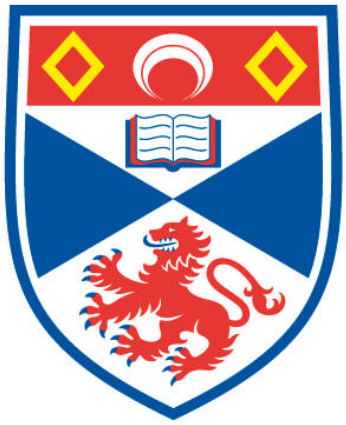
STRUCTURAL AND FUNCTIONAL STUDIES OF BACTERIAL OUTER MEMBRANE LIPOPOLYSACCHARIDE INSERTION AND SCHMALLENBERG VIRUS REPLICATION

Haohao Dong

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2015

Structural and functional studies of bacterial outer membrane lipopolysaccharide insertion and Schmallenberg virus replication

University of St Andrews

School of Chemistry

A thesis submitted for the degree of

Doctor of Philosophy

Haohao Dong

July 2015

Supervisor: Prof. Changjiang Dong and Prof. James H. Naismith

Declarations

I, Haohao Dong, hereby certify that this thesis, which is approximately 28,000 words in length, has been written by me, and that it is the record of work carried out by me, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in April, 2011 and as a candidate for the degree of Doctor of Philosophy in April, 2015; the higher study for which this is a record was carried out in the University of St Andrews between 2011 and 2015.

Date: Signature of candidate:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date: Signature of supervisor:

Date: Signature of supervisor:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Access to all of printed copy and electronic publication of thesis through the University of St Andrews.

Date: Signature of candidate:

Date: Signature of supervisor:

Date: Signature of supervisor:

Abstract

Lipopolysaccharide (LPS) is an essential component of the outer membrane (OM) of Gram-negative bacteria and plays a fundamental role in protecting the bacteria from harsh environments and toxic compounds. The LPS transport system is responsible for transporting LPS from the periplasmic side of the inner membrane (IM) to the OM, in a process involving seven LptA-LptG proteins. The current model for lipopolysaccharide transport (Lpt) suggests that LPS is initially extracted by a four-protein complex, LptBCFG, from the inner membrane to the periplasm, where LptA mediates further transport to the OM. Another two protein complex, LptD/E, catalyses the assembly of LPS at the OM cell surface. However, the details of this transport mechanism have remained unknown, mainly due to a lack of structural information.

In chapter 1 and 2 of this thesis, I report materials and methods for all LptD/E, and Schmallenberg virus (SBV) nucleoprotein (NP) experiments and the theories and softwares that were used in determining structures of LptD/E, SBV NP and the SBV NP/RNA complex.

In chapter 3 of this thesis, I report the first crystal structure of the outer membrane protein LptD/E complex. LptD forms a 26-strand β-barrel in a closed form and LptE is a roll-like structure located inside LptD to form “barrel and plug” architecture. Through structural analysis, function assay and molecular dynamics simulation, we proposed a mechanism in which the hydrophilic head of LPS molecule, including the oligosaccharide core and the O antigen, directly penetrates through the hydrophilic β-

barrel whilst the hydrophobic lipid A tail is inserted into an intramembrane hole, with a lateral opening between strand β1 and β26 of the LptD. LptE may assist this process.

In chapter 4, I report the crystal structure of the SBV NP in two conformations: tetrameric when the protein was purified under native conditions, and trimeric when denatured and refolded during purification. The SBV NP has a novel fold and we have also identified that the N-terminal arm is crucial for RNA binding, and the N- and the C-terminal arm is essential for RNA multimerisation with adjacent protomers and for viral RNA encapsidation.

Chapter 5 describes the crystal structure of SBV NP in complex with a 42 nucleotide long RNA (polyU). This ribonucleoprotein (RNP) complex was crystallized as a ring-like tetramer with each protomer bound to 11 ribonucleotides. Eight of these nucleotides are bound in a positively charged cleft between N- and C- terminal domains and three are bound in the N-terminal arm. I also compared the structure to that of other NPs from negative-sense RNA viruses, and found that SBV NP sequesters RNA using a different mechanism. Furthermore, the structure suggests that when RNA binds the protein, there are conformational changes in the RNA-binding cleft, and in the N- and C-terminal arms. Thus our results reveal a novel mechanism of RNA encapsidation by orthobunyaviruses NP.

Acknowledgements

I would like to express my sincerest gratitude to my supervisors, Professor Changjiang Dong and Professor Jim Naismith, for their outstanding and excellent supervision. Their guidance, enthusiasm and inspiration, as well as their academic experience have all contributed to my enjoyment of working in this very exciting area of research. Under such outstanding supervision, I was able to solve these protein structures.

My special and sincere thank goes to Prof. Jim Naismith, who I have been being deeply indebted to for continuously supporting my studies in his laboratory. This thesis would have not been possible without his kind support, patience and profound advice.

A special thank you goes to our collaborator Prof. Richard M. Elliott and Dr. Ping Li for kindly providing the SBV NP plasmid, in addition to technical guidance and discussions.

I would also like to acknowledge the help that Dr. Stephen McMahon, Dr. Magnus Alphey, Dr. Judith Reeks, Dr. Emma Branigan, Dr. Gregor Wallat, Dr. Lucile Moynie, and Dr. Gregor Hagelüken, Dr. Hexian Huang, Dr. Lei Yang and Dr. Jane Potter have provided, in addition to all members of the CJD, JHN and the GLT groups. I owe a big thank you to Dr. Huanting Liu for encouraging me to be strong and passing on his knowledge and experience accumulated from years of academic research. I also would like to say a big thank you to Dr. Phedra Marius. I appreciate her a lot for discussing my project in depth with me throughout the year.

Last, but by no means least, I would like to express my gratitude for the BSRC support staff for providing media and plates.

Publications

Dong, H., Xiang, Q., Gu, Y., Wang, Z., Paterson, N. G., Stansfeld, P. J., He, C., Zhang, Y., Wang, W., Dong, C. (2014). Structural basis for outer membrane lipopolysaccharide insertion. *Nature*, *511*(7507), 52–6. doi:10.1038/nature13464

Dong, H., Li, P., Böttcher, B., Elliott, R. M., & Dong, C. (2013). Crystal structure of Schmallenberg orthobunyavirus nucleoprotein − RNA complex reveals a novel RNA sequestration mechanism. *RNA*, 19(8): 1129-36. doi:10.1261/rna.039057.113.5

Dong, H., Li, P., Elliott, R. M., & Dong, C. (2013). Structure of schmallenberg orthobunyavirus nucleoprotein suggests a novel mechanism of genome encapsidation. *Journal of Virology*, *87*(10), 5593–601. doi:10.1128/JVI.00223-13

Qi, X., Lan, S., Wang, W., Schelde, L. M., **Dong, H**., Wallat, G. D., Ly, H., Liang, Y., & Dong, C. (2010). Cap binding and immune evasion revealed by Lassa nucleoprotein structure. *Nature*, *468*(7325), 779–783. doi:10.1038/nature09605

Wallat, G. D., Huang, Q., Wang, W., **Dong, H**., Ly, H., Liang, Y., & Dong, C. (2014). High-resolution structure of the N-terminal endonuclease domain of the Lassa virus L polymerase in complex with magnesium ions. *PloS One*, *9*(2), e87577. doi:10.1371/journal.pone.0087577

Gu, Y., Stansfeld, P. J., Zeng, Y., **Dong, H**., Wang, W., & Dong, C. (2015). Lipopolysaccharide is Inserted into the Outer Membrane through An Intramembrane Hole, A Lumen Gate, and the Lateral Opening of LptD. *Structure*, *23*(3), 496–504. doi:10.1016/j.str.2015.01.001

Wang, Z., Xiang, Q., Zhu, X., **Dong, H**., He, C., Wang, H., … Dong, C. (2014). Structural and functional studies of conserved nucleotide-binding protein LptB in lipopolysaccharide transport. *Biochemical and Biophysical Research Communications*, *452*(3), 443–449. doi:10.1016/j.bbrc.2014.08.094

Jiang, X., Huang, Q., Wang, W., **Dong, H**., Ly, H., Liang, Y., & Dong, C. (2013). Structures of Arenaviral Nucleoproteins with Triphosphate dsRNA Reveal a Unique Mechanism of Immune Suppression. *The Journal of Biological Chemistry*, *288*(23), 16949–59. doi:10.1074/jbc.M112.420521

Table of contents

List of figures

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 5:

List of tables

Abbreviations

ABC	ATP-binding cassette
BAM	β-barrel assembly machinery
BUNV	Bunyamwera virus
C8E4	Tetraethylene Glycol Monooctyl Ether
CC	Correlation coefficient
CCHFV	Crimean-Congo hemorrhagic fever viruses
CTA	C-terminal arm
CTD	C-terminal domain
CV	Column volume
DDM	N-Dodecyl-β-D-Maltoside
EM	Electron microscopy
G	Glycoproteins
IFN	Interferon
IM	Inner membrane
IMPs	Inner membrane proteins
IPTG	Isopropyl β-*d*-thiogalactopyranoside
KDO	3-deoxy-D-manno-oct-2-ulosonic acid
L	Large
LB	Luria Broth
LDAO	lauryldimethylamine-oxide
Lol	Localization of lipoprotein
LPS	Lipopolysaccharide
Lpt	Lipopolysaccharide transport

M	Middle
MAD	Multi wavelength anomalous dispersion
MD	Molecular dynamic
NAG	N-acetylglucosamine
NAM	N-acetylmuramic acid
NBD	Nucleotide binding domain
Ni-NTA	Nickel-nitrilotriacetate affinity resin
NP	Nucleoprotein
NS	Nonstructural
nt	Nucleotide
NTA	N-terminal arm
NTD	N-terminal domain
OD600	Optical density at a wavelength of 600 nm
OM	Outer membrane
OMPs	Outer membrane proteins
ORFs	Open reading frames
PDB	Protein data bank
PEG	Polyethylene glycol
PMSF	phenylmethylsulphonyl fluoride
RdRp	RNA-dependent RNA polymerase
RMSD	Root-mean-square deviation
RNPs	Ribonucleoprotein complexes
RVFV	Rift Valley fever phlebovirus
S	Small
SBV	Schmallenberg virus

SBV NP	Schmallenberg virus nucleoprotein
SeMet	Selenomethionine
SRP	Signal recognition particle
TEV	Tobacco etch virus
TF	Trigger factor
TMD	Transmembrane domain
WHO	World Health Organization
β-OG	*N*-Octyl- β -D-glucopyranoside

Alanine	A	Ala	Leucine	L	Leu
Arginine	R	Arg	Lysine	K	Lys
Asparagine	N	Asn	Methionine	M	Met
Aspartic acid	D	Asp	Phenylalanine	F	Phe
Cysteine	C	Cys	Proline	P	Pro
Glutamine	Q	Gln	Serine	S	Ser
Glutamic acid	E	Glu	Threonine	T	Thr
Glycine	G	Gly	Tryptophan	W	Trp
Histidine	H	His	Tyrosine	Y	Tyr
Isoleucine	I	Ile	Valine	V	Val

Chapter 1

Materials and methods for LptD/E, SBV NP and SBV NP/RNA complexes

1.1 Materials and methods for LptD/E

1.1.1 Medium and buffers

LB medium (Luria-Bertani)	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl.
M9 medium (1000 ml)	1 g NH_4Cl, 3 g KH_2PO_4, 6 g Na_2HPO_4, 0.5 g NaCl 20 g Glucose, 0.3 g $MgSO_4$, 0.01 g $Fe_2(SO_4)_3$ and 0.01 g Thiamine
Cell lysis buffer	20 mM Tris-HCl, pH 7.8, and 150 mM NaCl, supplemented with complete protease inhibitor mixture tablet (Roche), 1 μM DNAse (Sigma), 1 μM lysozyme (Fluka), 0.1 mM phenylmethylsulphonyl fluoride (PMSF, Sigma-Aldrich).
IM Solubilization buffer	The cell lysis buffer with 0.5% (w/v) N-lauroylsarcosine sodium salt (Sigma-Aldrich).
OM Solubilization buffer	20 mM Tris-HCl, pH 7.8, 300 mM NaCl, 10 mM imidazole，5% (v/v) glycerol and 2 % (w/v) 3-(N,N-Dimethylmyristyl-ammonio- propanesulfonate) (Sigma-Aldrich).
Wash buffer	20 mM Tris-HCl, pH 7.8, 300 mM NaCl, 30 mM imidazole, 5 % glycerol (v/v) and 1% (w/v) *N*-Octyl- β -D-glucopyranoside (β-OG; Anatrace).
Elution buffer	20 mM Tris-HCl, pH 7.8, 300 mM NaCl, 300 mM imidazole, 5 % glycerol (v/v) and 1% (w/v) β-OG).
Gel filtration buffer	20 mM Tris-HCl, pH 7.8, 100 mM NaCl, 5 % glycerol (v/v) and 1% (w/v) β-OG).

1.1.2 Generation of the LptD/E constructs

The genes *lptD and lptE* of *Salmonella enterica typhimurium* strain LT2 were amplified by PCR and inserted into pET28a (+) and pACYC-Duet-1 (Novagen), respectively. The hexahistidine tag was introduced into the C-terminus of LptE.

1.1.3 LptD/E transformation

Plasmids pET-28a-lptD and pACYC-Duet-1-lptE were co-transformed into an expression strain of *E. coli* subtype C43 (DE3) cells (Avidis). 100 ng of plasmids DNA was added to 50 µl of competent C43 (DE3) cells. The mixture was incubated on ice for 30 minutes, heat-shocked at 42°C in a water-bath for 90 seconds and then chilled on ice for 3 minutes. After that, the mixture was added 100 µl of Luria Broth (LB) and incubated at 37°C for 1 hour. Subsequently, the cell culture was plated onto a L-agar plate containing 30 µg ml^{-1} kanamycin and 34 µg ml^{-1} chloramphenicol and incubated at 37°C overnight.

A single colony was picked and propagated in 10 ml LB medium (200 RPM, 37°C) containing the antibiotics kanamycin (30 µg ml^{-1}) and chloramphenicol (34 µg ml^{-1}). After 9 hours, a glycerol stock was prepared (15% glycerol), flash frozen in liquid nitrogen and stored at -80°C.

