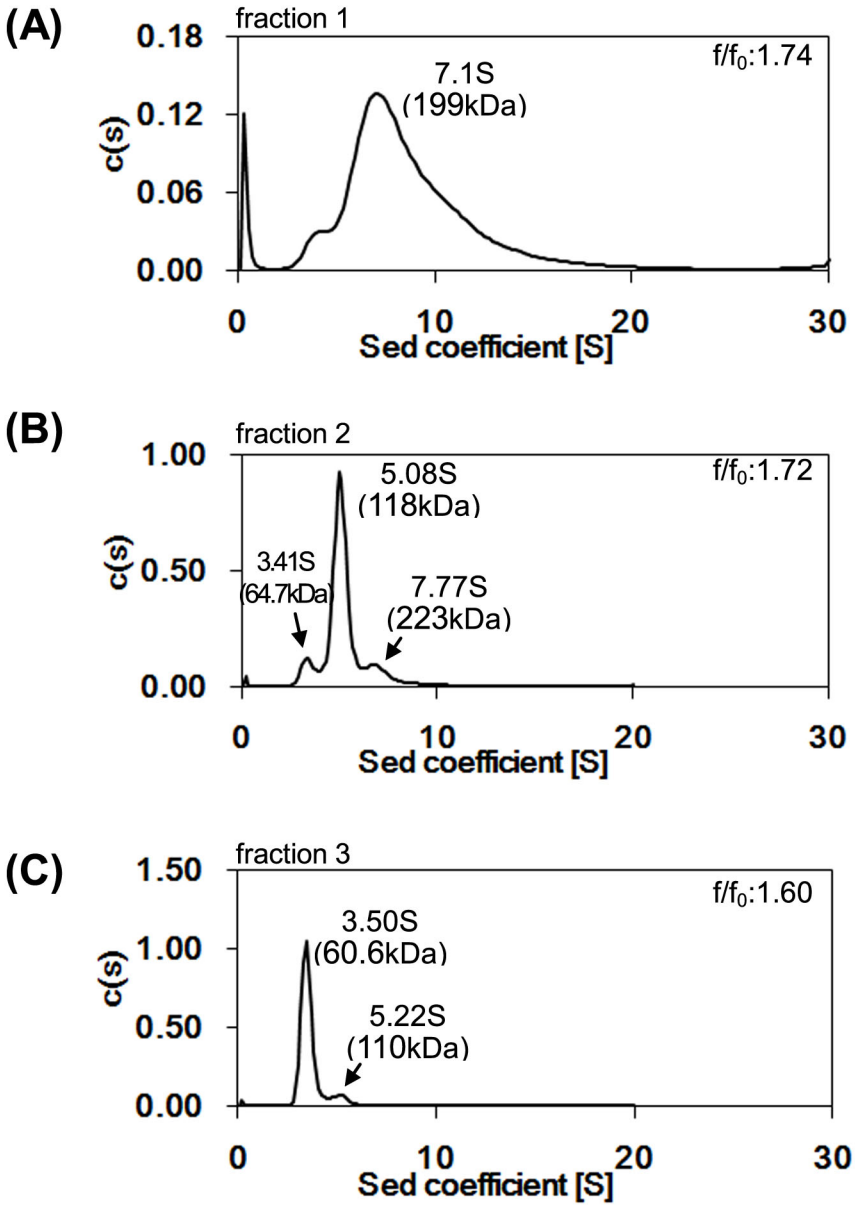

Figure 4-15. Sedimentation velocity analysis of *$(VXG)_{6\times2}$Cf. The sedimentation velocity data of fractions 1-3 in Figure 4-14 (A) were analyzed by SEDFIT. The moving boundaries were measured at 280 nm at 20°C. The rotor speed was 40,000 rpm. The molecular weight was obtained by converting *c(s)* to *c(M)*.

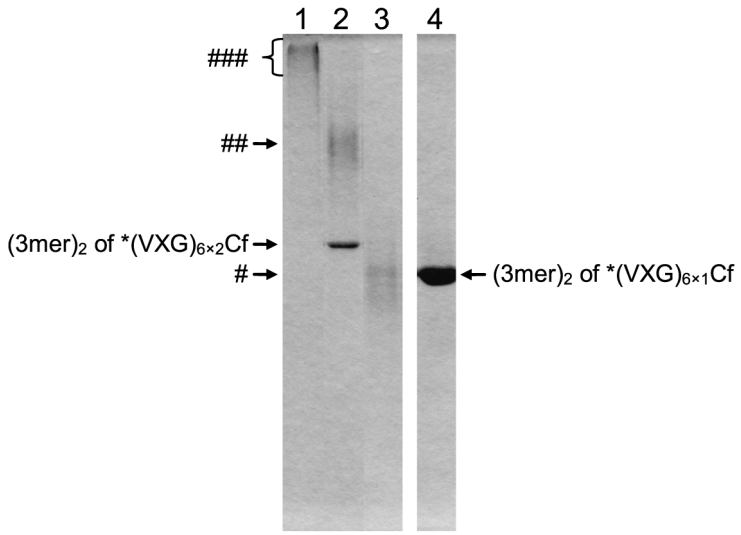

Figure 4-16. Native PAGE of *$(VXG)_{6\times2}$Cf and *$(VXG)_{6\times1}$Cf. Lane 1, the peak top of the first peak of size exclusion chromatography; lane 2, the peak top of the second peak; lane 3, the peak top of the third peak; lane 4, *$(VXG)_{6\times1}$Cf in dimer of trimer (for comparison). The 10% polyacrylamide gel was stained with Coomassie brilliant blue. "###" indicates aggregates of partially unfolded *$(VXG)_{6\times2}$Cf, "##" indicates dimer of trimer of partially unfolded *$(VXG)_{6\times2}$Cf and "#" indicates trimer of partially unfolded *$(VXG)_{6\times2}$Cf.

