

# Structure and properties of the C-terminal $\beta$ -helical domain of VgrG protein from *Escherichia coli* O157

Received October 18, 2013; accepted November 26, 2013; published online December 3, 2013

Kazuya Uchida<sup>1</sup>, Petr G. Leiman<sup>2</sup>,  
Fumio Arisaka<sup>1</sup> and Shuji Kanamaru<sup>1,\*</sup>

<sup>1</sup>Department of Bioengineering, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259-B9 Nagatsuta-cho, Midori-ku, Yokohama 226-8501 Japan; and <sup>2</sup>École Polytechnique Fédérale de Lausanne (EPFL), Laboratory of Structural Biology and Biophysics, BSP-415, CH-1015 Lausanne, Switzerland

\*Shuji Kanamaru, 4259-B9 Nagatsuta-cho, Midori-ku, Yokohama 226-8501 Japan. Tel: +81-45-924-5712, Fax: +81-45-924-5713, email: skanamar@bio.titech.ac.jp

**Protein Data Bank accession number.** The atomic coordinate and structure factor of VgrG1C<sup>G561</sup> are deposited in the PDB with the accession number of 3WIT.

The bacterial Type 6 secretion system (T6SS) translocates protein toxins (also called effectors) from the cytosol of a T6SS-carrying cell to a target cell by a syringe-like supramolecular complex resembling a contractile tail of bacteriophages. Valine-glycine repeat protein G (VgrG) proteins, which are the homologues of the gp27-gp5 (gene product) cell puncturing complex of bacteriophage T4, are considered to be located at the attacking tip of the bacterial T6SS apparatus. Here, we over-expressed six VgrG proteins from pathogenic *Escherichia coli* O157 and CFT073 strains. Purified VgrG1 of *E. coli* O157 and c3393 of *E. coli* CFT073 form trimer in solution and are rich in  $\beta$ -structure. We also solved the crystal structure of a trypsin-resistant C-terminal fragment of *E. coli* O157 VgrG1 (VgrG1C<sup>G561</sup>) at 1.95 Å resolution. VgrG1C<sup>G561</sup> forms a three-stranded antiparallel  $\beta$ -helix which is structurally similar to the  $\beta$ -helix domain of the central spike protein (gp138) of phi92 phage, indicating a possible evolutionary relationship. Comparison of four different three-stranded  $\beta$ -helix proteins shows how their amino acid composition determines the protein fold.

**Keywords:** bacteria/structure/three-stranded  $\beta$ -helix/Type 6 secretion system/X-ray crystallography.

**Abbreviations:** CD, circular dichroism; CHES, *N*-Cyclohexyl-2-aminoethanesulphonic acid; *c*(*s*), sedimentation coefficient distribution function; gp, gene product; LB, Luria-Bertani; OB fold, oligonucleotide-/oligosaccharide-binding fold; PAAR motif, proline-alanine-alanine-arginine motif; RHS, rearrangement hot spot; SeMet, seleno methionine; T6SS, Type 6 secretion system; TEV, tobacco etch virus; VgrG, valine-glycine repeat protein G.

The related genes of the secretion system form a cluster. Most genes encoding T6SS proteins often cluster together, and this cluster consists of highly conserved 13 core genes and several strain-specific genes (2). The structural proteins of T6SS assemble into a contractile supramolecular structure with the length of  $\sim 0.7 \mu\text{m}$  (1). Upon contraction of the sheath structure, it shortens down to  $\sim 0.4 \mu\text{m}$  and drives secretion of protein toxins or effector molecules into target cells (1). T6SS plays various physiological functions, e.g. cytotoxicity against eukaryotic cells by actin crosslinking [*Vibrio cholerae* (3, 4)] or actin ribosylation [*Aeromonas hydrophila* (5)], interbacterial competition [*Pseudomonas aeruginosa* (6, 7) and *V. cholerae* (8)], biofilm formation [enteroaggregative *Escherichia coli* (9)] and virulence [*Burkholderia mallei* (10, 11) and *P. aeruginosa* (12, 13)].

The Type 6 secretion apparatus consists of two parts, namely a syringe-like structure extending to the cell membrane and the membrane-associated protein complex. Recent studies indicated that 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  $\mu\text{m}$ ). The syringe-like structure is assumed to contain (among others) the following proteins—VgrG, Hcp and VipA/VipB—which are the homologues of the gp27–gp5 (gene product) complex, gp19 (tail tube protein) and gp18 (tail sheath protein) of T4 phage, respectively (1). On the other hand, most of the membrane-associated proteins are not homologous to phage proteins. Interestingly, a homologue 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 T6SS (14).

T4 phage is a very well-characterized bacteriophage that infects *E. coli* (15). It injects its genomic DNA into the host cell by a contractile tail which is constituted of a co-cylinder structure and the baseplate at the distal end. The co-cylinder structure consists of two parts namely, tail tube (passage of DNA, gp19) and contractile tail sheath (gp18) surrounding the tube. The tip of the tail tube carries a needle- or spike-shaped structure called the gp27–gp5 tail lysozyme complex (16). Upon infection, the gp27–gp5 complex penetrates the outer membrane of the host cell and locally degrades the peptidoglycan layer with the help of the lysozyme domain that is part of gp5 (16). The C-terminal domain of T4 gp5 contains a tandem 8 amino acid repeat of (VXGXXXXX), which forms a three-stranded  $\beta$ -helix in gp5 trimer (16). It is a highly stable structure and shows resistance to 10% sodium dodecyl sulfate (SDS) and 2 M guanidine HCl (16). The three-stranded  $\beta$ -helix is conserved in the central

Bacterial Type 6 secretion system (T6SS) is one of the largest organelles localized in the cell membrane (1).

spike proteins of *Myoviridae* such as phi92, P2 and Mu phages, though there are two types of three-stranded  $\beta$ -helix; one is made of fully intertwined  $\beta$ -helix and the other is made of three antiparallel  $\beta$ -sheets (17–19).

In T6SS, valine-glycine repeat protein G (VgrG) protein is thought to function similar to the gp27–gp5 complex in T4 tail. VgrG is an essential component of T6SS and found in the culture supernatants of the T6SS expressing cells (3, 8, 20). Multiple *vgrG* genes are often found in many bacterial genomes. They often contain a conserved part, which corresponds to genes 27 and 5 of T4 fused together, and 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) (3). Interestingly, VgrG protein is predicted to contain triple-stranded  $\beta$ -helix as in gp5 of T4 phage (3), but the number of predicted  $\beta$ -strands and their amino acid sequences are highly diverse in VgrG proteins. Some VgrGs display a VXGXXXXX repeat (X, any amino acid), which characterizes the T4 gp5  $\beta$ -helix with others displaying a VGXXXXXX repeat and yet others showing another repeat or no apparent repeat at all. How these sequence variations are reflected in the three dimensional structure is of great interest.

In this study, we characterize six VgrG proteins from *E. coli* O157 and CFT073 strains and show that both *E. coli* VgrG1 and *E. coli* CFT073 c3393 are trimer in solution and rich in  $\beta$ -structure as in T4 gp27–gp5 complex, except for the lysozyme domain which is absent in these VgrG proteins. We also solved the crystal structure of the C-terminal fragment of *E. coli* O157 VgrG1 and proved that it forms a three-stranded  $\beta$ -helix. It resembles the antiparallel  $\beta$ -helix domain of the central spike proteins from phi92 phage (17). Finally, we propose a method for prediction of  $\beta$ -helix topology and show that it is possible to predict the structure and topology of several diverse VgrG proteins to some extent.