1.1.4 Protein expression of LptD/E

500 ml LB medium was inoculated with the glycerol stock and supplemented with antibiotics kanamycin (30 μg ml^{-1}) and chloramphenicol (34 μg ml^{-1}). This culture was incubated overnight at 37 °C with shaking at 200 RPM. Each litre of LB medium (supplemented with antibiotics) was inoculated with 40 ml of the overnight culture. The cells were incubated at 200 RPM, 26 °C until the optical density (OD) at 600nm reached 0.8. Protein expression was induced by addition of isopropyl β-*d*-thiogalactopyranoside (IPTG) to a final concentration of 0.1mM with shaking at 200 RPM, for 24 hours at 26 °C. The cells were harvested by centrifugation at 6,000 × g for 15 minutes at 4 °C.

(L)- selenomethionine labelled LptD/E was expressed in M9 medium. The next day, the cells were pelleted by centrifugation at 6,000 × g for 10 minutes (Beckman Coulter). The pellet was then washed with 500 ml of sterilised PBS and re-suspended in 200 ml of PBS. 20 ml of the suspension was used to inoculate 1 L of M9 medium with 50 ml of SeMet nutrient mix (Molecular Dimensions) according to the manufacturer's instructions (5.1g SeMet was dissolved in 50 ml steriled water, and then filtered with 0.22 μm sterile filter). The cultures were incubated at 26°C, 200 RPM until the optical density (OD) of the bacterial cell culture reached mid-log phase (OD_{600} = 0.6-0.7) at a wavelength of 600 nm. Then, 10 ml (100x) amino acids lysine, phenylanine, threonine, isoleucine, leucine and valine were added to the cell cultures to inhibit methionine biosynthesis. After 25 minutes, SeMet (Generon) was added to the culture at a final concentration of 100 μg ml^{-1}. The cells were incubated at 200 RPM and 26°C for a further 25 minutes. Protein overexpression was induced by adding isopropyl β-*d*-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM. The cells were cultured at 26°C for 24 hours before they were harvested by

centrifugation.

1.1.5 Purification of LptD/E

Cell pellets were resuspended in 200 ml of lysis buffer (20 mM Tris-HCl, pH 7.8, and 150 mM NaCl), supplemented with cOmplete protease inhibitor mixture tablets (Roche), 1 μM DNAse (Sigma), 1 μM Lysozyme (Fluka) and 0.1 mM phenylmethylsulphonyl fluoride (PMSF, Sigma-Aldrich). The cells were lysed by passing through a cell disruptor twice at 30,000 psi (Constant Systems Ltd). The cell debris was removed by centrifugation at 4000 × g for 25 minutes at 4°C. The cell membranes were pelleted by ultracentrifugation at 120,000 × g for 1 hour at 4°C. The inner membrane fraction was solubilized by suspending the pellet in the IM solubilization buffer containing 0.5% (w/v) N-lauroylsarcosine sodium salt (Sigma-Aldrich) with rocking for 3 hours at room temperature. The outer membrane was pelleted by ultracentrifugation at 120,000 × g for 1 hour, and then solubilized in the OM solubilization buffer (20 mM Tris-HCl, pH 7.8, 300 mM NaCl, 10 mM imidazole，5% (v/v) glycerol and 2% (w/v) 3-(N,N-Dimethylmyristyl-ammonio-propanesulfonate) (Sigma-Aldrich) and incubated for 5 hours at 4°C.

The suspension of solubilized outer membrane was then ultra-centrifuged at 100,000 × g for 1 hour, and the supernatant was loaded onto a nickel-nitrilotriacetate affinity resin (Ni-NTA, Qiagen) column (5ml). The resin was washed with 10 column volumes (CV) of wash buffer (20 mM Tris-HCl, pH 7.8, 300 mM NaCl, 30 mM imidazole, 5 % glycerol (v/v) and 1% (w/v) *N*-Octyl- β -D-glucopyranoside (β-OG; Anatrace). The LptD/E complex was eluted with 2 CV elution buffer (20 mM Tris-HCl, pH 7.8, 300 mM NaCl, 300 mM imidazole, 5 % glycerol (v/v) and 1% (w/v) β-OG.

The protein was further purified using size exclusion chromatography using a HiLoad 16/60 Superdex 200 prep grade column (GE Healthcare) with gel filtration buffer containing (20 mM Tris-HCl, pH 7.8, 100 mM NaCl, 5 % glycerol (v/v) and 1% (w/v) β-OG). Two peaks appeared. The detection of the LptD/E complex in fractions was carried out by SDS-polyacrylamide gel electrophoresis (NuPAGE 4-12% Bis-Tris) using Mark12 protein molecular marker (Invitrogen) with conditions: 1x MES running buffer at 200 V, 120 mA for 35 minutes. The gel was stained with Coomassie blue R250 (Coomassie brilliant blue R250, Methanol, Acetic acid). Gels bands corresponding to the LptD/E molecular weight were cut and identification was confirmed by mass spectroscopy (University of St Andrews).

1.1.6 Crystallization and data collection of LptD/E

1.1.6.1 Crystallization of LptD/E

Crystallization trials were performed by Cartesian^{TM} Honeybee robot (Genomic solutions LTD). The protein was screened in 96-well sitting drop crystallization plates. Each crystallization drop was built up using 0.15 µl of protein and 0.15 µl of the crystallization solution, sitting in a reservoir containing 70 µl of crystallization solution. The crystallization screening kits attempted include PEG/ION (Hampton research), MemGold 1&2, MemSys & MemStart, MemPlus (Molecular Dimensions Ltd.) and crystallization kit for OM (Sigma). The crystallization plates were incubated at 4°C and 20°C separately.

1.1.6.2 Crystallization failed

After two weeks, the crystal trials were checked under microscope, but no crystals were formed. In order to obtain crystals, several detergents have been tried including

tetraethylene glycol monooctyl ether (C8E4), lauryldimethylamine-oxide (LDAO) and n-Dodecyl-β-D-Maltoside (DDM). The LptD/E protein complex was purified using the method described above and crystallization trials were set up with different detergents. However, these detergents did not help to form crystals. I then carried out limited proteolysis on the purified protein in an attempt to obtain crystals.

1.1.6.3 Limited proteolysis

It was desalted with gel filtration buffer using a desalting column Hi-Prep^{TM} 26/10, (GE Healthcare). The protein was pooled and concentrated to approximately 5 mg ml^{-1}, using Vivaspin concentrators with 100-kDa molecular weight cut-off (MWCO). Firstly, small-scale experiments were performed using proteases α-chymotrypsin, trypsin, V8, papain or thermolysin (Sigma-Aldrich). They were individually added to purified LptD/E protein at 1:1000 and 1:100 (protease-to-protein molar ratios). These mixtures were incubated at room temperature for digestion at the time intervals of 30, 60, 90, 120, 150, 180 minutes. The samples were subjected to SDS-PAGE and the result showed that only α-Chymotrypsin and trypsin can be used to obtain stable LptD/E complex.

1.1.6.4 Crystallization of resulting LptD/E protein complex

The α-Chymotrypsin resulting LptD/E complex was screened for crystallization using the crystallization robot Honeybee. The LptD/E crystals grew within 7 days in a condition of 0.15 M zinc acetate, 0.08 M sodium cacodylate pH 6.5 and 15% (w/v) PEG 8000. The crystals were optimized by varying pH 6.0-6.9 (sodium cacodylate) and the concentration of 10-19% (w/v) Polyethylene glycol (PEG) 8000 using 96 well Crystal clear sitting–drop plates by mixing 1 µl of protein and 1 µl of crystallization precipitant with 100 µl reservoirs solution. The optimised plates were incubated at

room temperature (20°C).

1.1.6.5 Determination of the quality of LptD/E crystals

After 21 days incubation, the crystals were screened using an in-house X-ray source Rigaku micromax^{TM} – 007HF with a Rigaku Saturn 944+ CCD detector. The crystals were protected by a cryoprotectant containing the crystallization solution with 20% glycerol. The crystals were damaged and diffracted poorly. I optimized the cryoprotectants and found that adding the cryoprotectant directly to the crystallization wells could protect the crystals.

1.1.6.6 Data collection of SeMet labelled LptD/E crystals

Multi-wavelength anomalous dispersion (MAD) data were collected at Diamond Light Source, UK at beam station I02, I03, I04 and I24. Before MAD data collection, fluorescence scanning was performed on the SeMet LptD/E crystals to determine the wavelengths to be used for MAD data collection at peak, inflection and remote.

The best of four wavelengths for MAD datasets at peak, inflection, high remote and low remote were collected at I24 using a Pilatus3 6 M detector under 100K. The peak dataset was collected over 3600 images with an exposure of 0.02 second per image at a wavelength 0.9784 Å with oscillation 0.1 degree per image and 50% transmission.

The inflection dataset was recorded at wavelength 0.9788 Å, while the high remote and the low remote datasets were collected at wavelength 0.9775 Å and 0.9818 Å, respectively, using the same strategy as the peak dataset collection. Chapter 2 mainly described that how structures have been solved.

1.2 Materials and methods for SBV NP

1.2.1 Buffers

Lysis buffer	20 mM Na-phosphate, pH 7.2, and 0.5 M NaCl, 10% glycerol, 10 mM imidazole), supplemented with cOmplete protease inhibitor tablets (Roche), 1 µM DNAse (Sigma), 1 µM Lysozyme (Fluka), and 0.1 mM phenylmethylsulphonyl fluoride (PMSF, Sigma-Aldrich)
Binding buffer	20 mM Na-phosphate, pH 7.2, and 0.5 M NaCl, 10% glycerol, 10 mM imidazole
Wash buffer	20 mM Na-phosphate, pH 7.2, and 0.5 M NaCl, 10% glycerol and 30 mM imidazole
Elution buffer	20 mM Na-phosphate, pH 7.2, and 0.5 M NaCl, 10% glycerol and 500 mM imidazole
Gel filtration buffer	20 mM Tris-HCl, pH 7.5, 300 mM NaCl and 10 % glycerol (v/v)
Denaturation buffer (Line A)	20 mM Na phosphate pH 7.2, 1 M NaCl 10% glycerol and 8M urea
Refolding buffer (Line B)	20 mM Na phosphate pH 7.2, 1 M NaCl and10% glycerol

1.2.2 Generation of SBV NP expression plasmid

The gene encoding (Strain Na2) SBV NP was amplified and cloned into a modified pDEST14 vector (Invitrogen) with an N-terminal hexahistidine (6-His) tag and a tobacco etch virus (TEV) protease cleavage site (for removal of the 6-His tag) upstream of the NP gene. This plasmid p14TevSBV NP, was kindly provided by our collaborator Dr. Ping Li (University of Glasgow).

1.2.3 Transformation

The p14TevSBV NP plasmid was used to transform into expression strain of *E. coli*, specifically Rosetta cells (Novagen). 100 ng of plasmids DNA was added to 50 µl of competent Rosetta cells, which is a favoured expression strain for viral proteins as the cells contain a pRARE plasmid encoding several rare tRNAs. All transformation was as described in Chapter 1.3.1.

1.2.4 SBV NP over-expression

LB (500 ml) was inoculated with the glycerol stock of Rosetta cells harbouring the NP plasmid with antibiotics ampicillin (50 µg ml^{-1}) and chloramphenicol (34 µg ml^{-1}). The bacterial cultures were incubated at 37°C and 200 RPM until the optical density (OD) at 600nm reached mid-log phase (OD_{600} = 0.6-0.8). Protein expression was induced by addition of IPTG to a final concentration of 0.1 mM and shanking at 200 RPM for 16 hours at 20°C. The cells were harvested by centrifugation at 6,500 × g for 15 minutes at 4°C.

(L)- selenomethionine labelled SBV NP was expressed in M9 medium. All expression was the same as described in Chapter 1.1.4.

1.2.5 Protein purification

Cell pellets were re-suspended in 150 ml of lysis buffer (20mM Na-phosphate pH 7.2, 0.5 M NaCl, 10% glycerol and 10 mM imidazole), supplemented with three EDTA-free protease inhibitor tablets (Roche), 1 µM DNAse (Sigma), 1 µM Lysozyme (Fluka) and 0.1 mM phenylmethylsulfonyl fluoride (PMSF)(Sigma). The cells were lysed by passing the cell mixture through a cell disrupter at 30kpsi twice (Constant Systems Ltd). Cell debris was removed by centrifugation at 19,000 × g for 40 minutes at 4°C (Sorvall F21S-8x50y rotor).

The supernatant was decanted and loaded onto a pre-equilibrated (lysis buffer) 5 ml Ni-NTA agarose (Qiagen) column (EconoPac, Biorad). Once the supernatant was passed through the beads twice, the resin was washed with 12 column volume (CV) of wash buffer (20 mM Na-phosphate pH 7.2, 0.5 M NaCl, 10% glycerol and 30 mM imidazole) to remove non-specific proteins. The recombinant NP protein was eluted with 2 CV elution buffer (20 mM Na-phosphate pH 7.2, 0.5 M NaCl, 10% glycerol, and 500 mM imidazole). The eluted protein was immediately desalted with gel filtration buffer using a desalting column (Hi-Prep^{TM} 26/10, GE Healthcare) and Äkta Express. The protein was pooled and the His-tag was removed by TEV protease cleavage at room temperature overnight (600 µl of 7 mg ml^{-1} TEV protease was added).

TEV protease, uncleaved protein and contaminants were removed by applying the samples through a Ni-NTA column. Briefly, the Ni-NTA column was pre-equilibrated with 10 CV of elution buffer, followed by 5 CV of H_2O and 5 CV of lysis buffer. The detagged NP flows through the column. Subsequently, the cleaved NP was monitored by SDS-polyacrylamide gel (NuPAGE 4-12% Bis-Tris) electrophoresis with a protein

molecular marker (Mark12, Invitrogen). The gel was stained with Coomassie blue R250. Gel bands corresponding to the NP molecular weight were cut and sent for identification using mass spectroscopy.

1.2.5.1 Purification of RNase A treated SBV NP

The RNase A treatment protein purification procedure was very similar to the method described above, except that the RNase A was added to the protein after TEV cleavage and the mixture was incubated for 3 hours at room temperature. Subsequently, the RNase A and other contaminating proteins were removed by size-exclusion chromatography using an Äkta Xpress.

1.2.5.2 Purification of SeMet labelled of RNase A treated SBV NP

Selenomethionine (SeMet) labelled protein purification was performed following the same protocol as the native SBV NP described above, and the protein was treated with the RNAse A prior to the gel filtration.