*Purification and characterization of trypsin-resistant *$(VXG)_{6\times2}$Cf*

In Chapter 2, it was revealed that the disordered region is sensitive to trypsin but the three-stranded β-helix region is resistant to trypsin. Therefore I hypothesized that the fully folded artificial three-stranded β-helix would be resistant to trypsin, but partially unfolded species are not. To examine this hypothesis, *$(VXG)_{6\times2}$Cf was subjected to proteolysis by trypsin and the product was analyzed by native PAGE. The result indicated that the molecular species which gave smear bands were digested by trypsin but the molecular species which gave a clear band was trypsin-resistant (Figure 4-17).

The trypsin resistant species were purified by anion exchange chromatography and size exclusion chromatography as described in Materials and Methods 4-2 (Figure 4-18). The native PAGE of purified sample showed an almost single band, which indicates that *$(VXG)_{6\times2}$Cf was purified to near homogeneous (Figure 4-18 (B)). The AUC-SV of the same sample indicated that it contains 5.02 S species like those before trypsin treatment but ~3.3 S and ~7.3 S species significantly decreased (Figure 4-19), which also confirmed that trypsin treatment made *$(VXG)_{6\times2}$Cf more homogeneous.

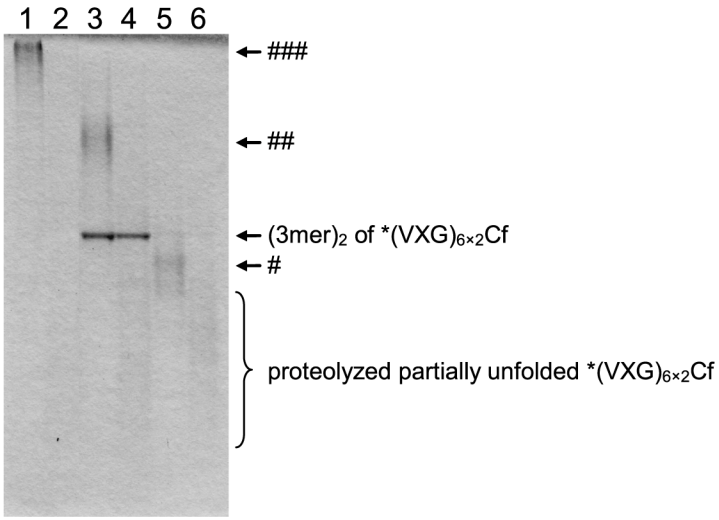

Figure 4-17. Native PAGE of trypsin treated *$(VXG)_{6\times2}$Cf. The peak fractions of the size exclusion chromatography were subjected to proteolysis by trypsin (37 °C for 6 h) and the products were analyzed by 10% native PAGE. Lane 1 and 2, before and after the trypsin digestion of the peak top of the first peak; lane 3 and 4, before and after the trypsin digestion of the peak top of the second peak; lane 5 and 6, before and after the trypsin digestion of the peak top of the third peak. The polyacrylamide gel was stained with Coomassie brilliant blue. "###" indicates aggregates of partially unfolded *$(VXG)_{6\times2}$Cf, "##" indicates dimer of trimer of partially unfolded *$(VXG)_{6\times2}$Cf and "#" indicates trimer of partially unfolded *$(VXG)_{6\times2}$Cf.

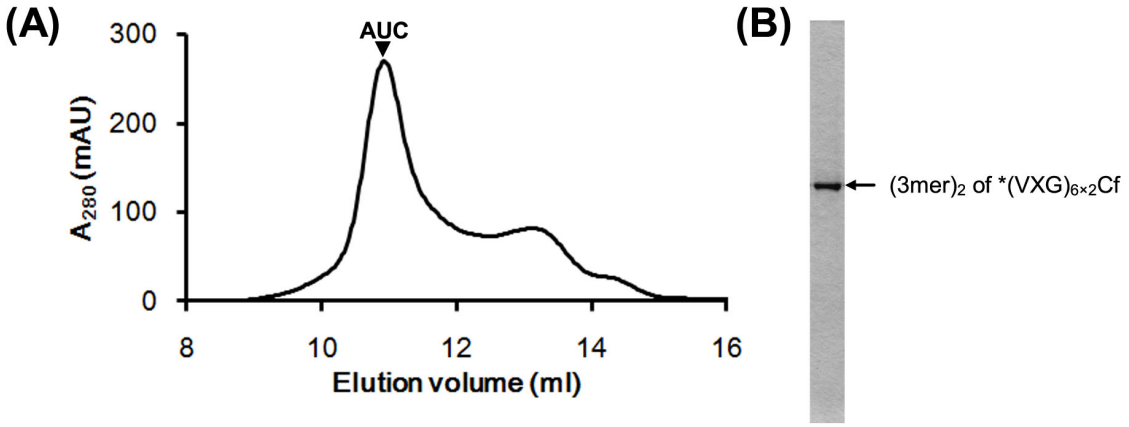

Figure 4-18. Purification of the trypsin-resistant *$(VXG)_{6\times2}$Cf. (**A**) Size exclusion chromatogram of trypsin treated *$(VXG)_{6\times2}$Cf. The chromatography was performed with Superdex200 HR 10/30 (GE Healthcare). The peak top fraction indicated with an arrowhead was subjected to AUC-SV analysis. (**B**) 10% native PAGE of the peak fraction. The polyacrylamide gel was stained with Coomassie brilliant blue.