## Materials and Methods

### Chemicals

All the chemicals were purchased from Nacalai Tesque (Japan) or Wako Pure Chemical Industries (Japan). Protein markers were purchased from Bio-Rad Laboratories.

### Plasmid construction

The *vgrG* genes, *vgrG1* (locus tag Z0707), *vgrG2* (locus tag Z2262), *vgrG3* (locus tag Z0267) of *E. coli* O157 EDL933 and c3393, c1888, c1883 of *E. coli* CFT073 were polymerase chain reaction (PCR)-amplified from genome DNA of each *E. coli* using the primers listed in Supplementary Table S1. The over-expression vectors of each full-length VgrG protein were constructed by inserting *vgrG* gene into pET29a vector using the restriction enzyme sites of NdeI and XhoI. The resulting plasmids express VgrG protein with C-terminal hexahistidine tag. The expression vector of the  $\beta$ -helix domain of VgrG1 (residues 467–633) with C-terminal hexahistidine tag was constructed in the same way but using the primer pairs of VgrG1-NdeI-467M(+) and Ec933Vgr1-Gly-Xho(-) (Supplementary Table S1 for details). The expression vectors of the  $\beta$ -helix domain of *E. coli* O157 VgrG2 and VgrG3 with N-terminal SlyD protein were constructed as follows. The VgrG2 residues 474–714 and VgrG3 residues 458–713 were amplified by PCR using the primers of VgrG2-BamHI-G474(+) and Ec933Vgr2-STP-Xho(-) or VgrG3-BamHI-G458 (+) and Ec933Vgr3-STP-Xho(-)

(Supplementary Table S1) and inserted to pSL vector by BamHI and XhoI restriction enzyme sites. pSL vector was made from pESL (21) by deleting a linker sequence between tobacco etch virus (TEV) protease site and multiple cloning site. The resulting plasmids express the  $\beta$ -helix domain of VgrG2 and VgrG3 with cleavable hexahistidine tag and SlyD at N-terminus. The C-terminal fragment of VgrG proteins are named as VgrGC with the starting residue in upper right (*e.g.* VgrG2C<sup>G474</sup>). The N-terminal and C-terminal hexahistidine tag are indicated with his in upper left and upper right, respectively (*e.g.* VgrG2<sup>his</sup>). A schematic view of the constructs used in this study is shown in Supplementary Fig. S4. All the nucleotide sequence of *vgrG* gene in the expression vector was confirmed by DNA sequencing.

### Protein expression

The expression of *E. coli* O157 VgrG1<sup>his</sup>, VgrG2<sup>his</sup>, VgrG3<sup>his</sup>, VgrG1C<sup>M467-his</sup>, hisSlyD-VgrG2C<sup>G474</sup> and hisSlyD-VgrG3C<sup>G458</sup> was performed as follows. BL21(DE3) was transformed by the expression vector described above and cultivated at 37°C in Luria-Bertani (LB) medium supplemented with 60  $\mu$ g/ml kanamycin. For the expression of VgrG1<sup>his</sup>, VgrG2<sup>his</sup> and VgrG3<sup>his</sup>, isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) was added to the final concentration of 0.1 mM when OD<sub>660</sub> was 0.4–0.5 and further cultivated at 25°C for 6 h. For the expression of VgrG1C<sup>M467-his</sup>, IPTG was added at a final concentration of 1 mM and cultivated at 20°C for 5 h. The SeMet derivative of VgrG1C<sup>M467-his</sup> was expressed basically in the same way as native protein but *E. coli* B384(DE3) and Le Master medium (22) supplemented with 25 mg/l Seleno-L-Methionine were used. For the expression of hisSlyD-VgrG2C<sup>G474</sup> and hisSlyD-VgrG3C<sup>G458</sup>, IPTG was added to 1 mM final concentration and further cultivated at 20°C for 10 h.

c3393<sup>his</sup>, c1888<sup>his</sup> and c1883<sup>his</sup> from *E. coli* CFT073 were expressed as follows. *E. coli* KRX (Promega Corporation, USA) was transformed by the expression plasmid and cultivated at 37°C in LB medium containing 60  $\mu$ g/ml kanamycin. The expression was induced by 0.1% rhamnose and 1 mM IPTG when OD<sub>660</sub> was about 0.5 and incubated at 16°C for 18 h. All cells were harvested by centrifugation at 2,500 $\times$ g for 10 min using Himac CR22G centrifuge with R9AF rotor (Hitachi Koki).

The soluble content of VgrG1<sup>his</sup> and c3393<sup>his</sup> were estimated by ImageJ (23) using scanned sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels by CanoScan D1250U2 (Canon Inc.).

### Protein purification

VgrG1<sup>his</sup>, c3393<sup>his</sup>, VgrG1C<sup>M467-his</sup>, hisSlyD-VgrG2C<sup>G474</sup> and hisSlyD-VgrG3C<sup>G458</sup> which have hexahistidine tag were purified as follows. The cell was suspended in 10 $\times$  volume of buffer A (50 mM Tris pH8.0, 100 mM NaCl, 20 mM imidazole) and lysed by sonication in the presence of 1 mM phenylmethylsulphonyl fluoride. The cell lysate was centrifuged at 20,000 $\times$ g (R18A rotor with Himac CR22G centrifuge, Hitachi Koki) for 20 min and supernatant was loaded to HisTrap FF column (GE Healthcare) which was equilibrated with buffer A. VgrG1<sup>his</sup>, c3393<sup>his</sup>, VgrG1C<sup>M467-his</sup>, hisSlyD-VgrG2C<sup>G474</sup> and hisSlyD-VgrG3C<sup>G458</sup> were eluted by imidazole gradient of 20–500 mM and ethylenediaminetetraacetic acid (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 (50 mM Tris pH8.0). The proteins were eluted with NaCl gradient of 0 to 1 M and further purified by Hiload 16/60 Superdex200 prep grade column (GE Healthcare) using buffer C (50 mM Tris pH8.0, 100 mM NaCl).

The purification of VgrG1C<sup>G561-his</sup> is similar to that of VgrG1C<sup>M467-his</sup> but modified as follows. First, VgrG1C<sup>M467-his</sup> was purified by HisTrap FF column (GE Healthcare) then trypsin was added on the weight ratio of 100 to 1 (VgrG1C<sup>M467-his</sup> to trypsin) and incubated at 25°C for 0.5–1 h in order to digest the N-terminal region of VgrG1C<sup>M467-his</sup>. The proteolysis was stopped by the addition of Tosyl-Lys-chloromethyl ketone (TLCK) at a final concentration of 0.1 mM. VgrG1C<sup>G561-his</sup> which had been cleaved at the C-terminus of Arg560 was purified by HiTrap Q HP (GE Healthcare) and the fractions containing the desired protein were mixed and loaded to Hiload 16/60 Superdex75 prep grade column (GE Healthcare). The SeMet derivative of VgrG1C<sup>G561-his</sup> was

purified in the same way as native protein. All the purification was performed at 4°C.

#### SDS–PAGE and N-terminal amino acid sequence determination

SDS–PAGE was carried out according to the method of Laemmli (24) using a minigel and stained with 0.1% Coomassie brilliant blue G-250 in 10% acetic acid. For the protein sequencing of trypsin-treated VgrG1<sup>M467-his</sup>, proteins were separated by SDS–PAGE and transferred to a polyvinylidene difluoride membrane by electrophoresis. The membrane was stained with amido black 10B in 10% acetic acid and target band was examined. The N-terminal amino acid sequence was determined by Edman degradation using the protein sequencer PPSQ-21 (SHIMADZU).