1.2.5.3 Purified SBV NP under denaturing and refolding conditions

The SBV NP was expressed using the same protocol as the native protein (chapter 1.2.4). The cells were suspended in cell lysis buffer and lysed by passing them through the cell disruptor twice. The cell debris was removed by centrifugation at 20,000g for 25 min. The supernatant was loaded onto a pre-equilibrated (lysis buffer) 5 ml HiTrap column (GE Healthcare). Then the column was washed with 6 CV wash buffer as described above. The SBV NP was denaturated on the HiTrap column by increasing the urea concentration from 0 to 8 M over a 10 CV gradient using the ÄKTA Xpress. Then, the SBV NP was refolded by decreasing the urea concentration from 8 to 0 M over a 10 CV gradient. This required two buffers containing 20 mM Na

phosphate pH 7.2, 1 M NaCl, 10% glycerol – one without urea (Line A) and one with 8 M urea (Line B) on a Äkta Xpress purifier.

To check whether RNA had been removed during denaturation, the 260/280 nm absorbance ratio of the flow-through was measured, which showed a high ratio of 2.10, clearly suggesting that some *E. coli* RNA had been removed. The refolded SBV NP was eluted using 500 mM imidazole, before immediately changing the buffer to TEV cleavage buffer by a desalting column (Hi-Prep^{TM} 26/10, GE Healthcare).

1.2.6 Crystallization and Data collection of SBV NP

1.2.6.1 Crystallization of native SBV NP

The crystallization conditions for the proteins from two peaks were screened separately using sitting drop in 96-well crystallization plates. Each crystallization drop was built up using 0.3 µl of protein and 0.3 µl of the crystallization solution with 70 µl of crystallization solution in the reservoir. The 96 well crystallization-screening kits, Index I & II, PEG/ION, Crystal Screen Cryo (Hampton Research), Wizard I & II (Emerald biosystems), JCSG+ (Molecular Dimensions Ltd.), StoPegs 1, 2, 3, 4 and Stochastic kits 16, 17 and 21 (JHN Lab) were used. All crystallization plates were incubated at room temperature.

1.2.6.2 Crystallization of RNase A treated SBV NP

Only the purest fractions were collected and concentrated to 7 mg ml^{-1} (Peak 1) and 12.2 mg ml^{-1} (Peak 2) respectively. The protein was flash frozen in liquid nitrogen and stored at -80°C. The crystallization trials were set using the Honeybee and used same crystallization screening as described above.

SBV NP crystals were obtained in several conditions within four days from the protein from peak 2. In order to find high quality crystals, all of the crystals were harvested and screened using the in-house x-ray source.

1.2.6.3 Crystallization of SeMet of RNase A treated SBV NP

The size exclusion chromatography pattern of the SeMet labelled SBV NP behaved similarly as the native SBV NP treated with RNase A. Therefore, I harvested the second peak of the pure protein, which was concentrated to 12.6 mg ml^{-1}. The

crystallization trials were setup using sitting-drop vapour diffusion, based on the original SBV NP crystallization.

1.2.6.4 Crystallization of denatured and refolded SBV NP

The two SBV NP samples pooled from different peaks were screened for crystallization conditions using Index I & II, PEG/ION and JCSG+ commercial screens. The Honeybee was used to setup 96 well plate sitting-drops, which were subsequently incubated at room temperature. Crystals were obtained after 7 days, the larger ones were flash frozen by supplementing the mother-liquor with 20% glycerol. The crystal diffractions were screened using in-house resource.

1.2.6.5 Determination of the quality of SBV NP crystals

To check whether the crystals were good enough for data collection, we decided to screen them using the in-house X-ray source. The protein crystals were mounted with litho-loops (Molecular Dimensions) before being flash frozen in liquid nitrogen using 1.4 M Sodium/potassium phosphate pH5.6 and 20% glycerol as cryoprotectant. The crystals were screened using a Rigaku micromax^{TM} – 007HF with a Rigaku Saturn 944+ CCD detector. However, all the crystals tested diffracted poorly.

1.2.6.6 Data collection of RNase A treated SBV NP crystals

Data was collected at beam station I24 of Diamond Light Source UK using a Pilatus3 6 M detector under 100K. 1200 images were collected with an exposure time of 0.5 second per image and oscillation angle of 0.15° per image, using a distance of crystal to detector of 498.3 mm and a wavelength of 0.9919 Å.

1.2.6.7 Data collection of RNase A treated SeMet of labelled SBV NP crystals

The crystals were cryoprotected by supplementing the crystallization solution with 20% glycerol, before being flash frozen in liquid nitrogen. The multi-wavelength anomalous dispersion (MAD) data were collected at Diamond Light Source, UK using ADSC Q315r detector at beam station I04. Before MAD data collection, fluorescence scanning was performed on seMet crystals to select the wavelengths to be used for MAD data collection at peak, inflection and remote.

All the three wavelengths datasets were collected from the same crystal at different positions. The Peak dataset was recorded over 1000 images with an exposure of 1 sec per image at the wavelength of 0.9797 Å. The oscillation angle was 1 degree per image and the crystal to detector distance was 421.7 mm. The inflection dataset was collected at the wavelength of 0.9799 Å with an oscillation angle of 1 degree per image, using 1 second exposures. A total of 360 images were recorded with a crystal to detector distance of 421.6 mm. The remote dataset was collected at the wavelength of 0.9218 Å with an oscillation angle of 1 degree per image and 1 second exposures. 360 images were collected with a crystal to detector distance of 449.9 mm.

1.2.6.8 Data collection of denatured and refolded SBV NP crystals

Crystals from the second peak were diffracted to a higher resolution than those from the first peak. We therefore decided to optimise the crystallization conditions of the protein pooled from the second peak. The best crystals were obtained with 0.075 M tris pH 8.5, 1.5 M ammonium sulphate, 25% glycerol after 12 days. They were frozen and tested in-house before collecting an entire dataset at Diamond beamline I24. 360

images were collected with 1-second exposure, with an oscillation angle of 0.5 degree per image at a wavelength of 0.9919 Å. The crystal to detector distance was 498.33 mm.

1.3 Materials and methods for SBV NP/RNA complex

All buffers were the same as described in Chapter 1.2.1.

1.3.1 SBV NP expression and purification

All the protein expression and purification were followed above method (section 1.2.4- 1.2.5) for the denatured and refolding SBV NP (section 1.2.5.3).

Protein concentration was determined from the absorbance at 280 nm using a Nanodrop (Thermo Scientific). The protein was flash frozen in liquid nitrogen and stored at -80°C.

1.3.2 SBV NP-RNA complex Crystallization and Data collection

1.3.2.1 Crystallization of NP-RNA complex

To reconstitute RNP, the RNA (poly U of 21-, 28-, 42- base-length; Eurogentec) was added into refolded SBV NP in a 1:1 molar ratio, and then the mixture was incubated on ice for 90 minutes. The protein-RNA complex was then screened for crystallization. All crystallization trails were same as described in Chapter 1.2.6.1.

1.3.2.2 Determination of structure of protein complexed with 21-, 28-, 42-nt RNA

The crystals of SBV NP in complex with 21-, 28-, 42-nt RNA were screened using in-house X-ray source, and these crystals were protected by supplementing 20% glycerol as a cyoprotectant before being flash frozen in liquid nitrogen. However, all crystals diffracted poorly, except the crystals of 42-nt RNA complex, which diffracted to high

resolution. Using in-house X-ray facility, 360 images were collected with an exposure time of 120 second per image, an oscillation angle of 0.5° per image and using a Rigaku micromax^{TM} – 007HF with a Rigaku Saturn 944+ CCD detector.

1.3.2.3 Data collection of 42-nt RNA-protein complex

Data was collected at beam station I24 of Diamond Light Source UK using a Pilatus3 6 M detector under 100K. 1000 images were collected with an exposure time of 2.0 second per image and oscillation angle of 0.5° per image, using a distance of crystal to detector of 498.3 mm and a wavelength of 0.9200 Å at 100 percent of transmission.

Chapter 2

The steps towards structure by X-ray crystallography

This chapter is mainly to describe the theories and softwares used in determining structures of LptD/E, SBV-NP and NP/RNA complex in my thesis following the workflow for solving the structure by X-ray crystallography (Figure 2.1).

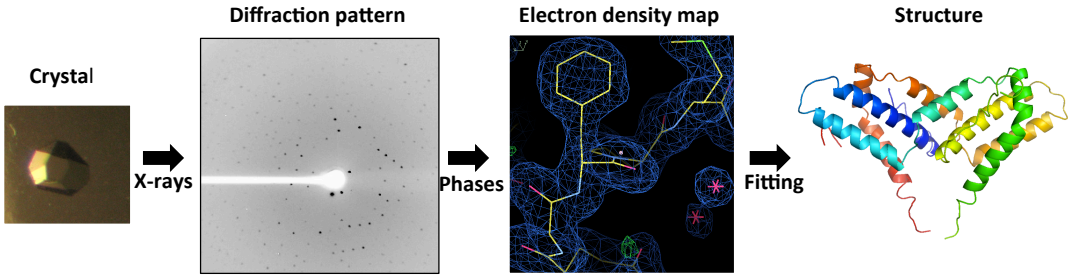


Figure 2.1. The step towards of structure by X-ray crystallography.

In brief, high quality of crystals are essential to determining the structure of a protein by X-ray crystallography, which measures the crystal diffraction directions and intensities by X-ray beams with 60 - 360 degree oscillations. Each diffraction image corresponds to crystal diffraction at a different oscillation angle. The diffraction data was indexed and integrated using iMosflim or HKL2000 or Xia2 or XDS. However, these data lacked the phase information, which can be obtained by molecular replacement or experiment phases (Battye et al. 2011; Kabsch 2010; Otwinowski et al. 1997; Kabsch 2010).

After solving the phase problems, the electron density map of the protein is generated by programs RESOLVE, or SHELX E (Sheldrick 2010; Terwilliger 2003). The molecular model was built according to the electron density map automatically or manually (Figure 2.1).

The strategy of collecting diffraction image is through oscillation of the crystal with a small angle by a starting position of rotation and recording all of reflections, then rotating crystal to a new starting point to re-record reflections. However, the new

oscillation range overlaps the previous one slightly, the reflections are recorded until all unique reflection has been collected (Battye et al. 2011).

2.1 Data Indexing

Each reflection is designated *h, k, l* to the position of an individual reflection in the reciprocal space of the diffraction pattern. The central reflection is used as the origin in reciprocal space and coordinates *hkl* = 000. Coordinates of other reflections are calculated from this origin and the indices *h*, *k*, *l* are integers.

Data processing can be indexed the reflections of the position *h*, *k*, *l* and the intensity I_{hkl} of each reflection. This data is used to calculate the dimensions of the unit cell and to determine the symmetry of the crystal of its space group.

2.1.1 iMosflm

iMOSFLM is extensively used to process diffraction images from a wide range of detectors, and produces an MTZ file of reflection indices with data intensities and standard deviations (Battye et al. 2011). The MTZ file is able to pass onto other programs of CCP4 program suite (POINTLESS, SORTMTZ, AIMLESS, CTRUNCATE) for further data reduction (Battye et al. 2011). The softwares HKL2000, Xia2, XDS are also used to process data (Kabsch 2010; Otwinowski et al. 1997; Kabsch 2010). Diffraction images are added to the iMosflm, the first image of sector is displayed as close as possible to a 90° rotation away in a new display window to show diffraction spots. The program is designed into a series of steps (Images, Indexing, Strategy, Cell refinement, Integration) and can be selected by clicking on the appropriate icons for data process. The spots finding parameter is set

by default to be between circles of radii between 5 and 95% of the radius of the inscribed circle centred on the direct beam position and the threshold value is default set to part of spot with 5 *sigma* above the background for a pixel and a various of rejection criteria is applied to distinguish the true Bragg spots in the diffraction image from noise.

Once all the images are integrated using iMosflm, the program POINTLESS (Winn et al., 2011) can be used to determine the true Laue symmetry and to determine the space group.

For example the pointless results the space group of $P2_1$ for SBV NP and the unit cell is defined by three lengths a = 76.21,b = 85.62, c = 77.03 Å and three angles $\alpha = \gamma = 90$ and $\beta = 101.9°$. The unit cell is the smallest unit that contains all of the structural and symmetry information and it can simply be stacked onto the next unit cell by simple translation to reconstitute the whole crystal.

2.2 Run Scaling

Scala merges multiple observations of reflection and produces a file, which contains averaged intensities for each reflection (Evans 2006). The Scala in CCP4 suite reads a sorted MTZ file of unmerged intensities, usually produce from iMosflm. The file also could be generated from other integration programs, such as Scalapack for XDS, and dtrek2scala for d*trek.

Scala calculates amplitude |F| from intensity, uses Ctruncate to output Wilson plot, puts all datasets into the same file, and generates FreeR set.

It is good to use R-merger during merging datasets from multiple reflection frames or

multiple crystals, which measure of same reflection with different measurement in being different frames of data among multiple measurements (Rhodes 2006). Rmerge is calculated as follows:

$$R_{merge} = \frac{\sum hkl \sum_{j=1}^{N} || Fhkl | - | Fhkl (j) ||}{\sum hkl \sum_{j=1}^{N} | Fhkl (j) |}$$

Where $| F_{hkl} |$ is the final value of the structure factor amplitude, Σ_j is scattering factor of atom j, *hkl* is index of diffraction.

The sum of diffraction reflects contributions of all atoms in the unit cell and all these individual atoms sum of the structure factor F_{hkl}. The structure factor sums all the reflection *hkl* for individual atoms, which is a Fourier sum, and has been treated as a sphere of electron density, which contributes each element of electron density of a volume element at centre position (*x, y, z*) and the average values of ρ (*x, y, z*) at this region (Rhodes 2006). The structure factor describe as follows:

$$F_{hkl} = \int_V \rho\,(x, y, z)\, e^{2\pi i(hx+ky+lz)} dV$$

Where the integral is carry out over volume element *dV,* the unit cell for the integral over all volume of *x, y, z*. Each volume element distribution to the structure factor F*hkl*, with a phase determined by its position (*x, y, z*) with each volume element.

The structure factor equation is the Fourier transform, and it is its inverse. The electron density ρ (*x, y, z*) is turned in structure factor with an inverse Fourier transform as follows:

$$\rho\,(x, y, z) = 1/v\ \Sigma_h\Sigma_k\Sigma_l F_{hkl} e^{-2\pi i\,(hx+ky+lz)}$$

Where V is the volume of the unit cell, $\Sigma_h\Sigma_k\Sigma_l$ the sum of diffraction with index diffraction (h, k, l), *i,* type of atom. The phase is unknown form the structure factor, and it is required to calculate the electron density.