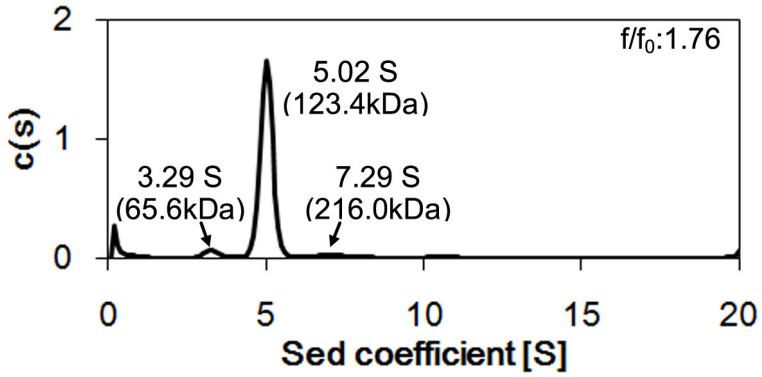

Figure 4-19. Sedimentation velocity analysis of trypsin-treated *$(VXG)_{6\times2}$Cf. The peak fraction indicated with "AUC" in Figure 4-18 was used for analysis. The moving boundaries were measured at 230nm at 20°C. The rotor speed was 45,000 rpm. The molecular weight was obtained by converting *c(s)* to *c(M)*.

4-4 Discussion

The three-stranded β-helix is a promising biomolecular nanomaterial for its high regularity and extraordinary stability. For the application of this remarkable structure, a protein nanotube has been developed in our laboratory, however the aggregation prone nature of engineered three-stranded β-helix prevented its practical use. To reduce the aggregation, I constructed the artificial three-stranded β-helix fused with SlyD and foldon, namely ^{his}SlyD-$(VXG)_{6\times1}$Cf and ^{his}SlyD-$(VXG)_{6\times2}$Cf.

The size exclusion chromatogram and the AUC-SV analysis indicated that *$(VXG)_{6\times1}$Cf mainly forms a dimer of trimer both in the presence and absence of SlyD. The dimerization of trimer is also reported in $(VXG)_{6\times1}$C (*66*) thus the fusion of SlyD does not seem to prevent the folding of the three-stranded β-helix.

The size exclusion chromatogram and the AUC-SV analysis of ^{his}SlyD-$(VXG)_{6\times2}$Cf indicated that they form a trimer and a dimer of trimer for main species. The analytical size exclusion chromatography of ^{his}SlyD-$(VXG)_{6\times2}$Cf after ^{his}SlyD removal revealed that a part of trimeric ^{his}SlyD-$(VXG)_{6\times2}$Cf gave the dimer of trimer of *$(VXG)_{6\times2}$Cf but the dimer of trimer of ^{his}SlyD-$(VXG)_{6\times2}$Cf did not give the trimeric *$(VXG)_{6\times2}$Cf. It indicates that $(VXG)_{6\times2}$Cf tends to dimerize and, once the dimer of trimer is formed, it doesn't dissociate. This dimerization was apparently mediated via the N-terminus, because they dimerized by removal of SlyD at the N-terminus. Several three-stranded β-helical proteins such as a part of phi92 gp138 (PDB 3PQH) (*16*), a part of T4 gp5C (PDB 3A1M) (*11*) and the *E. coli* O157 VgrG1C^{G561} (Chapter 2) (PDB 3WIT) associated in head to head manner via the exposed three-stranded β-helix in the crystal. Thus, $(VXG)_{6\times2}$Cf would associate in the similar way.

In both of ^{his}SlyD-$(VXG)_{6\times1}$Cf and ^{his}SlyD-$(VXG)_{6\times2}$Cf, the fusion of ^{his}SlyD could not completely prevent the dimerization of artificial three-stranded β-helical polypeptides. They contain a linker of 21 amino acids between SlyD and the three-stranded β-helix (including TEV protease site), hence it is possible that the linker was too long to keep the SlyD for the appropriate position to block the N-terminus of the three-stranded β-helix. Meanwhile, the size exclusion chromatogram and the AUC-SV analysis indicates that *$(VXG)_{6\times2}$Cf contains slightly less large aggregate

(larger than dimer of trimer) than $(VXG)_{6\times2}C^{his}$, which indicates that the fusion of SlyD and foldon would be effective to reduce the large aggregation.

The native PAGE of $*(VXG)_{6\times2}Cf$ indicated that the SEC fraction containing dimer of trimer gave a clear band and smear band but all the other fractions gave only smear bands, which appears from top to bottom of the gel depending on the elution volume. This is the first time to succeed in the visualization of the molecular species except for analytical ultracentrifugation. The molecular species, which gives a clear band, was trypsin-resistant but the molecular species, which gave smear bands, were digested by trypsin. It strongly suggests that the trypsin-resistant species is the fully folded dimer of trimer, which has a highly stable structure and the others are the oligomers of partially unfolded trimer, which has less stable structure. It is not conclusive whether the trimer of $*(VXG)_{6\times2}Cf$, which gave smear bands in native PAGE, is partially folded species, because it is possible that they are the "intermediate species" which would eventually assemble into dimer of trimer. However, the fact that SlyD-removed ^{his}SlyD-$(VXG)_{6\times2}$ gave two molecular species in analytical SEC (one elutes at 13.3ml and the other elutes at 13.0ml) supports the idea that they are partially unfolded species. The frictional ratio (f/f_0) of purified $*(VXG)_{6\times2}Cf$ was estimated to be 1.76 by SEDFIT, which is slightly but certainly bigger than that of $*(VXG)_{6\times1}$ (1.65). The frictional ratio becomes larger if the molecule has more prolonged shape. Thus $*(VXG)_{6\times2}$ would have more prolonged shape than $*(VXG)_{6\times1}$, which also supports that trypsin-resistant $*(VXG)_{6\times2}$ fully folds including the artificially introduced building block. The model of oligomerization and folding states of ^{his}SlyD-$(VXG)_{6\times2}$ is shown in Figure 4-20.