#### hisSlyD removal from hisSlyD-VgrG3C<sup>G458</sup>

Hexahistidine-tagged TEV protease harbouring S219V mutation was prepared from *E. coli* BL21(DE3)-RIL containing pRK793 (Addgene plasmid 8827) (25). The TEV protease was mixed with hisSlyD-VgrG3C<sup>G458</sup> at a protease to substrate weight ratio of 1:10 and incubated at 30°C for 6 h. The sample was taken in every 2 h and digestion profile was analysed by SDS–PAGE after the centrifugation at 20,000 × *g* to separate soluble protein and insoluble protein.

#### Far-UV CD spectrum

Far-UV CD (circular dichroism) spectrum of VgrG1<sup>his</sup> and c3393<sup>his</sup> was measured at 20°C with a J-720 spectropolarimeter (Jasco) and 1 mm-path-length quartz cell. Buffer C was used as a control buffer. The secondary structure was estimated by program CONTINLL (26) using the wavelength range of 200–240 nm.

#### Analytical ultracentrifugation

Analytical ultracentrifugation was performed by Optima XL-I (Beckman-Coulter) using a four-hole An60Ti or an eight-hole An50Ti rotor at 20°C. The dialysis buffer or equilibration buffer for size exclusion chromatography was used as a reference solution.

Sedimentation velocity data of VgrG1<sup>his</sup> and c3393<sup>his</sup> were collected with the rotor speed of 129,024 × *g*, those of VgrG1C<sup>G561-his</sup> were collected at 201,600 × *g* and those of hisSlyD-VgrG2C<sup>G474</sup> were collected at 72,576 × *g*. Moving boundaries were recorded at the wave length of 230 nm or 280 nm depending on the concentration of the protein without time intervals between each scan. The sedimentation coefficient distribution function, *c*(*s*), was obtained by SEDFIT program (27, 28). The distribution of molecular weight, *c*(*M*), was obtained by converting *c*(*s*) to *c*(*M*) as implemented in SEDFIT program. The protein partial spec volumes ( $\bar{v}$ ), buffer density ( $\rho$ ) and buffer viscosity ( $\eta$ ) were calculated by SEDNTERP program (29).

For the sedimentation equilibrium of VgrG1C<sup>G561-his</sup>, the samples with the absorbance at 230 nm of 0.2, 0.3 and 0.5 were centrifuged at 14,480, 20,644 and 58,787 × *g* until they reach equilibrium. The molecular weight was calculated using Origin program with a self-associating model.

#### Crystallization and structure determination

For the crystallization, VgrG1C<sup>G561-his</sup> was dialysed against 10 mM Tris pH8.0 and concentrated by Amicon Ultra 10K (Millipore) to 14 mg/ml. The initial crystal of VgrG1C<sup>G561-his</sup> was obtained in condition B12 of JBScreen Pentaerythritol [45% w/v Pentaerythritol propoxylate (average M.W. ~426), 100 mM Tris pH8.5 and 400 mM KCl] (Jena Biosciences) by a sitting drop vapour diffusion method using 96-well MRC 2-drop plate (Swissci). The crystallization condition was optimized in 24-well plate (Jena Biosciences) by hanging drop vapour diffusion method. The biggest crystal was obtained in a drop which consists of 1  $\mu$ l of 5.0 mg/ml protein solution and 1  $\mu$ l well solution containing 48% v/v Pentaerythritol propoxylate (average M.W. ~426), 100 mM CHES pH8.8, 700 mM CsCl. The SeMet derivative did not crystallize under the same condition as native protein; therefore, we performed crystallization screening again. The crystal of SeMet derivative was obtained in condition 7 of Cryo 1 Screen (40% ethanol, 100 mM Tris pH8.5, 50 mM MgCl<sub>2</sub>) (Emerald Biosystem) by a hanging drop vapour diffusion method using 6.7 mg/ml protein and 96-well polystyrene microplate (Greiner Bio-one). The crystallization condition was subsequently optimized in 24-well plate (HAMPTON). The best crystal was obtained when 1  $\mu$ l of 6.7 mg/ml protein was mixed with 1  $\mu$ l of mother liquid

containing 35% v/v ethanol, 100 mM Tris pH8.5, 50 mM MgCl<sub>2</sub> and equilibrated against 1 ml of same mother liquid. The crystal was soaked in cryoprotectant solution containing 40% v/v ethanol, 100 mM Tris pH8.5, 50 mM MgCl<sub>2</sub> for a few seconds and flash-frozen in liquid nitrogen. Single-wavelength anomalous data was collected from a single crystal under N<sub>2</sub> cryostream (100K) at BL-1A beamline of Photon Factory at KEK (Tsukuba, Japan) using a PILATUS 2M-F detector. The crystal belonged to H32, with cell dimension of *a* = *b* = 49.42 Å, *c* = 199.14 Å.

The X-ray diffraction data were indexed using X-ray detector software (XDS) (30) and the space group was determined using Pointless (31). The scaling was performed using SCALA (32). Then phenix.autosol (33) was used for the identification of the Se site and calculation of the phase. The initial model was built in phenix.autosol wizard and manually modified using Coot (34). The structure was refined by Refmac5 (35) and phenix.refine with manual modification in Coot. The final model includes the residues Gly561–Val569, Ile587–Asn633 and the C-terminal hexahistidine tag except for the last histidine residue. The electron density corresponding to the residues Asp570–Ala586 was missing, therefore, the connectivity of the polypeptide chain between Val569 and Ile587 was unconfident. The statistics of data collection and refinement are summarized in Supplementary Table S2.

#### Bioinformatics

Alignment of amino acid sequences was carried out using ClustalW2 (36) with default parameters. Structure superposition of VgrG1C<sup>G561</sup> to phi92 gp138 was performed with PDBeFold (Protein Data Bank in Europe) (37). The conservation of the amino acids was analysed by WebLogo (38) using aligned amino acid sequences.

#### Molecular graphics

All the images of protein structure were drawn using MolFeat v4.5 (FiatLux Corporation, Japan).