2.3 Phase determination

The structure factor possesses amplitude (F), frequency (h) and phase (α). The amplitude of F_{hkl} can be obtained by taking the square root of all measure reflection intensity I_{hkl}. Three frequencies (*h, k, l*) of three-dimension wave function produce the reflection that can be calculated, because the frequency of a structure factor is equal to $1/d_{hkl}$ that wavelength is the same as the space of planes producing the reflection. However, the phase of F_{hkl} is lost during data collection and is unable to detect from a single measurement of the reflection intensity(Rhodes 2006).

2.3.1 Molecular Replacement

Molecular replacement (MR) is a method of solving the phase problem (McCoy et al. 2007). It requires a homologue protein structure with protein sequence identity above 30%. The homology's structure can be used to calculate the initial phase of structure factor to obtain an electron density map as follows:

$$\rho\,(x, y, z) = 1/v\ \Sigma_h\Sigma_k\Sigma_l F_{hkl} e^{(2\pi i\,(hx+ky+lz)\,+\,i\alpha\,(hkl))}$$

Where α (*hkl*) is a phase with specific reflection (*hkl*) in the reciprocal space of the each diffraction pattern, which can obtain complete structure factor to real space electron density.

A Patterson map is calculated (F_{hkl}) from all the measured reflection intensity. By comparing to a Patterson map using the homologue structure coordinates previously solved in different orientations, the high correlation coefficients is given high score for two structures in similar orientation (Rhodes, 2006; Rius, 2011). In Patterson map all phases are set to zero, and the electron map contains a peak corresponding to distance vectors between each atoms. By rotating and then translating the Patterson

map closely correlated to each other in the correct orientation and position within the asymmetry unit cell, and then the correct phases would were obtained. The programs called Molrep and Phase in CCP4 can be used for MR (McCoy et al. 2007; Vagin et al. 2010).

Once obtained the phases with correctly orientation and translation, it is able to produce accurate electron density that can be used to build unknown protein structure with atom model.

2.3.2 Phase obtaining from heavy atom

For many proteins, there are no homologous structures available from protein data bank. To determinate these protein crystal structures, the phases are obtained from heavy atom binding protein crystals. There are two ways to obtain heavy atom derived protein crystals. One way is to incorporate the heavy atom into the protein from heavy atom derived amino acids, and the most popular amino acid is selenomethioine. The other way is to soak the heavy atoms to native protein crystals with heavy ions or ionic complex of Hg, Samarium and Pt in crystallization solution (Taylor, 2003). One or more heavy atoms bind to protein for phase determination and the heavy atom must not change crystal packing or conformation of protein.

A powerful method is the multi-wavelength anomalous diffraction (MAD) (Hendrickson et al. 1990). The multi-wavelength radiation dataset is collected using same crystal or different crystals that contain sufficient phasing information at different wavelengths (Normally, data was collected at peak, inflection, remote wavelengths) from heavy atom derivatives. The information can be used to accurately locate the positions of heavy atoms, and give the phases for the protein structure determination.

The Fourier sum called Patterson function P (u, v, w) is the most powerful tool in determining the heavy-atom coordinates, which is calculating electron density ρ (x, y, z) from structure factor (Rhodes, 2006; Rius, 2011). The P (u, v, w) coordinates sites in Patterson map are used same way as coordinates (x, y, z) in an electron density map. The Patterson function does not contain the phasing information and the amplitude of each term is the square of one structure factor as follows:

$$P(u, v, w) = 1/v \, \Sigma_h \Sigma_k \Sigma_l \, F^2_{hkl} e^{-2\pi i (hu+kv+lv)}$$

Although the Patterson function is without phases, the Patterson map P (u, v, w) can be calculated from location of heavy atoms in the unit cell from high density (peaks). The phases obtained by calculating structure factor form inverted hand or original hand and incorrect hand will not provide an interpretable map. The program of SHELX C/D/E (Grüne 2012; Sheldrick 2007) can extract the phase information from MAD and SHELXC analysis of date set resolution and is also prepared three files for SHELXD, which determine the heavy atoms locations and find the correlation coefficient at the signed anomalous differences for wavelength with highest anomalous signal, if the correlation coefficient reaches 40-50%, it indicates a reliable solution for MAD, and around 30% may be correct for single wavelength anomalous dispersion (SAD). The SHELXE is mainly for electron density modification. SAD is becoming a quick method to determine the protein structure, which is only required to collect data at peak wavelength.

2.4 Structure model building

The Phase α (*hkl*) is obtained from heavy atoms and the electron density map is obtained from complete structure factor of amplitude. An atom model of the structure is built into the electron density map.

The Buccaneer software is an automatic program for building atomic model from electron density map, even at low resolution around 3.5 angstrom, which connects the alpha-carbon positions using a density likelihood function and scores possible positions and orientation in the electron density map (Cowtan 2006). A list of amino acid oriented group is searched, which provide additional directional information to help the process of assembling the sequences into the protein subunit chains.

Once the initial model is built, the model of protein is needed to tidy up with Coot (Emsley et al. 2004). It is a graphic program for model manually building by moving the atoms and fragment, changing amino acid residues and contains many tools for electron density fitting to improve the model depend on |Fo| – |Fc| and 2|Fo| – |Fc| maps. The |Fo| - |Fc| map always shows the atoms within negative contours for problem areas and point out to correct location for these atoms. The 2|Fo| – |Fc| map is the main map for atoms building within positive contours. After manual adjusts of the model, which may result in unrealistic bond length and angles, part of model can be regularized with automatic correction of bond lengths and angles with minimal movement of residues.

2.5 Structure refinement

To improve the protein structure model and interpret the electron density map correctly, the iterative process of structure refinement is an important step for adjustment of atom coordinate to refine the model in good agreement with original diffraction data. After that the electron density map will be improved, which provides clearer and more detailed information to trace the continual density and side chains (Rhodes, 2006).

The least squares is calculated from the structure refinement according to the current model $|F_{cal}|$, which can be calculated from the current model of the structure factor amplitudes and observed amplitudes $|F_{obs}|$ that forms the original diffraction intensity (Rhodes, 2006). In the least squares, atom positions are selected in comparing minimize the squares difference between $|F_{cal}|$ and $|F_{obs}|$. The difference between observed amplitudes $|F_{obs}|$ and current model measured amplitudes $|F_{cal}|$ for reflection *hkl* is $(|F_{obs}| - |F_{cal}|)_{hkl}$, and the minimize function Φ is as follows:

$$\Phi = \Sigma_{hkl} W_{hkl} (|F_{obs}| - |F_{cal}|)^2_{hkl}$$

Where the function Φ is the sum of squares of differences between the observe amplitude and the calculated current model amplitude. The weighted W*hkl* is depended on the reliability of the measured intensity.

During the manual or automatic refinement, the constraints and restraints are used, which help to build the current model in agreement with the original intensity. The temperature factor (B-factor) measures the atom oscillations around positions in the models. In initial refinement, all temperature factors are assigned a starting value, however, the overall values is not constrained.

The R-factor is a measure of the agreement within structure refinement by calculating that the current model structure factor $|F_{cal}|$ and the observed structure factor amplitude $|F_{obs}|$ from original diffraction intensity (Brünger et al. 1987). It is defined by the following:

$$R_{factor} = \frac{\Sigma \, ||Fobs| - |Fcal||}{\Sigma \, |Fobs|}$$

Where $|F_{obs}|$ is structure factor and sum over all measured reflections and current model structure factor calculated $|F_{cal}|$.

The R-free is used as an important quality control, which measures how well the current model quality from entire data set and improvement in refinement (Brünger 1992). It is calculated with a randomly select 5- 10% data set, which is not used during refinement. The program of Refmac5 within CCp4 is used for refinement, which can perform rigid body, restrained and TLS refinement against data. Refmac5 carried out different likelihood functions during structure refinement..

2.6 Structure validation

Validation methods are used to check the protein structure's quality, and can give the suggestions about what the problems the structure has.

The peptide backbone conformational angles are Φ and Ψ. The Φ is the torsional angle alone the N - C*a* bond by the atoms C – N – C*a* – C and Ψ is the torsional angle alone the C*a* – C bond by the atoms N – C*a* – C – N (Rhodes, 2006).

The final model shows each amino acid with the pair of angles Φ and Ψ being restricted by steric repulsion. The allowed pairs of values are described on a Ramachandran plot (Ho et al. 2005). The conformational angles Φ and Ψ on either side represents the alpha carbon of one residue on the point (Φ, Ψ) in the diagram and

pink polygons represents enclose backbone conformation angle no steric repulsion, while yellow polygons represent only modest repulsion. The letters of α and β on the location represents conformational angles of residues in the structure with α helix and β sheet (Lovell et al. 2003). Ramachandran plots are very helpful in spotting conformationally unrealistic regions of the structure and it will show slight differences in the shapes of allowed regions from different sources (Ho et al. 2005; Lovell et al. 2003).

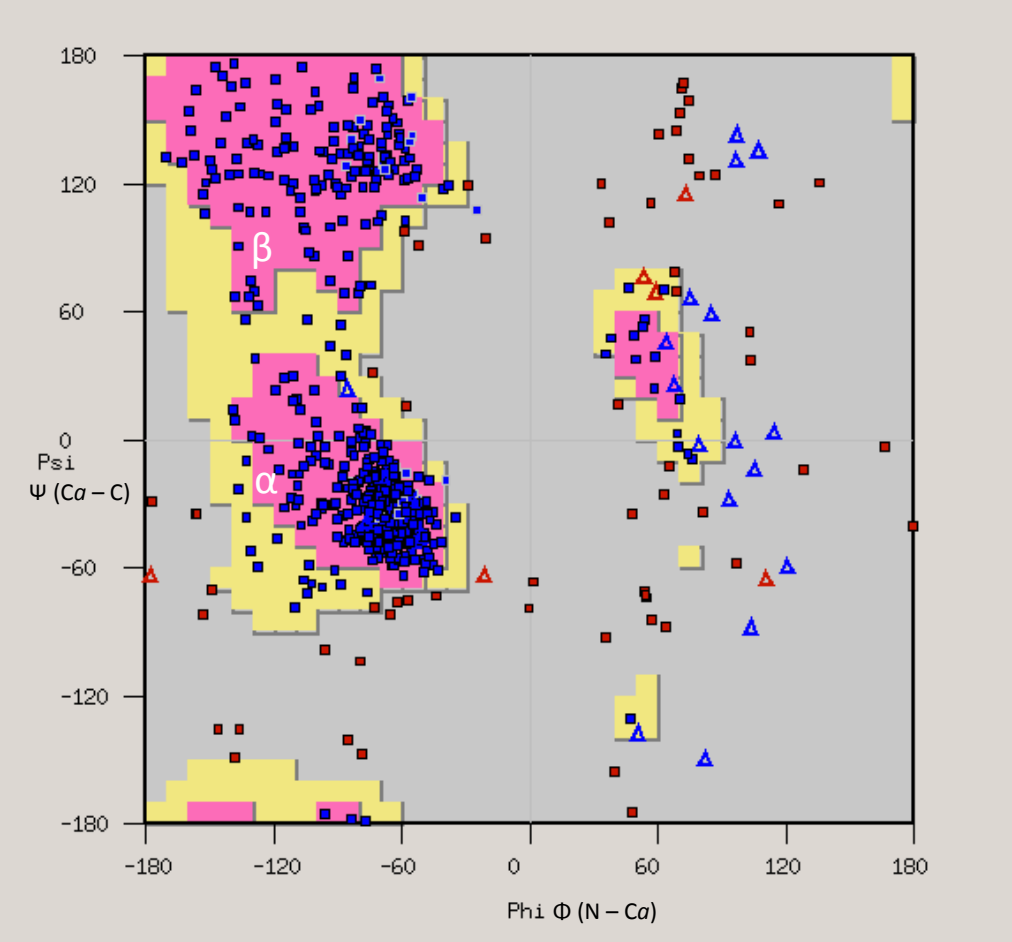


Figure 2.2. Ramachandran plot of structure of the denatured and refolded SBV NP. The most of amino acids located in favourite area.

Molprobity is a structure-validation web server in checking model quality. It provides detailed by optimized hydrogen placement and all atom contact analysis of steric problem within the structure and also update dihedral angle diagnostics (Chen et al. 2010; Davis et al. 2007). This program can rank the protein structure models against similar resolution models available in the protein data bank, which gives the ideas how much effort is required to obtain the final model.

Chapter 3

Crystal structure of lipopolysaccharide transport membrane protein complex LptD/E

3.1 Introduction

One hundred and twenty five years ago, Christian Gram developed a staining method, which could differentiate between two major classes of bacteria, the Gram negatives and Gram positives (Gram 1884). Microbiologists use this classification widely, which is based on the composition and structure of the cellular envelope of bacteria (Beveridge et al. 1983; Davies et al. 1983).

During the 1940s-1970s, much progress was made in determining the composition, structure and function of the different types of the bacterial envelopes. Many scientists have focused on the glycolipid of the outer membrane (OM) of Gram-negative bacteria since 1920s. The OM of Gram-negative bacteria is an asymmetric lipid bilayer in which the inner leaflet is composed of phospholipid and the outer leaflet is composed of the glycolipid, lipopolysaccharide (LPS) (Raetz et al. 2002; Ruiz et al. 2009; Nikaido 2003).

3.1.1 Gram-negative and Gram-positive bacteria staining

Gram-negative bacteria are referred to as Gram-negative because they do not retain the crystal violet stain, used in the Gram staining reaction, resulting in red or pink bacteria (Gram 1884). The Gram-negative bacteria have thin peptidoglycan layer of cell wall, which is sandwiched between an outer membrane and plasma membrane (Figure 3.1 A). In Gram staining the outer lipid layer of Gram-negative bacteria is dissolved by ethanol, thus increasing the permeability of the cell wall, and the crystal violet is easily decolorized by ethanol (Gram-negative bacteria cell wall peptidoglycan content is much lower compared with its lipid content). Subsequently a counterstain with the dye safranin will turn Gram-negative bacteria red or pink

(Figure 3.1 B) (Gram 1884). This test provides a way to distinguish between two different bacterial cell wall structures. Due to Gram-positive bacteria having a thicker peptidoglycan layer in bacteria cell wall (Figure 3.1 C), the ethanol in the stain is unable to permeate through to decolorize the crystal violet dye. As a result the Gram-positive bacteria retain the crystal violet stain. Although Gram-positive bacteria will also be counterstained, it will not be observed. (Figure 3.1 D) (Gram 1884).