The AUC-SV analysis of trypsin treated $*(VXG)_{6\times2}$ indicated that they contain 5.02 S species for ~90% of total *c(s)*. The high percentage of *c(s)* as well as the almost clear single band in native PAGE demonstrate that $*(VXG)_{6\times2}$ was purified to near homogeneity. It is expected that the purification procedure of ^{his}SlyD-$(VXG)_{6\times2}Cf$ would be applicative to longer artificial three-stranded β-helical polypeptides.

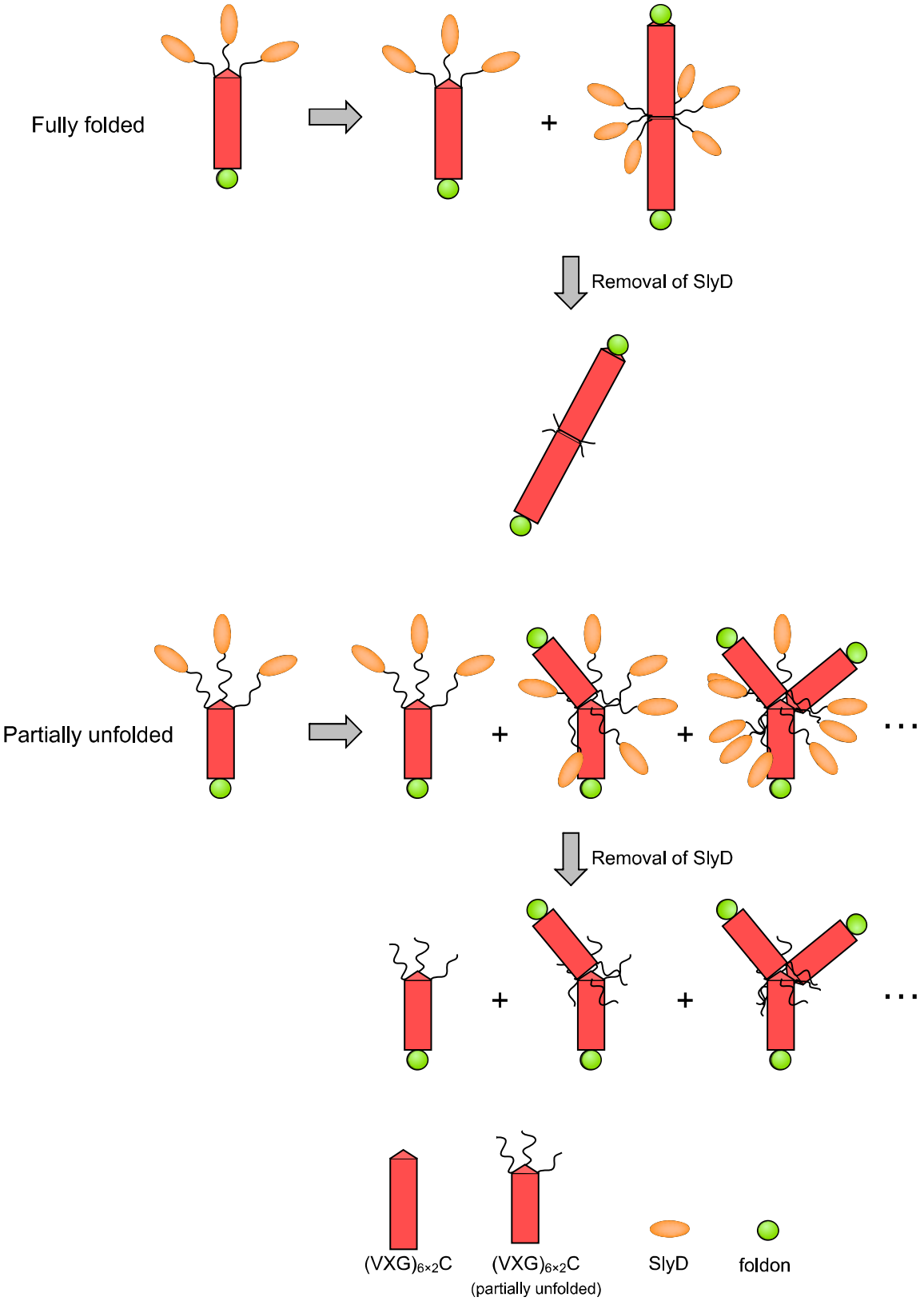

Figure 4-20. Schematic representation of the oligomerization state and folding state of ^{his}SlyD-(VXG)$_{6\times2}$Cf and *(VXG)$_{6\times2}$Cf.

Chapter 5

Concluding Remarks and Future Perspectives

5-1 Concluding Remarks

In order to elucidate the evolutionary relationship between VgrG proteins, which are a part of the type VI secretion system of bacteria, and T4 phage gp27-gp5 complex, which constitutes the cell-puncturing device of the phage, I have utilized the approach of structure biology and bioinformatics. As mentioned in Introduction, they are structurally highly similar in spite of the fact that they have quite distinct functions. For this purpose, I have focused on the structure of VgrG proteins from *E. coli* O157 and CFT073 strains and gp5 of bacteriophage KVP40, which is similar to T4 gp5, but does not have lysozyme domain. The results of these proteins towards structural determinations are summarized in Table 5-1. I have also tried to create artificial three-stranded β-helix in trimeric form, where we could precisely regulate the length. Finally, I attempted to purify the artificial three-stranded β-helix to near homogeneity by limited proteolysis.