## Results

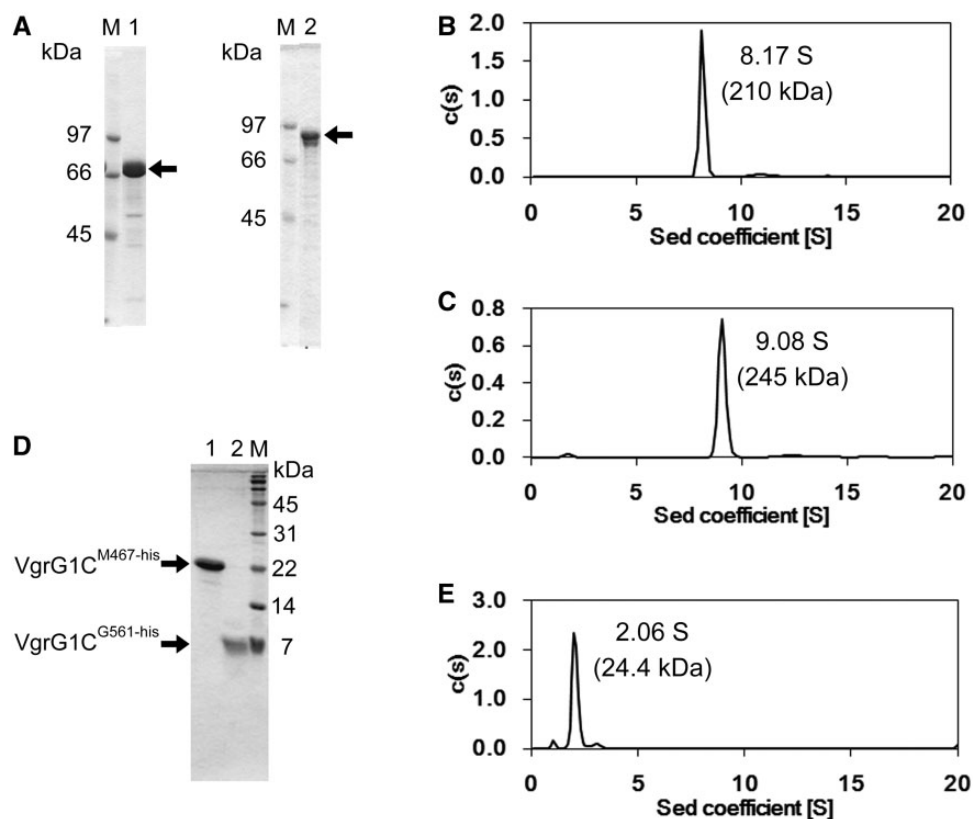
### Expression and purification of VgrG proteins from *E. coli* O157 and CFT073 strains

We constructed the over-expression vectors of VgrG1<sup>his</sup>, VgrG2<sup>his</sup>, VgrG3<sup>his</sup> from *E. coli* O157 and c3393<sup>his</sup>, c1888<sup>his</sup>, c1883<sup>his</sup> from *E. coli* CFT073 (see Materials and Methods section for details) and examined their expression profile in *E. coli* cells (Supplementary Fig. S1). All of these six VgrG proteins were expressed in high level, but most of them tend to become insoluble. However, VgrG1<sup>his</sup> of *E. coli* O157 and c3393<sup>his</sup> from *E. coli* CFT073 were obtained in soluble form by ~25% and ~16% of the total protein, respectively, which was enough to purify for further investigations.

VgrG1<sup>his</sup> and c3393<sup>his</sup> were purified by Ni affinity chromatography by using their hexahistidine tag at C-terminus, anion exchange chromatography and size exclusion chromatography at 4°C (see Materials and Methods section for details). Both VgrG1<sup>his</sup> and c3393<sup>his</sup> eluted as a single peak from size exclusion column and gave nearly a single band in SDS–PAGE (Fig. 1A).

### Both VgrG1<sup>his</sup> and c3393<sup>his</sup> are trimers in solution and rich in $\beta$ -structure

Sedimentation velocity analysis revealed that VgrG1<sup>his</sup> and c3393<sup>his</sup> form a homogeneous oligomer with the sedimentation coefficients of 8.17S and 9.08S, respectively (Fig. 1B and C). The molecular weight of VgrG1<sup>his</sup> was estimated to be 210,000, which is 2.9 times the molar mass of monomer from amino acid



**Fig. 1** Purification and characterization of *E. coli* O157 VgrG1<sup>his</sup> and *E. coli* CFT073 c3393<sup>his</sup>. (A) SDS-PAGE of *E. coli* O157 VgrG1<sup>his</sup> and *E. coli* CFT073 c3393<sup>his</sup>. Lane M, molecular weight marker; lane 1, *E. coli* O157 VgrG1<sup>his</sup> after size exclusion chromatography; lane 2, *E. coli* CFT073 c3393<sup>his</sup> after size exclusion chromatography. The polyacrylamide gel (10%) was stained with Coomassie brilliant blue. The arrows indicate VgrG1<sup>his</sup> and c3393<sup>his</sup>. Faint bands below VgrG1<sup>his</sup> and c3393<sup>his</sup> would be partially degraded VgrG1<sup>his</sup> and c3393<sup>his</sup>. (B) Sedimentation velocity analysis of *E. coli* O157 VgrG1<sup>his</sup>. The sedimentation velocity data was analysed 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  $c(M)$ . (C) Sedimentation velocity analysis of *E. coli* CFT073 c3393<sup>his</sup>. The sedimentation velocity data was analysed by SEDFIT. The measurement condition is same as (B). (D) Limited proteolysis of VgrG1<sup>M467-his</sup> by trypsin. VgrG1<sup>M467-his</sup> was digested by trypsin at 25°C for 30 min and analysed by SDS-PAGE (20%). Lane M, molecular weight marker; lane 1, VgrG1<sup>M467-his</sup> before trypsin digestion; lane 2, VgrG1<sup>M467-his</sup> after trypsin digestion. The stable fragment around 7 kDa marker (VgrG1<sup>G561-his</sup>) was identified as the C-terminal region of VgrG1 (residues Gly561–Asn633) with hexahistidine tag by protein sequencing and Ni affinity chromatography. (E) Sedimentation velocity analysis of *E. coli* O157 VgrG1<sup>G561-his</sup>. Sedimentation velocity data was analysed by SEDFIT. The moving boundaries were measured at 280 nm at 20°C. The rotor speed was 50,000 rpm. The molecular weight was determined by the sedimentation equilibrium analysis (data not shown).

sequence (72.6 kDa). Similarly, the molecular weight of c3393<sup>his</sup> was estimated as 245,000, which is 2.7 times the molar mass of monomer (92.1 kDa). We, therefore, concluded that both VgrG1<sup>his</sup> and c3393<sup>his</sup> are trimers in solution.

We also examined the secondary structure of VgrG1<sup>his</sup> and c3393<sup>his</sup> using far-UV CD Spectrum. The secondary structure estimation indicated that both VgrG1<sup>his</sup> and c3393<sup>his</sup> are rich in  $\beta$ -structure ( $\beta$ -sheet content of 33% for VgrG1 and 34% for c3393) (Table I). As it was reported earlier, VgrG protein is a homologue of the T4 phage gp27–gp5 complex without the lysozyme domain of gp5 (3, 14, 39), we compared their secondary structure contents. The secondary structure contents of VgrG1<sup>his</sup> and c3393<sup>his</sup> were very similar to that of the crystal structure of gp27–gp5 without the lysozyme domain (Table I).

#### The $\beta$ -helix domain of VgrG2 and VgrG3 from *E. coli* O157 is prone to aggregation

It has been reported that the fusion of SlyD (*E. coli* prolyl cis-trans isomerase) could enhance the solubility of aggregation prone proteins (21, 40). Therefore, we

**Table I.** Secondary structure composition of VgrG1<sup>his</sup> and c3393<sup>his</sup>.

Protein	$\alpha$ -helix (%)	$\beta$ -sheet (%)	Turn (%)
VgrG1 <sup>his</sup>	17	33	23
c3393 <sup>his</sup>	15	34	21
T4 gp27-gp5 without gp5-lysozyme domain <sup>b</sup>	11	35	29

<sup>a</sup>The secondary structure was estimated by CONTINLL program using the CD spectrum data at wavelength range of 200–240 nm.