Both Gram-negative and Gram-positive bacteria possess a cell wall, which is composed of peptidoglycans, surrounds the plasma membrane and acts to protect the cell against increased water pressure. Peptidoglycan is a polymer composed of N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM) and short amino acid chains. Alternative NAG and NAM molecules form carbohydrate backbones that are cross-linked by polypeptides (Vollmer et al. 2008). Although the structure of polypeptides will vary in different polypeptide cross bridges, the tetrapeptide always composes four amino acids attached to NAMs (Figure 3.1 E).

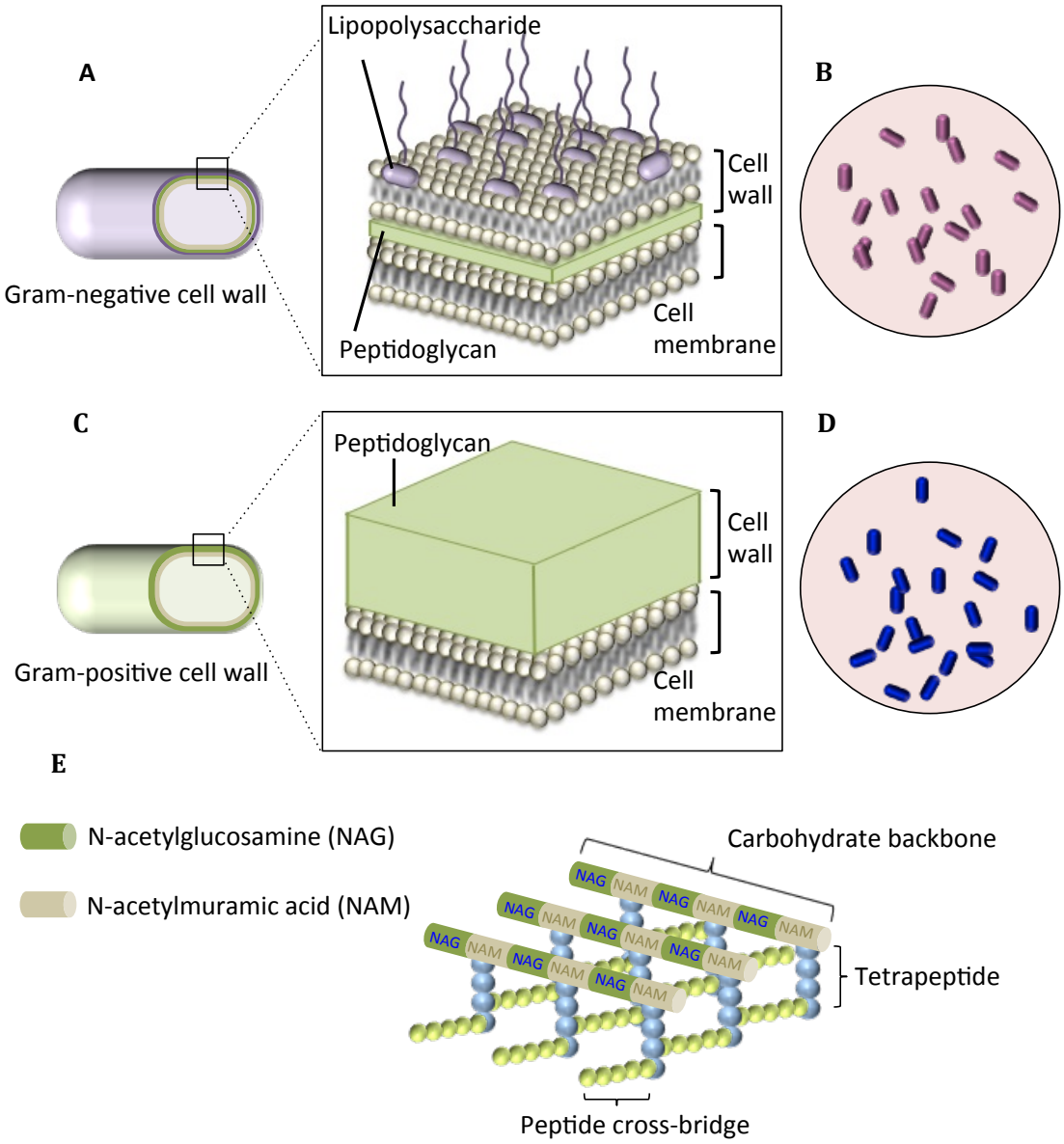


Figure 3.1 Gram-negative and Gram-positive cell wall and staining. A. Gram-negative bacteria cell wall. B. Gram-negative bacteria that stained in pink colour. C. Gram-positive bacteria cell wall. D. Gram-positive bacteria that stained in darker violet colour. E. The structure of peptidoglycan in cell wall.

3.1.2 Structure of Gram-negative bacterial cell envelope

Gram-negative bacteria, such as *Pseudomonas* and *Escherichia coli*, possess cell walls consisting of three layers, namely, the OM, the peptidoglycan, and the plasma membrane or IM (Figure 3.1 A). The peptidoglycan is much thinner than those in Gram-positive bacteria, but remains strong and elastic to protect the bacteria against extreme environmental conditions. Unlike the IM, the OM is an asymmetric lipid bilayer, which consists of phospholipid in the inner leaflet and the outer leaflet is composed of LPS (Raetz et al. 2002; Ruiz et al. 2009). The space between the outer membrane and the inner membrane is termed periplasmic space.

3.1.2.1 The outer membrane of Gram-negative bacteria

The OM functions as a barrier protecting bacteria from antibiotics and prevents diffusion of small hydrophobic molecules into the bacterial cell (Nikaido 2003). Like other biological membranes, the OM also has a lipid bilayer structure, but not a phospholipid bilayer. The distinct feature of the OM in Gram-negative bacteria is that it possesses an asymmetry arrangement with a layer of glycolipid, primarily LPS, on the outer leaflet and phospholipids in inner leaflet (Raetz et al. 2002; Ruiz et al. 2009). The protein embedded in the OM can be normally divided into two classes: lipoproteins and β-barrel proteins. Lipoproteins contain lipid moieties that are attached to amino terminal cysteine residues. These lipid moieties are thought to embed lipoproteins (rather than transmembrane) in the inner leaflet of OM (Sankaran et al. 1994). In contrast, nearly all OM transmembrane proteins are β-sheet proteins, whose conformations are wrapped into barrels (with a few exceptions, such as Wza, which uses helix domain to cross outer membrane) (Dong et al. 2006). Several outer membrane proteins (OMPs), such as the porins, OmpF and OmpC, function to allow the passive diffusion of small molecules across the OM (Silhavy et al. 2010). Crystal

structures show that porin proteins are transmembrane β barrel proteins and all of them exist as trimers (Cowan et al. 1992). For examples, LamB porins have 18 transmembrane β strands (Schirmer et al. 1995) and PhoE has16 transmembrane β strands (Cowan et al. 1992). Both of these protein structures exist as trimers and are involved in the diffusion of specific small molecules (maltose or maltodextrins) and anions (phosphate) respectively, across the OM.

3.1.2.2 The periplasm

The cellular compartment between the OM and IM of Gram-negative bacteria is called the periplasm or periplasmic space. The periplasm contains many proteins that function in nutrient acquisition, which is responsible for the transport of nutrition materials into the cell (Silhavy et al. 2010).

3.1.2.3 The inner membrane

The IM is composed of a phospholipid bilayer. The main phospholipids are phosphatidylethanolamine, phosphatidyl glycerol and lesser amounts of cardiolipin in *E. coli*. The phospholipids are suggested to maintain the permeability barrier of membrane and serve as supporting matrix for membrane proteins (Silhavy et al. 2010). The IM cytoplasmic surface is the site of biosynthesis of all membrane lipids and proteins (Raetz et al. 1990). Most of the membrane proteins of the IM function in energy production, lipid biosynthesis and protein secretion (Silhavy et al. 2010).

Inner membrane proteins (IMPs) are synthesized on ribosomes where the signal recognition particle (SRP) binds to newly synthesized peptide on the N-terminal signal sequence, which then delivers them to the Sec machinery. IMPs are directly inserted into the IM (Figure 1.2) (Von Heijne 1990; Hagan et al. 2011). Trigger factor (TF) is a ribosome-associate molecular chaperone in bacteria, which assists in newly

synthesized protein folding and prevents premature protein secretion through the SRP pathway (Hoffmann et al. 2010).

OMPs are also ribosome-synthesized in the cytoplasm but these bind to the cytoplasmic chaperone (SecB) before being transferred to the Sec machinery (Figure 1.2). After they pass through the Sec channel, their signal sequences are removed by the signal peptidase (Hagan et al. 2011). OMPs in unfolded form are thought to be protected by periplasmic chaperones, primarily SurA, and then transported and inserted into the OM by the β-barrel assembly machinery Bam ABCDE (Figure 3.2 B). Periplasmic chaperones, SurA, Skp and DegP are involved in protection and transport of the unfolded OMPs (Kim et al. 2007; Vertommen et al. 2009; Hagan et al. 2011) .

Lipoproteins with N-terminal signal sequence are also synthesized on ribosomes and transported by Sec machinery. However, the signal sequence is cleaved by signal peptidase II at the periplasmic face of the IM (Hagan et al. 2011) and lipid motifs are added (Figure 3.2 A). Lipoproteins interact with ABC transporter Lol CDE (localization of lipoprotein) at the outer leaflet of the IM and are then delivered them to the periplasmic chaperon LolA. LolA transports them to lipoprotein LolB on the OM for assembly (Okuda et al. 2011; Tokuda 2009; Hagan et al. 2011) (Figure 3.2 A).

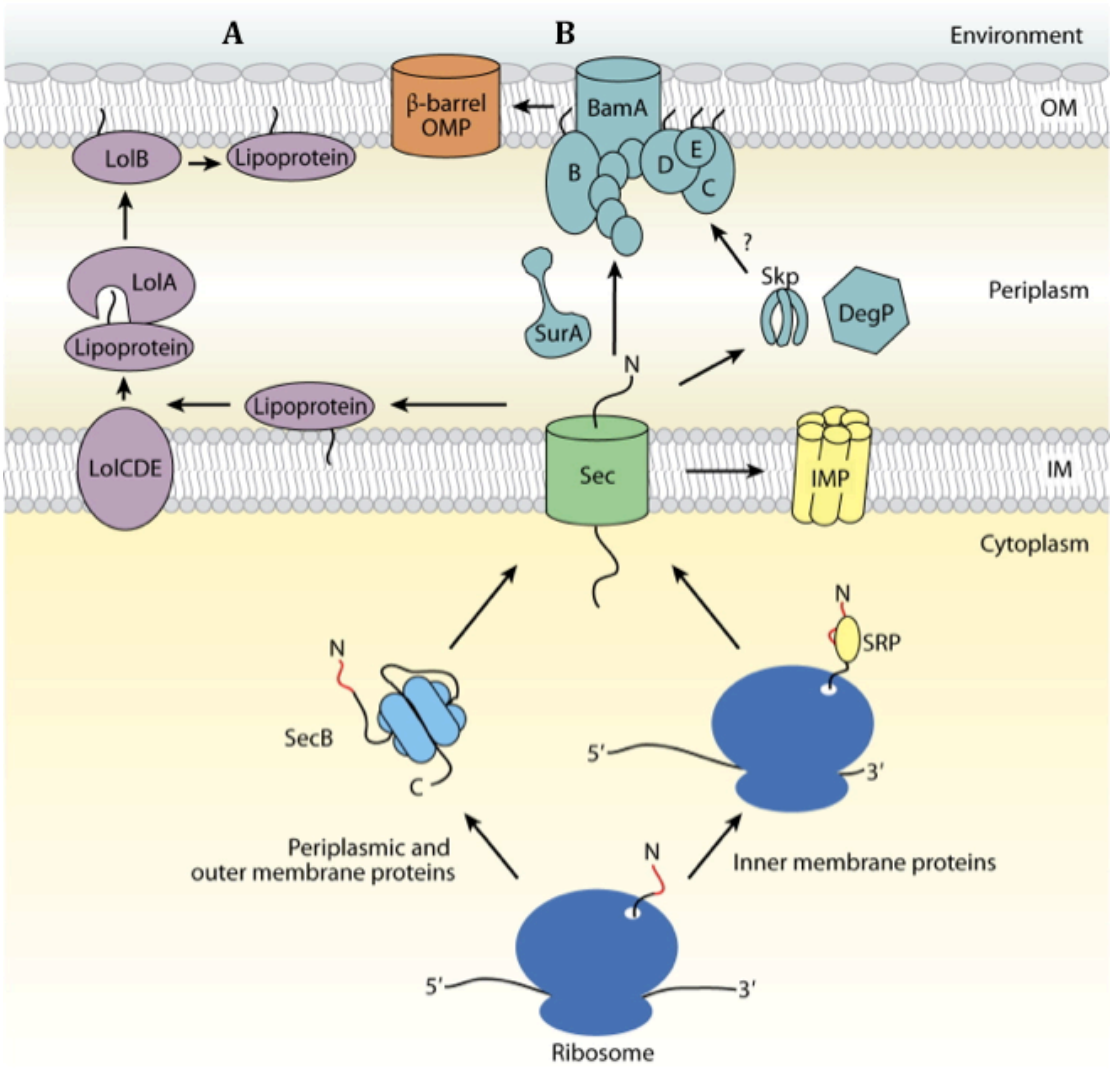


Figure 3.2 Cell envelope proteins synthesize pathway. All proteins (OM, Periplasm, IM) are synthesized on ribosomes in the cytoplasm and delivered to the Sec machinery. IMPs are co-translated with SRP and are inserted into the IM. Periplasm proteins (soluble) pass the Sec channel and are folded in periplasm compartment. A, Lipoproteins are delivered by Lol machines from the IM to the OM. B, The OM β-barrel proteins are inserted to the OM from periplasm by BamABCDE machinery (Figure adapted from Hagan et al. 2011) .