1. *E. coli* O157 VgrG1^{his} and *E. coli* O157 c3393^{his} were purified to near homogeneity and were shown to be trimer in solution and rich in β-structure by analytical ultracentrifugation and far-UV CD analysis (Figure 2-9, 2-10 and Table 2-3). The characterization of the putative β-helix region of *E. coli* O157 VgrG2 and VgrG3 and *E. coli* CFT073 c1888 and c1883 indicated that the three-stranded β-helix of VgrG proteins is prone to aggregate.
2. The crystal structure of the C-terminal trypsin-resistant fragment of VgrG1 (VgrG1C^{G561}) was determined at 1.95Å. It forms a prism-shaped antiparallel three-stranded β-helix (Figure 2-22). The topology and structure of VgrG1C^{G561} were similar to the three-stranded β-helix domain of gp138 of phi92 phage, which indicates a possible evolutional relationship between them. VgrG1C^{G561} has a long loop of 17 amino acid residues between Asp570 and Ala586. I surmised that some VgrG proteins would have an enzymatic domain in this loop or a certain Type 6

secretion system protein would bind to the loop. VgrG1C^{G561} is stabilized by the hydrogen bonds and a salt bridge at the sharp kinks in addition to the hydrophobic core at the center of the β-helix (Figure 2-23 and 2-24).

3. In order to reveal the determinant of the topology of the three-stranded β-helix, structure-based sequence alignment of intertwined three-stranded β-helices and antiparallel three-stranded β-helices were performed. It indicated the different preference and restriction of the amino acids due to their structural difference, which makes it possible to predict the topology of three-stranded β-helix to some extent (Figure 2-28 and 2-29).
4. I concentrated on the crystal structural analysis of *E. coli* CFT073 c3393, which was predicted to contain both intertwined and parallel three-stranded β-helix (Figure 2-29). ^{his}c3393 was successfully crystallized and the X-ray diffraction data was collected at 3.35Å resolution (Figure 2-26, 2-27 and Table 2-5). The phase was probably determined by molecular replacement using the crystals structure of the N-terminal region of c3393 (PDB 2p5z) as a search model but the search model needs to be improved for complete structure.
5. I have attempted to determine the crystal structure of KVP40 gp5. It was crystallized in two different forms (hexagonal-shape and rhombohedral-shape (Figure 3-2)), and the full X-ray diffraction data was collected from rhombohedral-shaped crystals at the resolution of 2.59Å (Table 3-1).
6. In parallel with the structural analysis of three-stranded β-helical proteins, I also attempted to obtain the artificial three-strained β-helix in more homogenous form. It was found that the fusion of SlyD could not completely prevent the association of artificial three-stranded β-helix in my constructs, but reduced the large aggregates. Native PAGE made it possible to distinguish the oligomerization state and, more importantly, detect partially unfolded species (Figure 4-16). It was also found that partially unfolded species could be eliminated by the limited proteolysis by trypsin (Figure 4-17).
7. Finally, I succeeded in purifying the artificial three-stranded β-helix to near homogeneity in dimer of trimer form by using trypsin treatment (Figure 4-18 and 4-19). I expect the purification method established in this study would be applicative to longer construct.

5-2 Future Perspectives

1. The X-ray diffraction data of ^{his}c3393 and KVP40 gp5^{his} were obtained in the present study. In the case of KVP40 gp5^{his}, the expression and purification procedure of the SeMet derivative is already established and its crystallization trials are ongoing. The T4 gp5 has the longest three-stranded β-helix with known three-dimensional structure but KVP40 gp5 is expected to have even longer three-stranded β-helix. Hence, I hope that the crystal structure of KVP40 gp5 will promote the understanding of three-stranded β-helical structure and also give us a clue to design new artificial three-stranded β-helices.
2. Another experiment to be performed is the examination if the RHS (rearrangement hot spot, Chapter2 2-4) protein could facilitate the folding of VgrG proteins. The experiment has been already started in our laboratory.
3. In the present work, the presence of partially unfolded artificial three-stranded β-helix was detected. One possible method to facilitate the folding of artificial three-stranded β-helix and eliminate these species would be the introduction of the hydrophobic core of gp5C (called C or Cgp5C in this study) to before or after the building block. Cgp5C is expected to fold very rigidly thus I expect it would stabilize three-stranded β-helix.

Table 5-1 Summary of the results towards structural determination

Protein constructs		Expression	Purification	Crystallization	Collection of X-ray diffraction data	Structure determination
E. coli O157	$VgrG1^{his}$	○	○			
E. coli O157	$VgrG1C^{M467\text{-}his}$	○	○			
E. coli O157	$VgrG1C^{G561\text{-}his}$	○	○	○	○ (1.95Å)	○
E. coli O157	$VgrG2^{his}$	○				
E. coli O157	$^{his}SlyD\text{-}VgrG2C^{G474}$	○	○[a]			
E. coli O157	$VgrG3^{his}$	○				
E. coli O157	$^{his}SlyD\text{-}VgrG3C^{G458}$	○	○[a]			
E. coli CFT073	$c3393^{his}$	○	○			
E. coli CFT073	$^{his}c3393$	○	○	○	○ (3.35Å)	
E. coli CFT073	$c1888^{his}$	○				
E. coli CFT073	$^{his}SlyD\text{-}c1888C^{L461}$	○	○[a]			
E. coli CFT073	$c1883^{his}$	○				
E. coli CFT073	$^{his}SlyD\text{-}c1883C^{G459}$	○	○[a]			
KVP40 phage	$gp5^{his}$	○[b]	○[b]	○	○ (2.59Å)	

○ has been achieved.