<sup>b</sup>The secondary structure composition was calculated from the crystal structure of T4 gp27–gp5 complex (Protein Data Bank code 1k28).

constructed the over-expression vector of the putative  $\beta$ -helix domain of VgrG2 (residues 474–712) and VgrG3 (residues 458–713) with N-terminal cleavable hisSlyD (see Materials and Methods section for details). Both hisSlyD-VgrG2<sup>G474</sup> and hisSlyD-VgrG3<sup>G458</sup> were expressed in soluble form for more than 80% and 70% of the total protein, respectively. We attempted to purify them by Ni affinity chromatography using the N-terminal histidine tag. However,

>95% of <sup>his</sup>SlyD-VgrG2C<sup>G474</sup> and <sup>his</sup>SlyD-VgrG3C<sup>G458</sup> did not bind to the Ni column (data not shown). In addition, sedimentation velocity analysis of <sup>his</sup>SlyD-VgrG2C<sup>G474</sup>, which had been further purified by anion exchange chromatography and size exclusion chromatography, showed a broad peak in sedimentation coefficient distribution function,  $c(s)$ . Furthermore, in the case of <sup>his</sup>SlyD-VgrG3C<sup>G458</sup>, removal of <sup>his</sup>SlyD resulted in precipitation of VgrG3C<sup>G458</sup> (Supplementary Fig. S2). Hence, we concluded that the  $\beta$ -helix of these VgrG proteins tends to aggregate.

#### Limited proteolysis of VgrG1C<sup>M467-his</sup> by trypsin

Our attempts to crystallize the full-length VgrG1<sup>his</sup> protein from *E. coli* O157 were unsuccessful. Therefore, we created VgrG1C<sup>M467-his</sup> which consists of residues 467–633 of VgrG1 (expected  $\beta$ -helix region) with C-terminal hexahistidine tag. VgrG1C<sup>M467-his</sup> was expressed in high level with ~70% of the total protein in soluble form. However, the far-UV CD spectroscopy analysis of purified VgrG1C<sup>M467-his</sup> indicated that it contains a significant amount of disordered region (data not shown). Then we performed limited proteolysis of VgrG1C<sup>M467-his</sup> in order to obtain a stable fragment suitable for crystallization. Trypsin digestion of VgrG1C<sup>M467-his</sup> gave a stable fragment in 30 min to 1 h at 25°C (Fig. 1D). Protein sequencing and Ni affinity chromatography revealed that the stable fragment starts with GTIAGSV (from residue Gly561) and contains the C-terminal histidine tag. Thus, the fragment was identified as the C-terminal 73 residues of VgrG1 with C-terminal hexahistidine tag (VgrG1C<sup>G561-his</sup>). Incompletely digested species were occasionally seen depending on the batch of purification, but they could be easily separated by size exclusion chromatography (data not shown). The sedimentation velocity analysis indicated that VgrG1C<sup>G561-his</sup> forms a homogeneous molecular species with a sedimentation coefficients of 2.06S (Fig. 1E). Sedimentation equilibrium analysis gave the molecular weight of 24,400, which is 2.8 times the molar mass of monomer (8.6 kDa, calculated from amino acid sequence). We thus concluded that VgrG1C<sup>G561-his</sup> forms a trimer in the same fashion as full-length VgrG1<sup>his</sup>.

#### Overall crystal structure of VgrG1C<sup>G561</sup>

The structure of VgrG1C<sup>G561</sup> was solved at 1.9 Å resolution by using the single-wavelength anomalous diffraction method with the help of a Se-methionine derivative that contained only one Se atom in the asymmetric unit of H32 space group (Supplementary Table S2). VgrG1C<sup>G561</sup> forms a three-stranded  $\beta$ -helix with the width of 24 Å and height of 29 Å (Fig. 2A and B). The  $\beta$ -helix of VgrG1C<sup>G561</sup> is constituted with three antiparallel  $\beta$ -sheets each consisting of six  $\beta$ -strands and one intertwined parallel  $\beta$ -sheet at C-terminus (Fig. 2A). Each  $\beta$ -strand of antiparallel  $\beta$ -sheet is constituted of five or six amino acids and connected by  $\beta$ -turn.

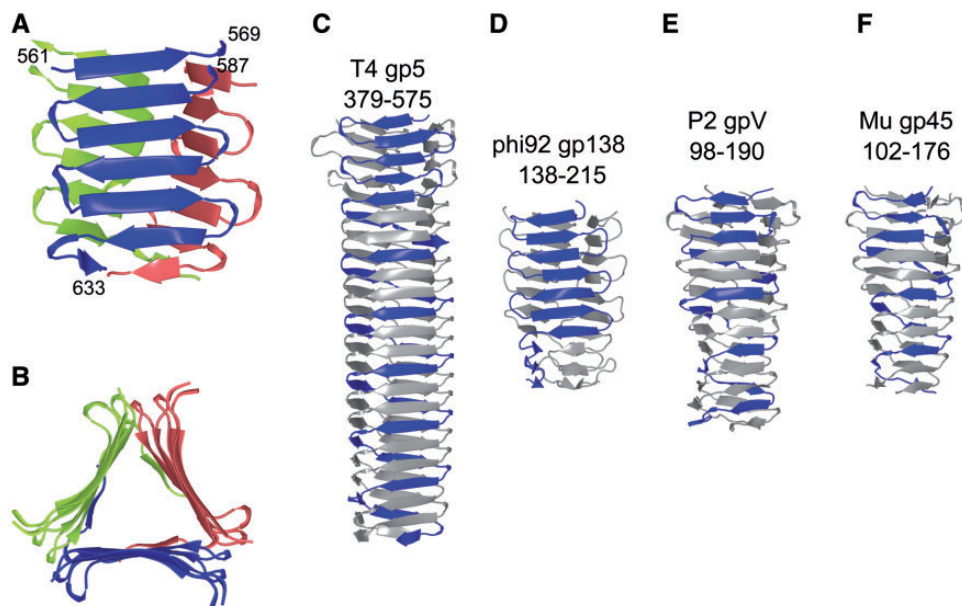
The electron density was of high quality and the atomic model for residues 561–569 and 587–641 and

the C-terminal hexahistidine tag could be easily built. However, no interpretable density for residues 570–586 was observed. Clearly, these residues are disordered because their orientation is different in different molecules comprising the crystal. The 17 residues missing from the model (570–586) raise a question about the topology of the first strand relative to the rest of the structure—the first strand can be part of any of the three molecules. We assigned the first  $\beta$ -strand to the molecule that forms a  $\beta$ -sheet together with. In this configuration the missing residues create a large loop, which might be required for binding effectors to VgrG, upon which this loop becomes ordered.