3.1.3 Structure of LPS and function

LPS typically consists of three domains: lipid A, core oligosaccharide and O-antigen (Figure 3.3). The core oligosaccharide is covalently linked to lipid A, and is divided into two parts, inner core and outer core. The inner core is composed of heptose and KDO (3-deoxy-D-manno-oct-2-ulosonic acid) and the outer core consists of a complex polymer of oligosaccharide which determines the specificity of the LPS. KDO is connected to lipid A without any additional saccharides called as Re-LPS, while LPS containing the lipid A and core oligosaccharide without O-antigen oligosaccharide that is called as Ra-LPS (Raetz et al. 2002). LPS plays an essential role in protecting the bacteria from harsh environments and toxic compounds including detergent and antibiotics and in limiting entry of hydrophobic molecules. It is a major endotoxin of Gram negative bacteria which elicits immune responses via toll-like receptor 4 in the host (Ruiz et al. 2009; Bos et al. 2004). LPS's cell surface localization and physicochemical properties make it an important OM barrier and it is difficult to develop antibiotics against these organisms because of the presence of LPS (Delcour 2009).

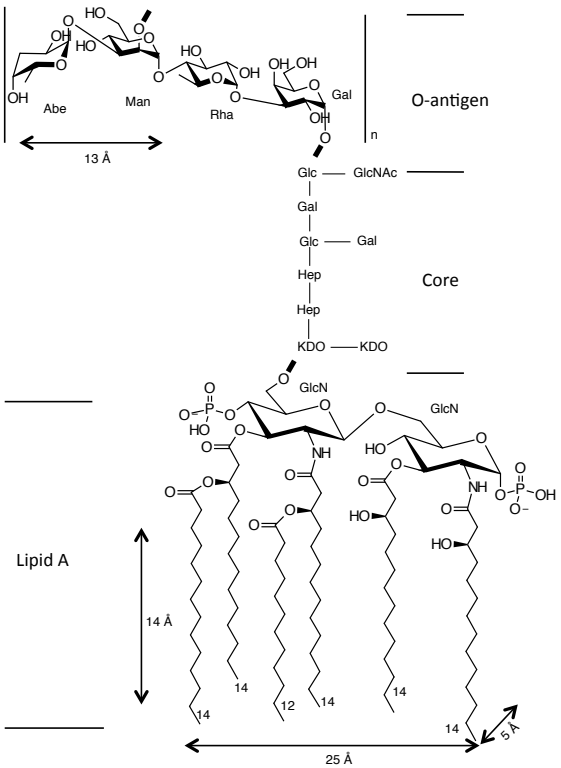


Figure 3.3 Schematic structure of LPS of *salmonella enterica typhimurium LT2* . LPS molecules consist of three domains: the core, O-antigen and lipid A. KDO, 3-deoxy-D-manno-oct-2-ulosonic acid; Hep, Heptose; Glc, D-glucose; Gal, D-galactose; GlcNac, N-acetylglucosamine. The lipid A diameter of high is about 14 Å, long 25 Å and width 5 Å. The O-antigen diameter of linear polymer is around 13 Å.

3.1.4 LPS synthesis and assembly pathway

The lipid A-core segment of LPS is synthesized at the cytoplasmic side of IM and then is flipped over the IM by the ATP-binding cassette (ABC) transporter, MsbA (Raetz et al. 2002). The O-antigen units are independently synthesized at an alternate site in the cytoplasm and then translocated across the IM by the Wzx flippase to the periplasm. The O-antigen is polymerized by Wzy with Wzz mediating the lengthening of O-antigen. The O-antigen unit and lipid A-core are then ligated by WaaL ligase at the periplasmic face of the IM to form mature LPS (Raetz et al. 2002; Ruiz et al. 2009; Whitfield, 2010).

There are 16 predicted ORFs within the gene cluster (Figure 1.4 A), and all these genes in the same transcriptional direction form *galf* to *hisI* gene (Shao et al. 2003). It has been well established that a hexose-1-phosphate is transferred from nucleotide diphospho-linked sugar (typically NDP-glycoses) to an undecaprenol diphosphate (Und-PP) acceptor to form an Und-PP linked tetrasaccharide repeat unit. This step is initiated by a polyisoprenyl-phosphate hexose-1-phosphate transferase (PHPT) or polyisoprenyl-phosphate *N*-acetylhexosamine-1-phosphate transferase (PNPT) and completed by glycosyltransferase (Whitfield, 2010). The lipid-linked repeat unit is then exported across the cytoplasmic membrane through a process involving flippase, Wzx flippase to the periplasm. A polymerization reaction is catalyzed by polymerase Wzy (blue) which involves in putting the O-antigen units together to form the polymers. Finally, a polysaccharide copolymerase protein Wzz (purple) enhances the processivity of the polymerization reaction, ensuring that the majority of the glycan products fall within a relatively narrow size range or modal chain-length (Figure 3.4 B) (Raetz et al. 2002; Ruiz et al. 2009; Whitfield, 2010).

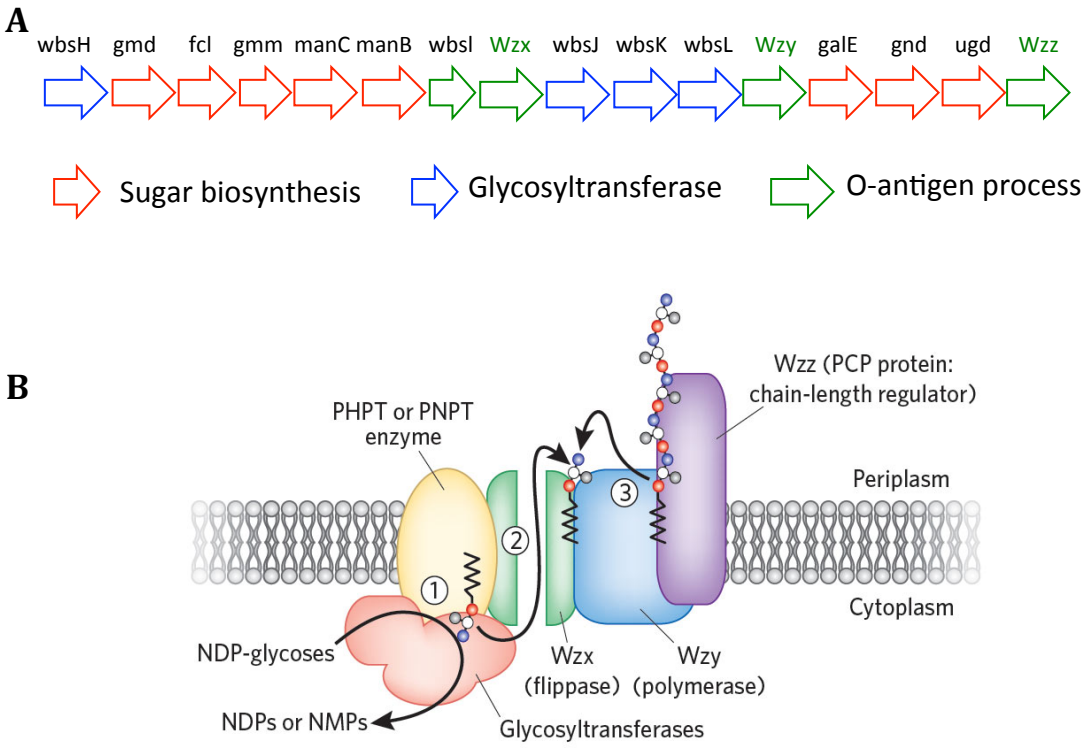


Figure 3.4 The O-antigen gene cluster of *E. coli* O128 and assembly in Wzx (RfbX), Wzy (Rfc) and Wzz (Rol, Cld) process. A. The O-antigen gene cluster of *E. coli* O128 biosynthesis. B. (1), A hexose-1-phosphate is transferred from nucleotide diphospho-linked sugar (typically NDP-glycoses) to an undecaprenol diphosphate (Und-PP) acceptor to form an Und-PP linked tetrasaccharide repeat unit. This step is started from a polyisoprenyl-phosphate hexose-1-phosphate transferase (PHPT) or polyisoprenyl-phosphate *N*-acetylhexosamine-1-phosphate transferase (PNPT) (yellow) and finished by glycosyltransferase (red). (2), The lipid-linked repeat unit is then exported across the cytoplasmic membrane through flippase, Wzx (green). (3), A polymerization reaction is catalyzed by polymerase Wzy (blue) which involves the transfer of growing polymer from its Und-PP carrier to the incoming Und-PP-repeat unit. Finally, a polysaccharide copolymerase protein Wzz (purple) controls the length of the O-antigen (Figure adapted from Whitfield, 2010).

Seven proteins (LptA-G) are required for LPS transport from the IM to the OM of the cell surface. LptB, LptC, LptF and LptG form an ABC transporter, and these proteins are essential for LPS transport (Villa et al. 2013; Sperandeo et al. 2008; Freinkman et al. 2012). LptA is a periplasmic protein that mediates LPS transport across the periplasm to its final destination, the OM (Sperandeo et al. 2011). LPS is delivered to the OM from the IM through a bridge formed by LptC, LptA and the N-terminal domain of LptD. The OM proteins LptD and LptE form a stable complex (Freinkman

et al. 2011; Wu et al. 2006; Chimalakonda et al. 2011; Malojčić et al. 2014), which is responsible for LPS assembly at the cell surface (Figure 3.4). Disulphide bonds are formed between the N and C-terminal domain of LptD, which have been shown to be critical for LPS assembly (Ruiz et al. 2010).

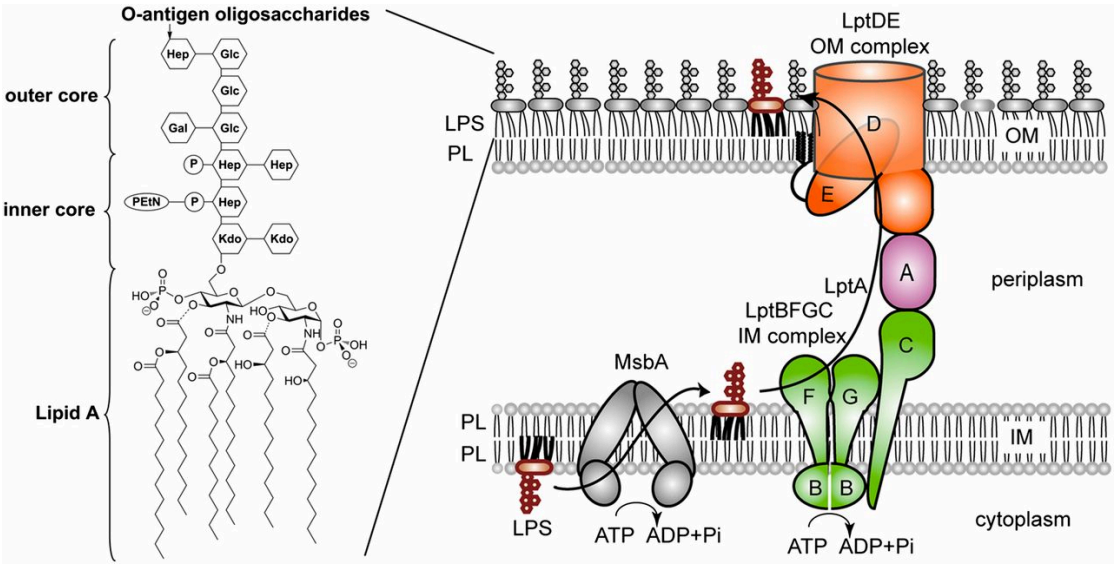


Figure 3.5 LPS assembly pathway. LPS structure shown on the left hand side. Once a mature LPS is synthesized, it is extracted from the IM by LptBFGC, and passed to periplasmic chaperone LptA. LPS is delivered to the LptD/E complex, which inserts LPS into the outer membrane (Figure adapted from Malojčić et al. 2014).

3.1.5 Lipid A and the oligosaccharide core of LPS are flipped across the IM

Lipid A and the oligosaccharide core of LPS are synthesized in the cytoplasm, and are transported across the IM by membrane protein MsbA. MsbA gene was first identified in 1993 (Karow et al. 1993) and codes a 64-kDa membrane protein. MsbA is an essential ATP-binding cassette (ABC) transporter which transfers a variety of substrates such as ions, lipids, peptides, metabolites across the cell membrane and

also plays a role in multidrug resistance in bacteria (Ward et al. 2007).

3.1.6 Structure and function of LPS transport proteins

Lipopolysaccharide transport proteins (LptA –LptG) are involved in transport of LPS from the IM to the final destination OM of the cell surface (Figure 1.4). The structures of LptA and LptC have been determined, which help us to understand the two proteins' functions in LPS transport (Suits et al. 2008; Tran et al. 2010).

3.1.6.1 LptA

LptA is a perisplasmic protein. In *E. coli K-12*, the precursor protein has 185 residues with a molecular weight of 18.6 kDa. LptA protein is processed after the 23 amino acid signal peptide. LptA is proposed to act as a periplasmic chaperone for LPS translocation across the periplasm (Sperandeo et al. 2007). It has been reported that LptA interacts with the IM protein LptC and the N-terminal domain of the OM protein LptD to form a continuous bridge between the IM and the OM (Freinkman et al. 2012).

Recently, structures of LptA of *E. coli* have been determined (Suits et al. 2008). When the protein was crystallized without LPS, the LptA protomers are packed in a head to tail fashion with two LptA molecules in an asymmetric unit (Figure 3.5 A). When LPS or Ra-LPS was used in protein crystallization, four molecules of LptA are organized in head-to-tail fashion in a fiber like arrangement (Figure 3.5 B). It is unknown whether the structural conformation in the four-molecule-oligomer was induced by the interaction between LptA and LPS as no LPS molecule was observed in the structure. A possible explanation for this arrangement is that LptA is induced to change a conformation for lipid A binding. Although the crystal structure of LptA in complex with LPS has not been demonstrated, it is evident that LptA is essential in

participating in protein-protein interacts with LptC and LptD and in LPS cellular translocation (Suits et al. 2008).

The LptA has a roll-like structure, composed of 16 antiparallel β-strands with a novel fold. The 16 β-strands form two β-sheets, where hydrophobic residues are located in the core of the LptA structure. Using crosslinking experiments, the LptA head to tail interaction arrangement has been validated (Okuda et al. 2012). The interaction between C-terminal residues of LptC and the N-terminal residues of LptA as well as C-terminal residues of LptA interact with the N-terminal domain of LptD form a bridge cross the periplasm to transport the LPS from the inner membrane to the outer membrane (Figure 3.4). To understand how LptA transports LPS, Okuda performed an UV light mediated crosslinking experiment (Okuda et al. 2012). In this experiment LptA was mutated by substituting with unnatural amino acid p-benzoylphenylalanine which contains photo-cross-linker. Cross-links between LptA and LPS were detected upon UV radiation at T32, I36, F95, Y114 and L116. All these hydrophobic residues locate at the inner core of LptA involve in LPS transportation. On the other hand no cross-links were detected at the residues that are at outer side of the core. This suggests that the LPS is transported along the hydrophobic core of LptA (Okuda et al. 2012).