[a] The proteins form aggregation.

[b] The expression and purification of KVP40 gp5 were reported by Nemoto et al. (*61*)

References

1. Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., and Phillips, D.C. (1958) A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*. **181**, 662-666
2. Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G., and North, A.C. (1960) Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. *Nature*. **185**, 416-422
3. van Raaij, M.J., Mitraki, A., Lavigne, G., and Cusack, S. (1999) A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature*. **401**, 935-938
4. Williams, P.A., Fulop, V., Garman, E.F., Saunders, N.F., Ferguson, S.J., and Hajdu, J. (1997) Haem-ligand switching during catalysis in crystals of a nitrogen-cycle enzyme. *Nature*. **389**, 406-412
5. Yoder, M.D., Keen, N.T., and Jurnak, F. (1993) New domain motif: the structure of pectate lyase C, a secreted plant virulence factor. *Science*. **260**, 1503-1507
6. Baumann, U., Wu, S., Flaherty, K.M., and McKay, D.B. (1993) Three-dimensional structure of the alkaline protease of Pseudomonas aeruginosa: a two-domain protein with a calcium binding parallel beta roll motif. *EMBO J.* **12**, 3357-3364
7. van Raaij, M.J., Schoehn, G., Burda, M.R., and Miller, S. (2001) Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre. *J. Mol. Biol.* **314**, 1137-1146
8. Kanamaru, S., Leiman, P.G., Kostyuchenko, V.A., Chipman, P.R., Mesyanzhinov, V.V., Arisaka, F., and Rossmann, M.G. (2002) Structure of the cell-puncturing device of bacteriophage T4. *Nature*. **415**, 553-557
9. Sugimoto, K., Kanamaru, S., Iwasaki, K., Arisaka, F., and Yamashita, I. (2006) Construction of a ball-and-spike protein supramolecule. *Angewandte Chemie*. **45**, 2725-2728
10. Ueno, T., Koshiyama, T., Tsuruga, T., Goto, T., Kanamaru, S., Arisaka, F., and Watanabe, Y. (2006) Bionanotube tetrapod assembly by in situ synthesis of a gold nanocluster with (Gp5-His6)3 from bacteriophage T4. *Angewandte Chemie*. **45**, 4508-4512
11. Yokoi, N., Inaba, H., Terauchi, M., Stieg, A.Z., Sanghamitra, N.J., Koshiyama, T., Yutani, K., Kanamaru, S., Arisaka, F., Hikage, T., Suzuki, A., Yamane, T., Gimzewski, J.K., Watanabe, Y., Kitagawa, S., and Ueno, T. (2010) Construction of robust

bio-nanotubes using the controlled self-assembly of component proteins of bacteriophage T4. *Small*. **6**, 1873-1879

12. Tetart, F., Desplats, C., Kutateladze, M., Monod, C., Ackermann, H.W., and Krisch, H.M. (2001) Phylogeny of the major head and tail genes of the wide-ranging T4-type bacteriophages. *J. Bacteriol.* **183**, 358-366
13. Hambly, E., Tetart, F., Desplats, C., Wilson, W.H., Krisch, H.M., and Mann, N.H. (2001) A conserved genetic module that encodes the major virion components in both the coliphage T4 and the marine cyanophage S-PM2. *Proceedings of the National Academy of Sciences of the United States of America*. **98**, 11411-11416
14. Rossmann, M.G., Mesyanzhinov, V.V., Arisaka, F., and Leiman, P.G. (2004) The bacteriophage T4 DNA injection machine. *Curr. Opin. Struct. Biol.* **14**, 171-180
15. Yamashita, E., Nakagawa, A., Takahashi, J., Tsunoda, K., Yamada, S., and Takeda, S. (2011) The host-binding domain of the P2 phage tail spike reveals a trimeric iron-binding structure. *Acta crystallographica. Section F, Structural biology and crystallization communications*. **67**, 837-841
16. Browning, C., Shneider, M.M., Bowman, V.D., Schwarzer, D., and Leiman, P.G. (2012) Phage pierces the host cell membrane with the iron-loaded spike. *Structure*. **20**, 326-339
17. Harada, K., Yamashita, E., Nakagawa, A., Miyafusa, T., Tsumoto, K., Ueno, T., Toyama, Y., and Takeda, S. (2013) Crystal structure of the C-terminal domain of Mu phage central spike and functions of bound calcium ion. *Biochim. Biophys. Acta*. **1834**, 284-291
18. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*. **23**, 2947-2948
19. Shneider, M.M., Buth, S.A., Ho, B.T., Basler, M., Mekalanos, J.J., and Leiman, P.G. (2013) PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. *Nature*. **500**, 350-353
20. Mougous, J.D., Cuff, M.E., Raunser, S., Shen, A., Zhou, M., Gifford, C.A., Goodman, A.L., Joachimiak, G., Ordonez, C.L., Lory, S., Walz, T., Joachimiak, A., and Mekalanos, J.J. (2006) A virulence locus of Pseudomonas aeruginosa encodes a protein secretion apparatus. *Science*. **312**, 1526-1530
21. Pukatzki, S., Ma, A.T., Sturtevant, D., Krastins, B., Sarracino, D., Nelson, W.C., Heidelberg, J.F., and Mekalanos, J.J. (2006) Identification of a conserved bacterial protein secretion system in Vibrio cholerae using the Dictyostelium host model

system. *Proceedings of the National Academy of Sciences of the United States of America*. **103**, 1528-1533