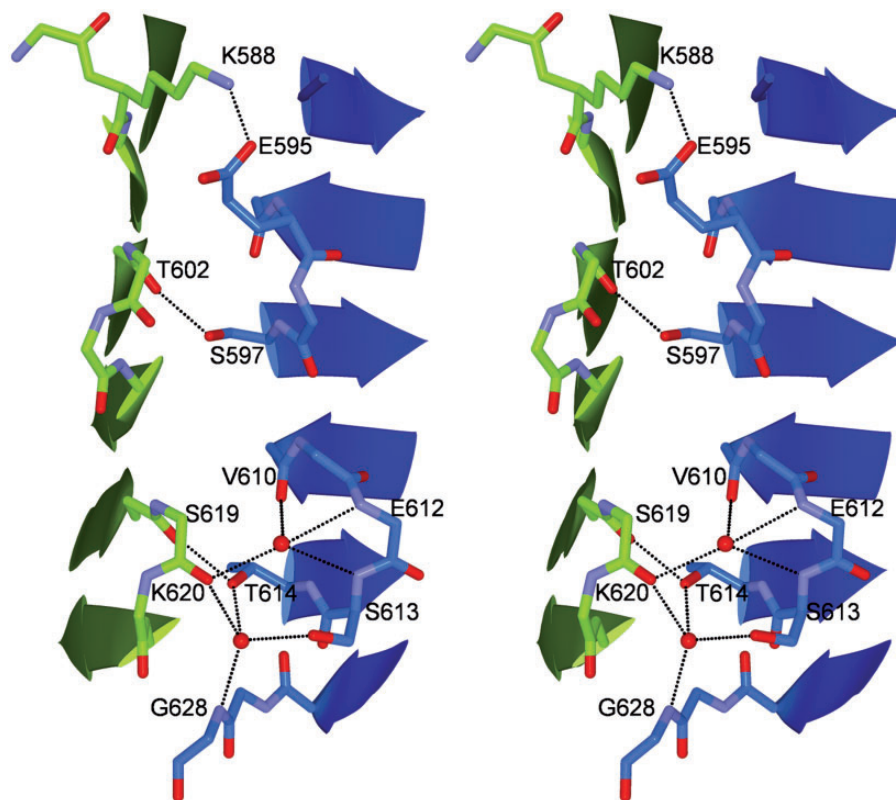
The overall structure of VgrG1C<sup>G561</sup> is similar to that of the three-stranded  $\beta$ -helix of T4 phage gp5 and its homologues—the central spike proteins from different phages (phi92 phage gp138, P2 phage gpV and phage Mu gp45) (Fig. 2); they all form three-stranded  $\beta$ -helix which is about 28 Å wide (16, 17). The amino acid sequence of VgrG1C<sup>G561</sup> was aligned with the amino acid sequences of these three-stranded  $\beta$ -helices by clustalW2 (36) but the identities were all <20%, indicating no primary sequence similarity between them. Recently, the crystal structure of PA0091 VgrG1a from *P. aeruginosa* PAO1 was solved (Sycheva L.V., Shneider M.M., Basler M., Ho B.T., Mekalanos J.J. and Leiman P.G., in preparation) and its  $\beta$ -helix is found to mainly consist of intertwined  $\beta$ -sheet. Thus the overall topology of the  $\beta$ -helix is not conserved among VgrG proteins. However, the topology of the last strand of the  $\beta$ -helix in both VgrGs is similar.

#### Sharp turns of three-stranded antiparallel $\beta$ -helix are stabilized by hydrogen bonds and salt bridge

To examine the nature of antiparallel  $\beta$ -helix of VgrG1C<sup>G561</sup>, we first focused on the turns at the corner of the  $\beta$ -helix triangular prism. In those turns, the side chains of serine and threonine face inwards where they interact with the main chain of another polypeptide chain in the trimer (Fig. 3). The side chain hydroxyl group of Ser597 and Thr614 make intermolecular hydrogen bonds with the main chain carbonyl oxygen of Thr602 and that of Ser619 from another strand, respectively (Fig. 3). In addition, the side chain hydroxyl group of Thr614 makes an intermolecular hydrogen bond with the main chain carbonyl oxygen of Lys620 from neighbouring polypeptide chain via a water molecule which is further coordinated by the side chain oxygen of Ser613 and main chain nitrogen of Gly628 (Fig. 3). Although Val594, which is present at the corner, cannot form intermolecular hydrogen bond through its side chain, Glu595 instead makes an intermolecular salt bridge with the side chain of Lys588 (Fig. 3). In the case of Val610, the main chain carbonyl oxygen of Val610 and main chain amide nitrogen of Glu612 and main chain amide nitrogen of Ser613 make an intermolecular hydrogen bond with the main chain carbonyl oxygen of Lys620 via the same water molecule (Fig. 3). This hydrogen bond network and salt bridge are likely to be important for stability of the trimer.



**Fig. 2** Crystal Structure of VgrG1C<sup>G561</sup> and comparison with the  $\beta$ -helix domain of central spike proteins of bacteriophage T4, phi92, P2 and Mu. (A) Structure of VgrG1C<sup>G561</sup> represented in ribbon diagram. Each polypeptide chain is drawn in different color. The strategic residues are labelled with their residue numbers. (B) Top-down view of VgrG1C<sup>G561</sup>. Triangular shape with 3-fold symmetry is confirmed. (C) The  $\beta$ -helix of T4 phage gp5 (residues 379–575, PDB code 1k28). (D) The  $\beta$ -helix of phi92 phage gp138 (residues 138–215, PDB code 3PQI). (E) The  $\beta$ -helix of P2 phage gpV (residues 98–190, PDB code 3QR7). (F) The  $\beta$ -helix of Mu phage gp45 (residues 102–176, PDB code 3VTO). One polypeptide chain is drawn in different colour (C–F) to make it easier to see the topology of the  $\beta$ -helix (C–E).



**Fig. 3** Stereo diagram of the hydrogen bonds and salt bridge network at the corner of the  $\beta$ -helix. The  $\beta$ -strands are drawn in ribbon model and the main chain of the  $\beta$ -turn and the side chains which make intermolecular hydrogen bond are represented in stick model. The hydrogen bonds and salt bridge are shown in dashed line with the residues numbers. To make it easy to see the interactions, only the interface of two polypeptide strands is depicted.

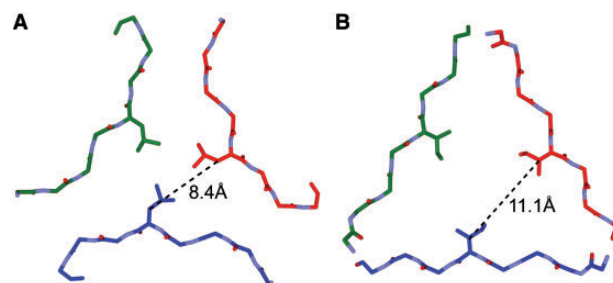
Similar intermolecular hydrogen bonds are observed in gp138 of phi92 phage (PDB code 3PQH), in which the hydrogen bonds are made between asparagine or aspartic acid and the main chain oxygen or nitrogen of neighbouring polypeptide chain (17). Furthermore, in the antiparallel  $\beta$ -sheet region of VgrG1a from *P. aeruginosa* PAO1, the corner is occupied by the side chain of cysteine and serine, which also creates an intermolecular hydrogen bond with the main chain oxygen of neighbouring polypeptide chain. Thus, the hydrogen bond network and salt bridges are conserved in antiparallel  $\beta$ -helix.

### The three-stranded antiparallel $\beta$ -helix is narrower than intertwined $\beta$ -helix

The inside of the VgrG1C<sup>G561</sup>  $\beta$ -helix is filled with the side chain of hydrophobic residues such as isoleucine and leucine that form its hydrophobic core. The hydrophobic core is also found in other triple-stranded  $\beta$ -helices such as T4 gp5, phi92 gp138, P2 gpV and VgrG1a of *P. aeruginosa* PAO1 regardless of their different topology (16, 17, Sycheva L.V., Shneider M.M., Basler M., Ho B.T., Mekalanos J.J. and Leiman P.G., in preparation). To further examine the structural feature of antiparallel  $\beta$ -helix, we measured the inter-residue distance between C $\beta$  of Ile and Leu which is located at the center of  $\beta$ -strand. The average of inter-residual C $\beta$  distances of Ile480, Ile496, Ile512 of T4 gp5 and Leu514, Leu546, Ile562 of *P. aeruginosa* PAO1 VgrG1a, which are located in intertwined  $\beta$ -helix, was  $10.6 \pm 0.7 \text{ \AA}$ . On the other hand, the average of inter-residual C $\beta$  distances of Leu570, Ile587 of *P. aeruginosa* PAO1 VgrG1a and Leu592, Ile599, Leu608, Leu616, Leu625 of VgrG1C<sup>G561</sup>, which are located in antiparallel  $\beta$ -helix, was  $8.5 \pm 0.8 \text{ \AA}$  which is about 2  $\text{\AA}$  shorter than that of intertwined  $\beta$ -helix (Fig. 4).