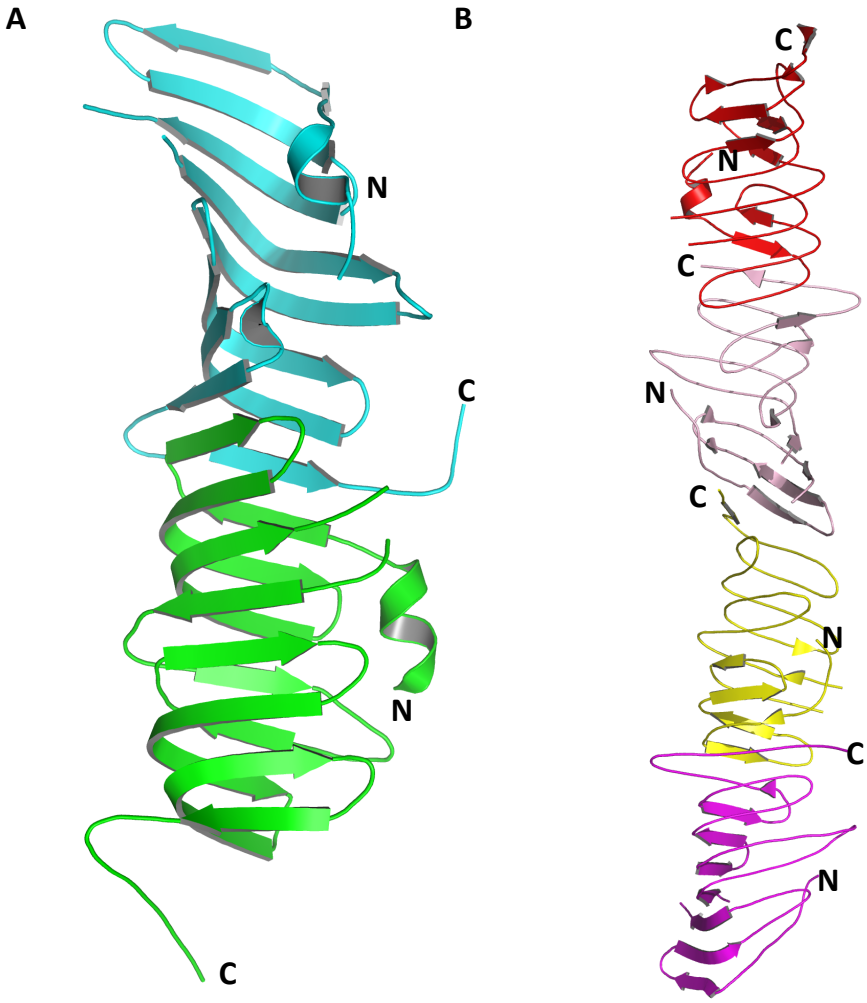


Figure 3.6 Crystal structure of LptA without LPS (A) and in presence of LPS or Ra-LPS (B) during crystallazation. A, Two LptA protomers (cyan, green) in an asymmetric unit, and the monomers are packed in head to tail fashion to form a dimer. N for N-terminus and C for C-terminus (PDB: 2R19). B, Four LptA protomers in an asymmetric unit. The crystals were obtained in presence of LPS or Ra-LPS. The N-terminal residues of LptA interact with adjacent C-terminal residues of β – strand of LptA to form tetramers (PDB: 2R1A) coloured in red, light pink, yellow, purple, respectively.

3.1.6.2 LptF and LptG

LptF and LptG are inner membrane proteins and each of them has been predicted to contain six transmembrane domains in the C-terminus. These proteins are essential to the OM biogenesis of Gram-negative bacteria. Deletion of LptF or LptG in bacteria

causes cell death (Ruiz et al. 2008).

3.1.6.3 LptB

LptB has 241 residues with a molecular weight of 26.7 kDa, which contains an ATP-binding cassette (ABC), and is believed to power the extraction of LPS from the IM to the OM (Okuda et al. 2012). Three of the Lpt proteins form an ABC transporter. Membrane proteins LptFG consist of transmembrane domains (TMD), while LptB possesses nucleotide binding domain (NBD). Together, they form the LptBFG transporter. Interestingly, this transporter does not transport substrate across the inner membrane. Instead, the transporter extracts LPS from the periplasmic side of the inner membrane, and passes it to another membrane protein LptC (Okuda et al. 2012; Narita et al. 2009).

LptB forms a dimer in solution. The protomeric structure is conserved amongst other nucleotide-binding proteins, which consists of ten α-helices and ten β-strands (Figure 3.6). The structure can be divided into two domains, the RecA-like domain and α-helical domain with the conserved motifs, Walker A, Walker B, H-loop, Q-loop, D-loop and Signature motifs (Wang et al., 2014). Energy is required for LPS transport from the LptC to LptA. Deletion of LptB results in impaired LPS transport to the OM and killing of the bacteria (Sperandeo et al. 2007). Mutations of the ATP binding residues and catalytic residues result in the cell death of *E. coli*.

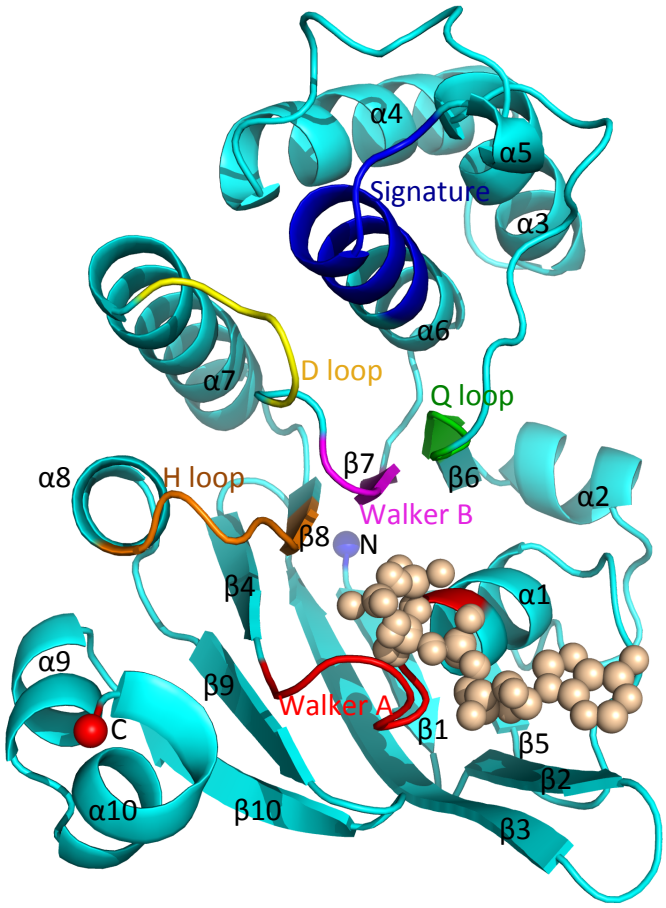


Figure 3.7 Protomer structure of LptB with ATP binding. LptB consists of ten α-helices and ten β-strands with typical ATP binding motifs, Walker A, Walker B, H-loop, D-loop, Q-loop and signature motifs (PDB: 4QC2). N for N-terminus and C for C-terminus, ATP in yellow white.

3.1.6.4 LptC

LptC is an essential inner membrane protein, which plays a role in exportation of LPS from the IM to the OM. LptC interacts with LptBFG to form a complex (Villa et al. 2013; Freinkman et al. 2012), but does not affect the ATPase activity of the transporter *in vitro* (Narita et al. 2009). The LptC gene encodes a membrane protein of 191 amino acids and a molecular weight of 21.7 kDa. LptC consists of two *β*-sheets of seven and eight antiparallel *β*-strands respectively in opposition to each

other (Figure 3.7). The crystal structure of LptC is similar to LptA, although the two proteins have low similarity in amino acid sequence. LptC, like LptA, can bind LPS *in vitro* (Tran et al. 2010), and LPS is transported only from LptC to LptA.

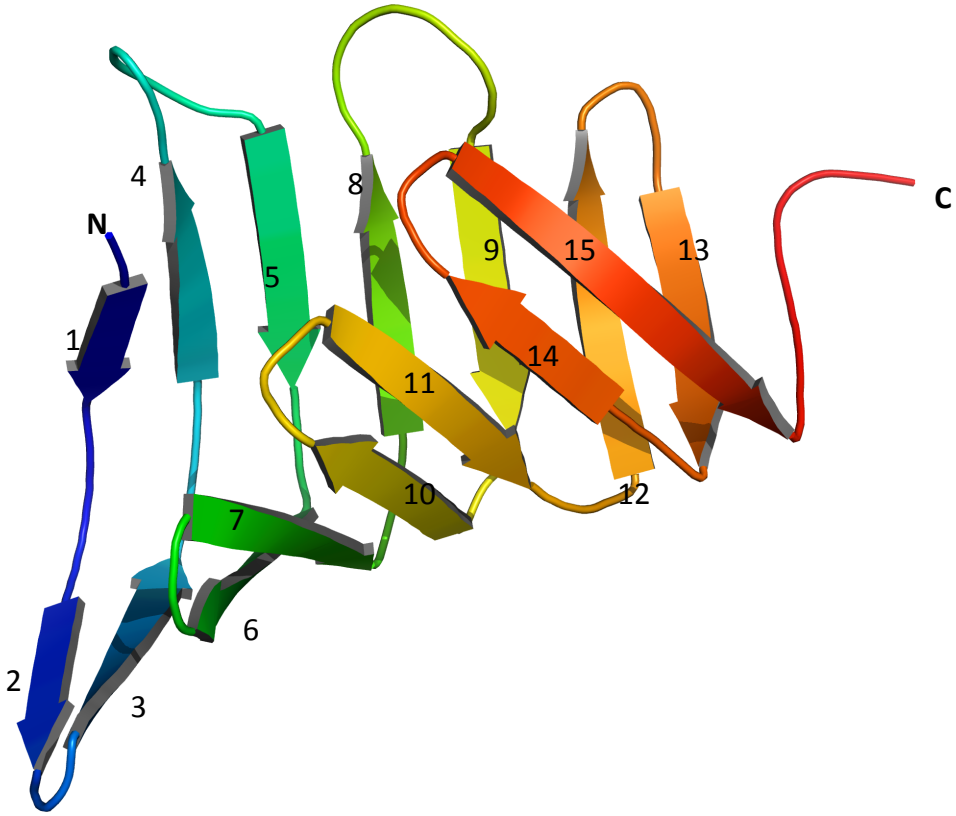


Figure 3.8 Crystal structure of LptC. The structure shows a *β*-jellyroll fold and consists of 15 antiparallel *β*-strands arranged two *β*-sheets in an opposite direction (PDB: 3MY2).

3.1.7 LptD/E form a complex for LPS insertion

Outer membrane protein LptD and lipoprotein LptE are responsible for LPS insertion. LptD/E can form a stable two-protein complex *in vitro* (Chimalakonda et al. 2011; Freinkman et al. 2011; Wu et al. 2006; Chng et al. 2010).

LptD has a molecular weight of 87-kDa. It is an essential β-barrel outer membrane

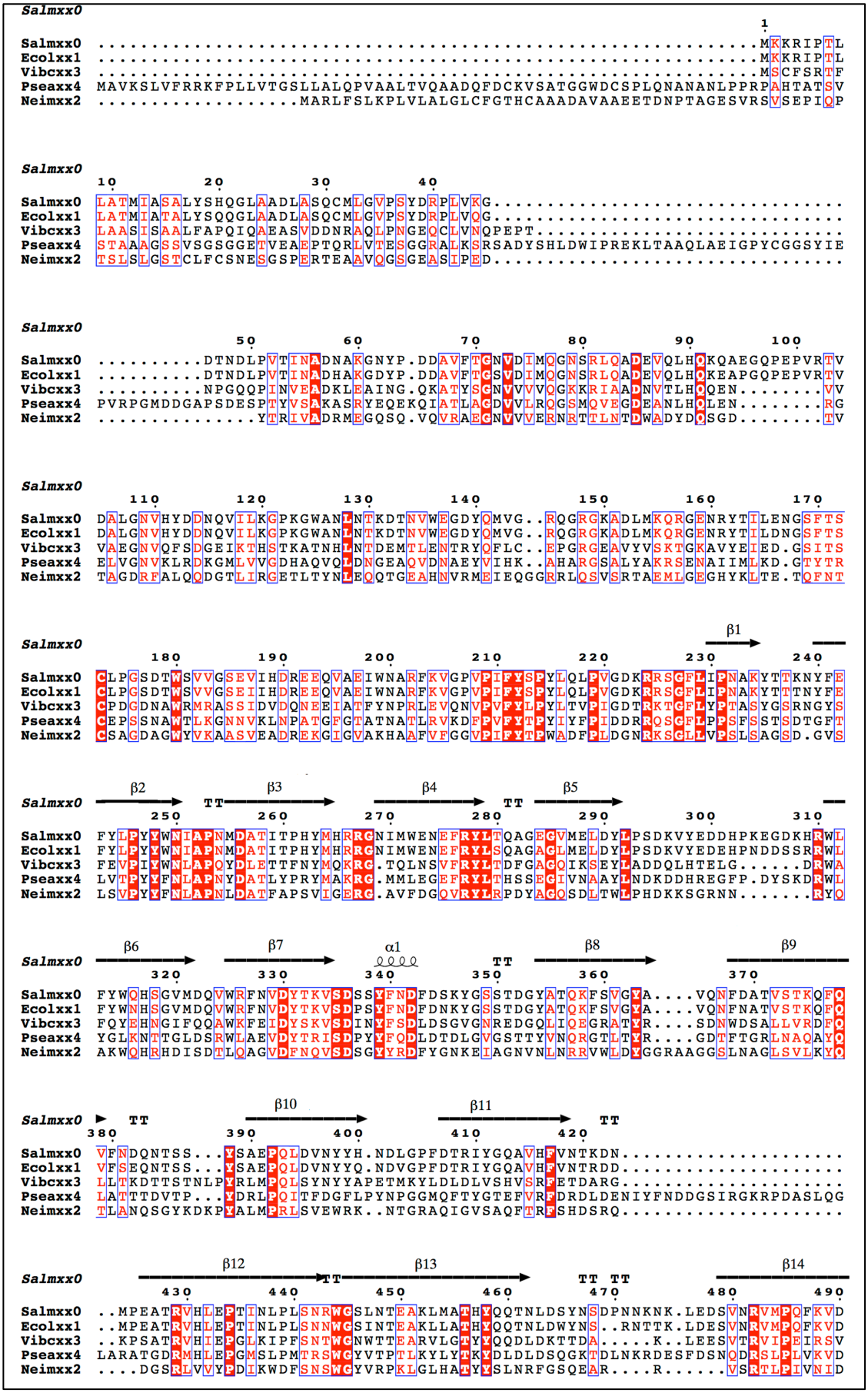
protein (OMP), whose function is to assemble LPS into the OM of the cell surface. LptD possesses a large periplasmic N-terminal domain (amino acids 25 to 202) and a C-terminal transmembrane β-barrel domain (amino acids 203 to 784) (Chng et al. 2010). There are four cysteine (Cys) residues in LptD, two (Cys31, Cys173) in the N-terminal domain and two (Cys726, Cys727) in the C-terminal domain. These cysteine residues form two non-consecutive disulphide bonds, Cys31 and Cys726 form the first disulphide bond, while Cys173 and Cys727 form the second disulphide bond, and both of them are connected from the N-terminal residues to the C-terminal residues (Figure 3.25). Disulphide bond formation in LptD is essential in transport of LPS to the OM; LptD is completely oxidised *in vivo* (Chng et al. 2012; Ruiz et al. 2010). LptD is synthesized on ribosomes in the cytoplasm with an N-terminal signal peptide, which allows the protein to pass to the Sec machinery (Figure 3.2 B). After passing through the Sec channel across the IM, its signal peptides are removed by signal peptidase (Hagan et al. 2011). LptD is thought to be protected by periplasmic chaperones (primarily SurA) in an unfolded form in periplasm, and is then finally transported into the OM by outer membrane protein assembly machinery BamABCDE (Figure 3.2 B). Periplasmic chaperones SurA, Skp and DegP are involved in transport and folding of outer membrane β-barrel proteins (Vertommen et al. 2009; Kim et al. 2007; Silhavy et al. 2010).