22. Bingle, L.E., Bailey, C.M., and Pallen, M.J. (2008) Type VI secretion: a beginner's guide. *Curr. Opin. Microbiol.* **11**, 3-8
23. Basler, M., Pilhofer, M., Henderson, G.P., Jensen, G.J., and Mekalanos, J.J. (2012) Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature*. **483**, 182-186
24. Boyer, F., Fichant, G., Berthod, J., Vandenbrouck, Y., and Attree, I. (2009) Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC Genomics*. **10**, 104
25. Pukatzki, S., Ma, A.T., Revel, A.T., Sturtevant, D., and Mekalanos, J.J. (2007) Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proceedings of the National Academy of Sciences of the United States of America*. **104**, 15508-15513
26. Ma, A.T., and Mekalanos, J.J. (2010) In vivo actin cross-linking induced by Vibrio cholerae type VI secretion system is associated with intestinal inflammation. *Proceedings of the National Academy of Sciences of the United States of America*. **107**, 4365-4370
27. Suarez, G., Sierra, J.C., Erova, T.E., Sha, J., Horneman, A.J., and Chopra, A.K. (2010) A type VI secretion system effector protein, VgrG1, from Aeromonas hydrophila that induces host cell toxicity by ADP ribosylation of actin. *J. Bacteriol.* **192**, 155-168
28. Hood, R.D., Singh, P., Hsu, F., Guvener, T., Carl, M.A., Trinidad, R.R., Silverman, J.M., Ohlson, B.B., Hicks, K.G., Plemel, R.L., Li, M., Schwarz, S., Wang, W.Y., Merz, A.J., Goodlett, D.R., and Mougous, J.D. (2010) A type VI secretion system of Pseudomonas aeruginosa targets a toxin to bacteria. *Cell Host Microbe*. **7**, 25-37
29. Russell, A.B., Hood, R.D., Bui, N.K., LeRoux, M., Vollmer, W., and Mougous, J.D. (2011) Type VI secretion delivers bacteriolytic effectors to target cells. *Nature*. **475**, 343-347
30. Dong, T.G., Ho, B.T., Yoder-Himes, D.R., and Mekalanos, J.J. (2013) Identification of T6SS-dependent effector and immunity proteins by Tn-seq in Vibrio cholerae. *Proceedings of the National Academy of Sciences of the United States of America*. **110**, 2623-2628
31. Aschtgen, M.S., Bernard, C.S., De Bentzmann, S., Lloubes, R., and Cascales, E. (2008) SciN is an outer membrane lipoprotein required for type VI secretion in enteroaggregative Escherichia coli. *J. Bacteriol.* **190**, 7523-7531

32. Burtnick, M.N., DeShazer, D., Nair, V., Gherardini, F.C., and Brett, P.J. (2010) Burkholderia mallei cluster 1 type VI secretion mutants exhibit growth and actin polymerization defects in RAW 264.7 murine macrophages. *Infect. Immun.* **78**, 88-99

33. Schell, M.A., Ulrich, R.L., Ribot, W.J., Brueggemann, E.E., Hines, H.B., Chen, D., Lipscomb, L., Kim, H.S., Mrazek, J., Nierman, W.C., and Deshazer, D. (2007) Type VI secretion is a major virulence determinant in Burkholderia mallei. *Mol. Microbiol.* **64**, 1466-1485

34. Lesic, B., Starkey, M., He, J., Hazan, R., and Rahme, L.G. (2009) Quorum sensing differentially regulates Pseudomonas aeruginosa type VI secretion locus I and homologous loci II and III, which are required for pathogenesis. *Microbiology*. **155**, 2845-2855

35. Leiman, P.G., Basler, M., Ramagopal, U.A., Bonanno, J.B., Sauder, J.M., Pukatzki, S., Burley, S.K., Almo, S.C., and Mekalanos, J.J. (2009) Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proceedings of the National Academy of Sciences of the United States of America*. **106**, 4154-4159

36. Hachani, A., Lossi, N.S., Hamilton, A., Jones, C., Bleves, S., Albesa-Jove, D., and Filloux, A. (2011) Type VI secretion system in Pseudomonas aeruginosa: secretion and multimerization of VgrG proteins. *J. Biol. Chem.* **286**, 12317-12327

37. Kanamaru, S. (2009) Structural similarity of tailed phages and pathogenic bacterial secretion systems. *Proceedings of the National Academy of Sciences of the United States of America*. **106**, 4067-4068

38. Sycheva, L.V., Shneider, M.M., Sykilinda, N.N., Ivanova, M.A., Miroshnikov, K.A., and Leiman, P.G. (2012) Crystal structure and location of gp131 in the bacteriophage phiKZ virion. *Virology*. **434**, 257-264

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

40. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*. **9**, 671-675

41. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. **227**, 680-685

42. Kapust, R.B., Tozser, J., Fox, J.D., Anderson, D.E., Cherry, S., Copeland, T.D., and Waugh, D.S. (2001) Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng*. **14**, 993-1000

43. Sreerama, N., and Woody, R.W. (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287**, 252-260

44. Schuck, P., Perugini, M.A., Gonzales, N.R., Howlett, G.J., and Schubert, D. (2002) Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. *Biophys. J.* **82**, 1096-1111

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

46. Laue, T.M., Shah, B.D., Ridgeway, T.M., and Pelletier, S.L. (1992) Computer-aided interpretation of analytical sedimentation data for proteins. *In S. E. Harding, A. J. Rowe and J. C. Horton (ed.), Analytical ultracentrifugation in biochemistry and polymer science. Royal Society of Chemistry, Cambridge, United Kingdom.*, 90-125

47. Kabsch, W. (2010) Xds. *Acta crystallographica. Section D, Biological crystallography*. **66**, 125-132

48. Evans, P. (2006) Scaling and assessment of data quality. *Acta crystallographica. Section D, Biological crystallography*. **62**, 72-82

49. Collaborative Computational Project, N. (1994) The CCP4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography*. **50**, 760-763

50. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., and Zwart, P.H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography*. **66**, 213-221

51. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010) Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography*. **66**, 486-501

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

53. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta crystallographica. Section D, Biological crystallography*. **60**, 2256-2268

54. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188-1190

55. Uchida, K., Leiman, P.G., Arisaka, F., and Kanamaru, S. (2013) Structure and properties of the C-terminal beta-helical domain of VgrG protein from Escherichia coli O157. *Journal of biochemistry*.