## Discussion

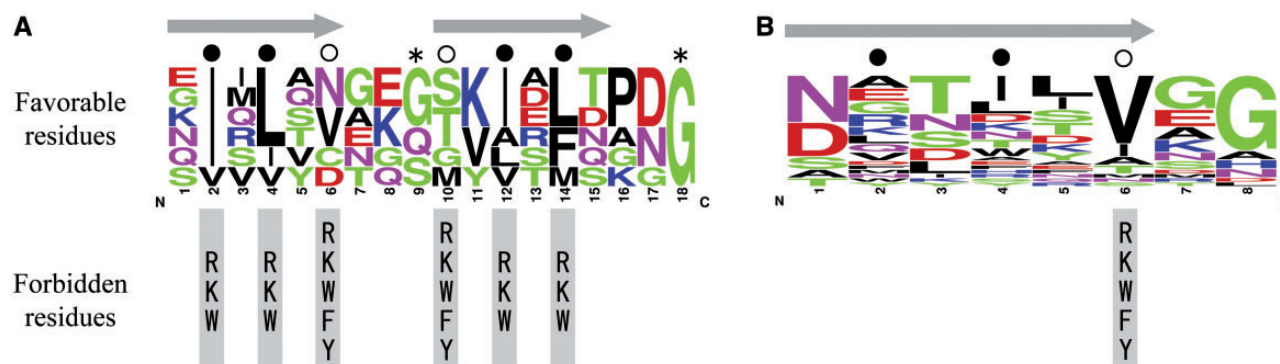
It is highly intriguing from the viewpoint of function and evolution that bacterial T6SS have a number of structurally similar proteins to contractile tails of bacteriophage (3, 14). VgrG protein in the T6SS shares three homologous domains with the gp27–gp5 cell puncturing device of T4 phage, namely, the gp27 domain, the OB fold domain and three-stranded  $\beta$ -helix (3, 14). In this study, we examined the over-expression profile of six VgrG proteins from *E. coli* O157 and CFT073 strains and found that most of them are insoluble indicating either completely incorrect folding or aggregation immediately after partially correct folding. It was unexpected, because the gp27–gp5 complex, which is a T4 phage orthologue of VgrG, is a stable and soluble protein complex (16). We hypothesized that folding of VgrG proteins can be facilitated by generic cellular chaperones—DnaK, DnaJ, GrpE, GroEL, GroES—and performed co-expression of VgrGs with these proteins. However, solubility of VgrG proteins was unaffected (data not shown). It is clear, however, that folding of VgrG proteins requires specialized molecular chaperones or partner molecules. A good candidate to perform this function would be proline-alanine-alanine-arginine



**Fig. 4** The core of antiparallel  $\beta$ -helix is narrower than intertwined  $\beta$ -helix. (A) The main chain of the antiparallel  $\beta$ -helix of VgrG1 (residues 604–611) and the side chain of Leu608 are drawn in stick model. The intermolecular C $\beta$  distance of Leu608 is displayed. (B) The main chain of the intertwined  $\beta$ -helix of T4 gp5 (residues 476–483) and the side chain of Ile480 is drawn in the stick model in the same scale as (A). The intermolecular C $\beta$  distance of Ile480 is displayed. Note that parallel  $\beta$ -helix is narrower than intertwined  $\beta$ -helix.

(PAAR)-repeat motif proteins that have been recently shown to bind to the C-terminal blunt end of the  $\beta$ -helix of T4 gp5–VgrG chimeric protein (41). The hypothesized function of these PAAR motif proteins is to form a nucleation point to facilitate the folding of the topologically complex  $\beta$ -helical domain of VgrG proteins. Interestingly, inactivation of all PAAR motif proteins led to functional defect of T6SS in *V. cholerae* and *Acitenobacter baylyi* (41). PAAR motifs are often a part of an RHS protein-encoding gene, which is frequently found downstream of a *vgrG* gene (41) and indeed, every T6SS locus of *E. coli* O157 contains an RHS gene after a *vgrG* gene. VgrG and RHS proteins are extremely diverse and not every bacterium with a functional T6SS contains an RHS protein, but at least in *E. coli* O157 RHS proteins may function as molecular chaperones of VgrGs.

The crystal structure of the N-terminal region of *E. coli* CFT073 c3393 has been already reported and was shown that it forms a trimer that is structurally similar to T4 gp27–gp5 complex, but the lysozyme domain is absent (14). In the present study, we showed that full-length VgrG1<sup>his</sup> of *E. coli* O157 and c3393<sup>his</sup> of *E. coli* CFT073 are also trimer in solution and contain similar secondary structure as that of the T4 gp27–gp5 complex without the lysozyme domain, which supports the idea that VgrG proteins contain three-stranded  $\beta$ -helix as the T4 gp5 does (3). We determined the crystal structure of C-terminal 73 residues of VgrG1 (VgrG1C<sup>G561</sup>) by X-ray crystallography and demonstrated that they indeed form a three-stranded  $\beta$ -helix. The overall structure of VgrG1C<sup>G561</sup> was similar to the three-stranded  $\beta$ -helix domain of T4 gp5, P2 gpV and phi92 gp138. Out of these three proteins, the  $\beta$ -helix of phi92 gp138 was most similar to VgrG1C<sup>G561</sup> in topology of the polypeptide chain. The rmsd (root mean square deviation) of residues 587–633 region of VgrG1C<sup>G561</sup> and the corresponding residues 155–202 of phi92 gp138 was 1.22  $\text{\AA}$  over 46 amino acids on C $\alpha$  atoms, which indicated that not only the topology but also the three-dimensional structure is similar. The VgrG1C<sup>G561</sup>



**Fig. 5 Consensus sequence and structurally forbidden amino acids of antiparallel and intertwined  $\beta$ -helix.** (A) The antiparallel  $\beta$ -helix regions of VgrG1C<sup>G561</sup>, VgrG1a of *P. aeruginosa* PAO1 and phi92 gp138 were aligned manually and conservation of the amino acid was visualized by WebLogo (38) with frequency plot option. The arrow indicates the  $\beta$ -strands. The closed circle indicates the inward pointing residues and open circle indicates the amino acid whose side chain occupies the inside of the corner. Asterisk indicates the occasional insertion in  $\beta$ -turn. Sterically forbidden amino acids were identified by manually introducing mutation using Coot. (B) The intertwined  $\beta$ -helix regions of T4 gp5 and VgrG1a of *P. aeruginosa* PAO1 were aligned manually and visualized by WebLogo with frequency plot option. Same symbols are used as (A). Sterically disallowed amino acids were identified in the same way as (A).

shows further similarity with phi92 gp138 in the stabilization mechanism of trimer; both of VgrG1C<sup>G561</sup> and phi92 gp138 are stabilized by the hydrophobic interaction and the intermolecular hydrogen bonds at the corner of  $\beta$ -helix. *Escherichia coli* is one of the hosts of phi92 phage. Therefore, it is possible that VgrG1C<sup>G561</sup> is evolutionarily related with gp138 of phi92 phage.

We found that VgrG1C<sup>G561</sup> contains a long flexible loop between Asp570 and Ala586, but it was not expected before the structure was determined, because the central spike proteins of bacteriophages with known structure do not have a loop insertion within the three-stranded  $\beta$ -helix. However, a long loop insertion before three-stranded  $\beta$ -helix was observed in T4 gp5 (16) and the loop after three-stranded  $\beta$ -helix was recently found in VgrG1a of *P. aeruginosa* (Sycheva L.V., Shneider M.M., Basler M., Ho B.T., Mekalanos J.J. and Leiman P.G., submitted for publication, in preparation). In the case of T4 gp5, a lysozyme domain is inserted before  $\beta$ -helix and it is connected to  $\beta$ -helix by a 51-residues loop. On the other hand, VgrG1a of *P. aeruginosa* contains a loop after  $\beta$ -helix though no enzymatic domain is present. It is, however, reported that VgrG proteins occasionally carry an enzyme domain, called effector domain, at the C-terminus (*e.g.* actin crosslinking domain of *V. cholerae* VgrG1 and lysozyme-like chitosanase domain of *V. cholerae* VgrG3) (3, 8) and those domains are predicted to be flanked by a long loop (42). Taking these facts into consideration, it is possible that a certain VgrG protein might carry an effector domain within  $\beta$ -helix though such a VgrG has not been reported yet.

The three-stranded  $\beta$ -helices can be classified into two categories, namely, antiparallel  $\beta$ -helix and intertwined  $\beta$ -helix. However, it is not obvious to decide which sequence dictates which conformation. Until now, a number of the structures of  $\beta$ -helices have been determined (16, 17) which can be categorized as two topologically different types. In order to formulate a consensus sequence for either antiparallel  $\beta$ -helix or intertwined  $\beta$ -helix, we performed structure-based sequence alignment of each category and examined the

conservation of the amino acids along each sequence (Fig. 5A and B).

The sequence alignment of the antiparallel  $\beta$ -helix region of VgrG1C<sup>G561</sup>, VgrG1a of *P. aeruginosa* PAO1 and phi92 gp138 revealed three types of amino acid preferences. The first type was observed in the inward pointing residues that form the core of the  $\beta$ -helix (Fig. 5A, closed circle). Hydrophobic residues such as leucine, isoleucine and valine are preferred and form a hydrophobic core to stabilize the trimer as we described in the ‘Results’ section. The second type was seen in the inward pointing residues whose side chain fills the inside of the corner (Fig. 5A, open circle). Small polar residues such as asparagine, serine and threonine are preferred at the positions and they form intermolecular hydrogen bonds. The last type was seen in the second  $\beta$ -turn (Fig. 5A, position 16 and 17). The preference here is proline, asparagine and aspartic acid, which are typical  $\beta$ -turn forming residues.

On the other hand, the sequence alignment of the intertwined  $\beta$ -helix region of T4 gp5 and that of VgrG1a of *P. aeruginosa* PAO1 indicated the preference of asparagine, aspartic acid at the first amino acid of the  $\beta$ -strand, valine, isoleucine at the last amino acid of the  $\beta$ -strand and glycine at the kink between each  $\beta$ -strand (Fig. 5B). Some of these preferences of the amino acids have been already reported and explained based on three dimensional structures (16, 17).

We next, examined the sterically disallowed amino acids, *i.e.* unacceptable amino acids due to the predicted steric hindrance, at inward pointing residues in order to make the prediction more accurate. As shown in Fig. 4, inside of the antiparallel  $\beta$ -helix is narrower as compared with that of the intertwined  $\beta$ -helix. Therefore, amino acids with bulky or long side chain, *i.e.* tryptophan, arginine and lysine, are not acceptable (Fig. 5A). Other bulky amino acids such as phenyl alanine and tyrosine are acceptable, but the inward pointing neighbor residue has to be small such as valine, alanine or glycine. On the other hand, the inside of intertwined  $\beta$ -helix is wider than that of the



antiparallel  $\beta$ -helix so that the amino acids with bulky or long side chain are acceptable. In fact, tryptophan, arginine and lysine are seen in the inside of T4 gp5  $\beta$ -helix (16). In both antiparallel and intertwined  $\beta$ -helix, however, the corner of the  $\beta$ -helix cannot accommodate bulky side chains. The side chains are flanked by the backbone and the side chain of the neighbouring  $\beta$ -sheet, thus bulky amino acids such as phenylalanine and tyrosine are not allowed in addition to tryptophan, arginine and lysine (Fig. 5).

We applied this consensus sequence with sterically forbidden amino acids (Fig. 5) to VgrG3 of *E. coli* O157 and c3393 of *E. coli* CFT073 and predicted that these VgrG proteins would form intertwined  $\beta$ -helix and mixture of intertwined and antiparallel  $\beta$ -helix, respectively (Supplementary Fig. S3). We are going to solve their atomic structure by X-ray crystallography and confirm their predicted topology of the  $\beta$ -helix.

## Supplementary Data

Supplementary Data are available at *JB* Online.

## Acknowledgements

We thank the 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. We thank Mikhail Shneider for his expression vectors containing the SlyD chaperone.

## Funding

Grant-in-Aid for Young Scientist (no. 21770164 to S.K.); Grant-in-Aid for Scientific Research (C) (no. 23570190 to F.A.); International Training Program from the Japan Society for the promotion of Science; the Swiss National Science Foundation (grant 31003A\_127092 to P.G.L.); and EPFL (for P.G.L.'s research work).

## Conflict of interest

None declared.

## References

- 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
- Boyer, F., Fichant, G., Berthod, J., Vandembrouck, 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
- 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. *Proc. Natl Acad. Sci. USA* **104**, 15508–15513
- 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. *Proc. Natl Acad. Sci. USA* **107**, 4365–4370
- 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

- 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
- 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
- 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*. *Proc. Natl Acad. Sci. USA* **110**, 2623–2628
- Aschtgen, M.S., Bernard, C.S., De Bentzmann, S., Llobes, 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
- 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
- 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
- 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
- Lescic, 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
- 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. *Proc. Natl Acad. Sci. USA* **106**, 4154–4159
- Leiman, P.G., Chipman, P.R., Kostyuchenko, V.A., Mesyanzhinov, V.V., and Rossmann, M.G. (2004) Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* **118**, 419–429
- 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
- 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
- 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

19. 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 Crystallogr Sect F Struct Biol Cryst Commun* **67**, 837–841
20. 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
21. 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
22. 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
23. 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
24. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685
25. 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
26. 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
27. 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
28. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619
29. Laue, T.M., Shah, B.D., Ridgeway, T.M., and Pelletier, S.L. (1992) Computer-aided interpretation of analytical sedimentation data for proteins in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S.E., Rowe, A.J., and Horton, J.C., eds.) pp. 90–125, Royal Society of Chemistry, Cambridge, UK
30. Kabsch, W. (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
31. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82
32. Collaborative Computational Project, N. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763
33. 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 Crystallogr. D Biol. Crystallogr.* **66**, 213–221
34. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
35. Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
36. 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
37. Krissinel, E. and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2256–2268
38. Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190
39. Kanamaru, S. (2009) Structural similarity of tailed phages and pathogenic bacterial secretion systems. *Proc. Natl Acad. Sci. USA* **106**, 4067–4068
40. 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
41. 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
42. 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