56. Han, K.Y., Song, J.A., Ahn, K.Y., Park, J.S., Seo, H.S., and Lee, J. (2007) Solubilization of aggregation-prone heterologous proteins by covalent fusion of stress-responsive Escherichia coli protein, SlyD. *Protein Eng. Des. Sel.* **20**, 543-549

57. Durand, E., Derrez, E., Audoly, G., Spinelli, S., Ortiz-Lombardia, M., Raoult, D., Cascales, E., and Cambillau, C. (2012) Crystal structure of the VgrG1 actin cross-linking domain of the Vibrio cholerae type VI secretion system. *J. Biol. Chem.* **287**, 38190-38199

58. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr.* **40**, 658-674

59. Matsuzaki, S., Tanaka, S., Koga, T., and Kawata, T. (1992) A broad-host-range vibriophage, KVP40, isolated from sea water. *Microbiol. Immunol.* **36**, 93-97

60. Miller, E.S., Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Durkin, A.S., Ciecko, A., Feldblyum, T.V., White, O., Paulsen, I.T., Nierman, W.C., Lee, J., Szczypinski, B., and Fraser, C.M. (2003) Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. *J. Bacteriol.* **185**, 5220-5233

61. Nemoto, M., Mio, K., Kanamaru, S., and Arisaka, F. (2008) ORF334 in Vibrio phage KVP40 plays the role of gp27 in T4 phage to form a heterohexameric complex. *J. Bacteriol.* **190**, 3606-3612

62. Pflugrath, J.W. (1999) The finer things in X-ray diffraction data collection. *Acta crystallographica. Section D, Biological crystallography*. **55**, 1718-1725

63. Mueller, M., Wang, M., and Schulze-Briese, C. (2012) Optimal fine phi-slicing for single-photon-counting pixel detectors. *Acta crystallographica. Section D, Biological crystallography*. **68**, 42-56

64. Bhardwaj, A., Walker-Kopp, N., Wilkens, S., and Cingolani, G. (2008) Foldon-guided self-assembly of ultra-stable protein fibers. *Protein Sci.* **17**, 1475-1485

65. Terauchi, M. (2008) Designing of tubular structures using building blocks derived from three-stranded β-helix protein with protein engineering (Master thesis).

66. Aburai, K. (2010) Design, Expression and Aggregation inhibition of the length controllable artificial triple beta-helix protein (Master thesis).

67. Davis, B.J. (1964) Disc Electrophoresis. Ii. Method and Application to Human Serum Proteins. *Ann. N. Y. Acad. Sci.* **121**, 404-427

68. 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

Acknowledgement

The present work has been under the direction of Professor Fumio Arisaka of Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology.

First of all, I express my deepest gratitude to Professor Fumio Arisaka. He always gave me precious advices when I faced difficulty not only in interpretation of the experiment data but also in the designing of career path. He also kindly read this doctoral thesis in his extraordinarily busy time. It was my great honor that I could learn from him for five years.

I express my sincere gratitude to Dr. Shuji Kanamaru for his kind guidance and valuable advices throughout this research. He always brought me a new point of view in the interpretation of experimental data and his suggestions were really helpful in the designing of next experiments.

I would like to express my sincere appreciation to Professor Petr G. Leiman, who accepted my three-month stay and instructed me by himself. I never forget that he taught me how to use the X-ray crystallographic software until the night of two days before my departure for Japan. He also kindly provided me the genome DNA of *E. coli* CFT073.

I also thank the lab members of Petr G. Leiman's laboratory. Dr. Christopher Browning gave me advices on protein crystallization experiments as well as X-ray diffraction experiments. Dr. Lada Sycheva kindly supported my experiments.

I sincerely thank the Professor Susumu Kajiwara and the staff of International Training Program, who gave me the opportunity to study in Professor Petr G Leiman's laboratory.

I express my gratitude to Dr. Mikhail Shneider, who provided us the expression vector containing the SlyD chaperone.

I wish to express my gratitude to the beam line staff of the Photon factory protein crystallography beamlines (BL-1A, BL5-A and BL-17A), Swiss Light Source PX III beamline and SPring-8 Structural Biology III beamline (BL38B1) for their help of X-ray diffraction analysis and data collection.

I would like to express my appreciation for Mr. Kaname Nishijo, who

constructed the over-expression vector of three-stranded β-helix and supported me to characterize them.

I would like to extend my gratitude to Mr. Kohei Aburai, Mr. Yasunori Monzaki and Dr. Yap Moh lan for their supports of my research. They taught me how to use the machineries in detail.

I am thankful for all the lab members, who gave me a joyful time in really hard lab life. The supper with them was the only rest time in the last few years.

Last, I express my heartfelt thanks to my parents for their supports and understandings. Thanks to their many helps, I could concentrate on my research every day.

Publication

Kazuya Uchida, Petr G Leiman, Fumio Arisaka and Shuji Kanamaru (2013): Structure and properties of the C-terminal β-helical domain of VgrG protein from Escherichia coli O157. *J Biochem.* (in press)