LptD is unable to fold properly without the interaction with the lipoprotein LptE. The C-terminal domain of LptD interacts tightly with LptE, and the two proteins form a stable 1:1 complex *in vitro*. LptD can protect LptE from proteolytic digestion *in vivo* (Wu et al. 2006; Chimalakonda et al. 2011; Chng et al. 2010). LptD/E complex forms a unique 'barrel and plug' architecture for LPS transport and insertion (Freinkman et al. 2011; Grabowicz et al. 2013). Deletion of LptD/E resulted in defect of the OM

biogenesis and caused the cell death of *E. coli* (Wu et al. 2006).

Although LPS biosynthetic pathway is well understood, how LPS transports from the IM to the OM and assembly at the cell surface still remains unknown. Lpt proteins are essential for OM biogenesis in most pathogenic Gram-negative bacteria, including *Salmonellae* and *Pseudomonades*. LptD in *pseudomonades* has been reported to be an ideal target for the development of novel peptidomimetic antibiotics against multi-drug resistant bacteria (Srinivas et al. 2010). Multi-drug resistant Gram-negative bacteria, such as *E. coli*, pose a global health threat. This limits the effectiveness of existing antibiotics in controlling infections. The World Health Organization (WHO) recently warned that antibiotic-resistance in bacteria becomes a global health problem (World Health Organization 2014).

The unique characteristic of Gram-negative bacteria carrying a LPS outer membrane endows these bacteria with antibiotic resistant properties. The fact that LptD and LptE are essential mediators for the biogenesis of this LPS outer membrane in most pathogenic Gram-negative bacteria and that they are highly conserved across species (Figure 3.8, 3.9) made us particularly interested in these two proteins.


Salmxx0
Ecolxx1
Vibcxx3
Pseaxx4
Neimxx2
β1
β2
β3
β4
β5
β6
β7
α1
β8
β9
β10
β11
β12
β13
β14

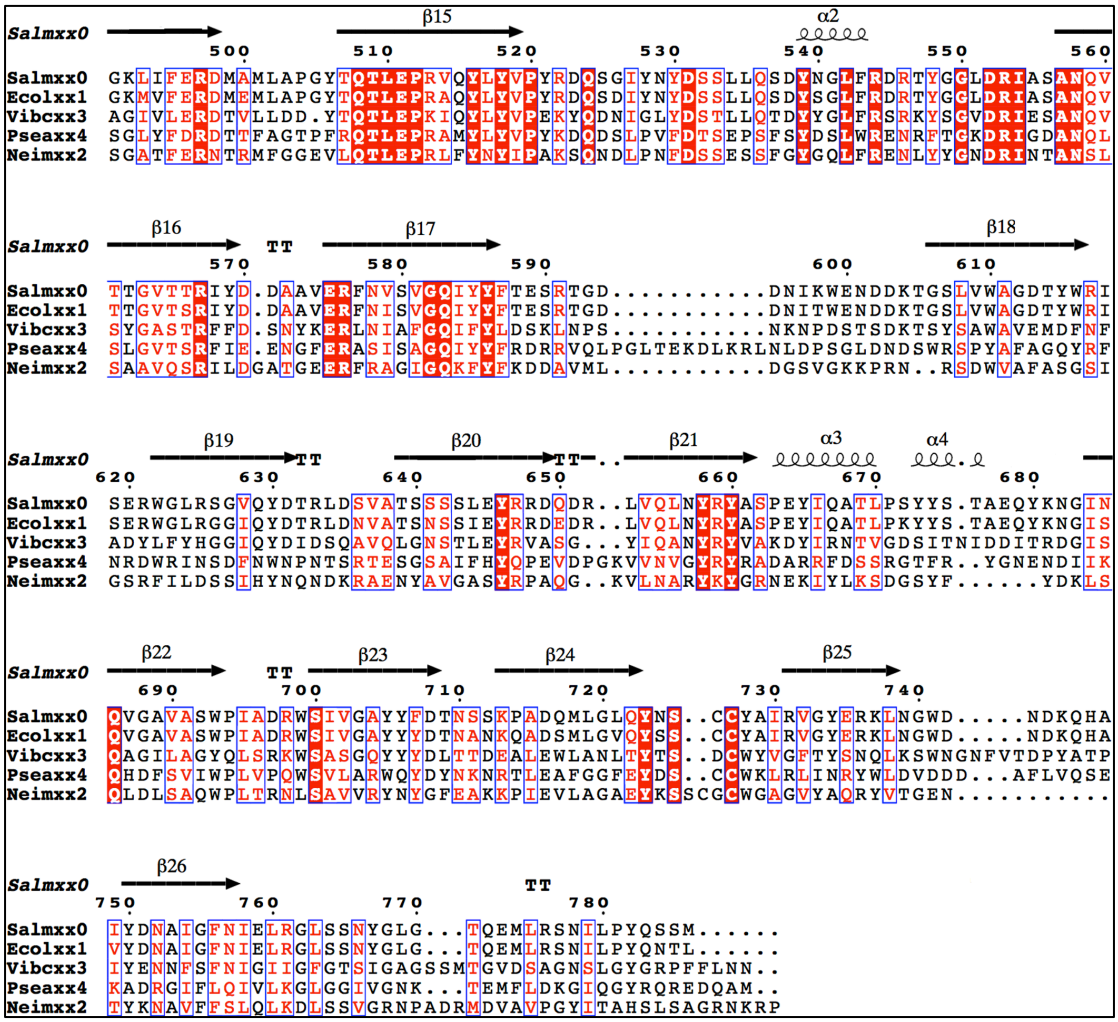


Figure 3.9 Amino acid sequence alignment of LptD. The C-terminal domain of LptD forms a 26-stranded β barrel which is highly conserved in Salmxx0, *Salmonella* Typhimurium, accession GI25008880; Ecolxx1, *E. coli*, accession GI 2507089; Vibcxx3, *Vibrio cholera*, accession GI67477419; Pseaxx4, *Pseudomonas aeruginosa*, accession GI25008883; Neimxx2: *Neisseria meningitides*, accession GI134034978. The dot lines represents theses LptD amino acid sequences lack the residues, The N-terminal of LptD from the *Pseudomonas* strain is much longer than other species.

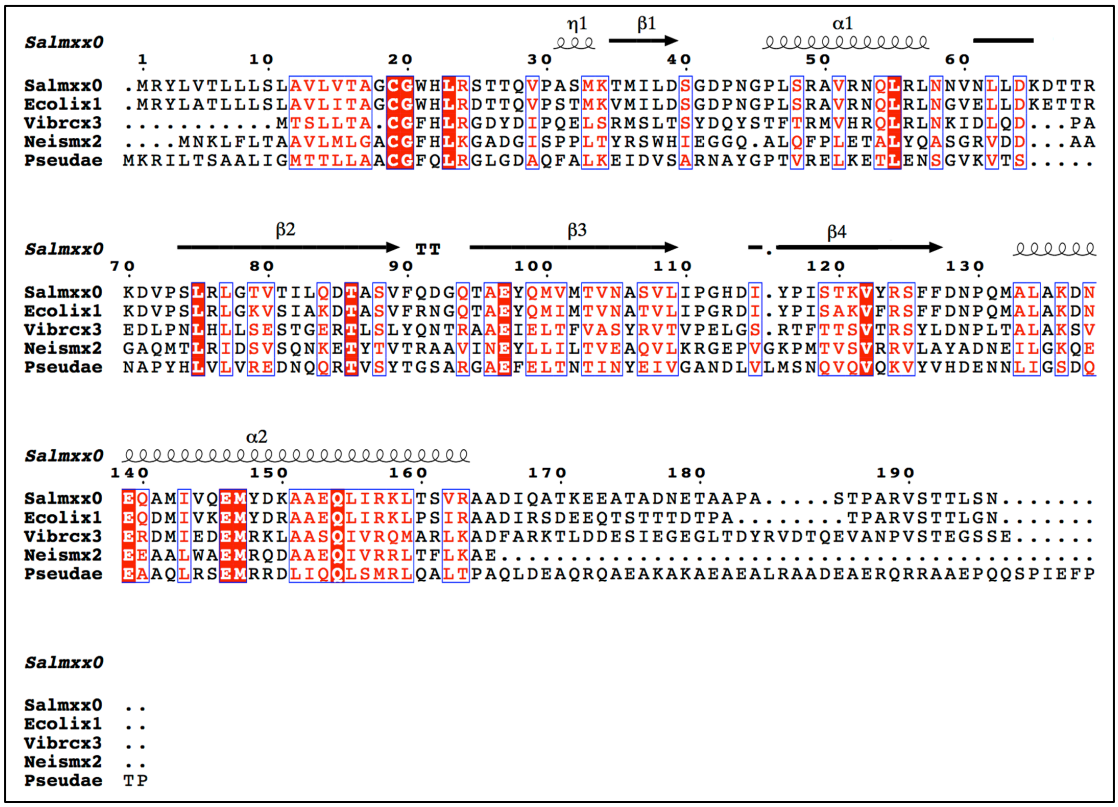


Figure 3.10 Amino acid sequence alignment of LptE. The C-terminal residues of LptE from different bacteria are highly variable. Although the sequence identity of LptE is low, secondary and tertiary structure is conserved. Salmxx0, *Salmonella* Typhimurium, accession GI81523600; Ecolix1, *E. coli*, accession GI259491800; Vibrcx3, *Vibrio cholera*, accession GI469684423; Neismx2, *Neisseria bacilliformis*, accession GI389606221; Pseudae, *Pseudomonas aeruginosa*, GI15599183. The dot lines represents these amino acid are absent in these proteins .

3.2 Aims

The LptD/E complex represents a particularly attractive drug target, because drug candidates would not need to penetrate the bacterial cell wall. The development of such new antibiotics has been hampered by not having a detailed model of the LptD/E complex. This study aims to determine the crystal structure of the LptD/E through combined approaches including cloning, protein expression, purification, crystallization the LptD/E complex, and using X-ray diffraction.

3.3 Results

3.3.1 Crystallization of LptD/E

The LptD/E protein complex was successfully co-expressed in *E. coli* subtype C43 (DE3) cells (Avidis) and was purified with nickel-nitrilotriacetate affinity resin (Ni-NTA, Qiagen) (Figure 3. 10 B). The protein was further purified using size exclusion chromatography (GE Healthcare) (Figure 3.10 A). The crystal trails were set up with different detergents and no crystals were obtained.

3.3.1.1 Limited proteolysis of LptD/E

In general, an average of only 30% of purified protein can form crystals and amongst which only 15% form high quality crystals that can be used for crystal structure determination (A. Dong et al. 2007; Wernimont et al. 2009). Low crystallization efficiency is the bottleneck in the field of protein crystallographic study. Dong et al. reported that adding small amounts of protease to crystallization trials to remove disordered regions of protein could increase chances of forming good quality crystals for diffraction (A. Dong et al. 2007).

In Wernimont and Edwards' study, 270 purified proteins failed to produce crystals or high quality crystals for structural determination. After proteolysis, 34 produced sufficient quality crystals with an average 1.8 angstrom diffraction resolution (Wernimont et al. 2009). Proteinases chymotrypsin and trypsin are the most successful proteinases used for crystallization. Among the 34 crystal structures, 12 structures were determined by using trypsin protease, 14 structures were solved by using chymotrypsin and 5 structures were solved by using V8 protease (Wernimont et al. 2009). Proteolysis has been proven to be a successful method to obtain crystal structures. Therefore, I decided to use proteases to treat the purified LptD/E protein

in an attempt to obtain LptD/E complex crystals.

The most stable LptD/E fragment was obtained by using α-Chymotrypsin (Sigma-Aldrich) at ratio 1:100 for 180 minutes. Subsequently a large scale of LptD/E protein preparation was mixed with α-Chymotrypsin at ratio 1:100 for 180 minutes. After limited proteolysis, the protein was further purified using size exclusion chromatography (Figure 3. 10 B). The main peak is the LptD/E complex. The corresponding fractions were pooled and concentrated to 8 mg ml^{-1} for crystallization. Protein concentration was determined by measuring the absorbance at 280 nm using a Nanodrop (Thermo Scientific).

The resulting LptD/E proteins were confirmed by SDS-PAGE and mass spectrometry (University of St Andrews), which revealed that the N-terminal domain residues 25–211 of LptD (Figure 3. 10 C) and the C-terminal residues 170–194 of LptE were removed by α-chymotrypsin.

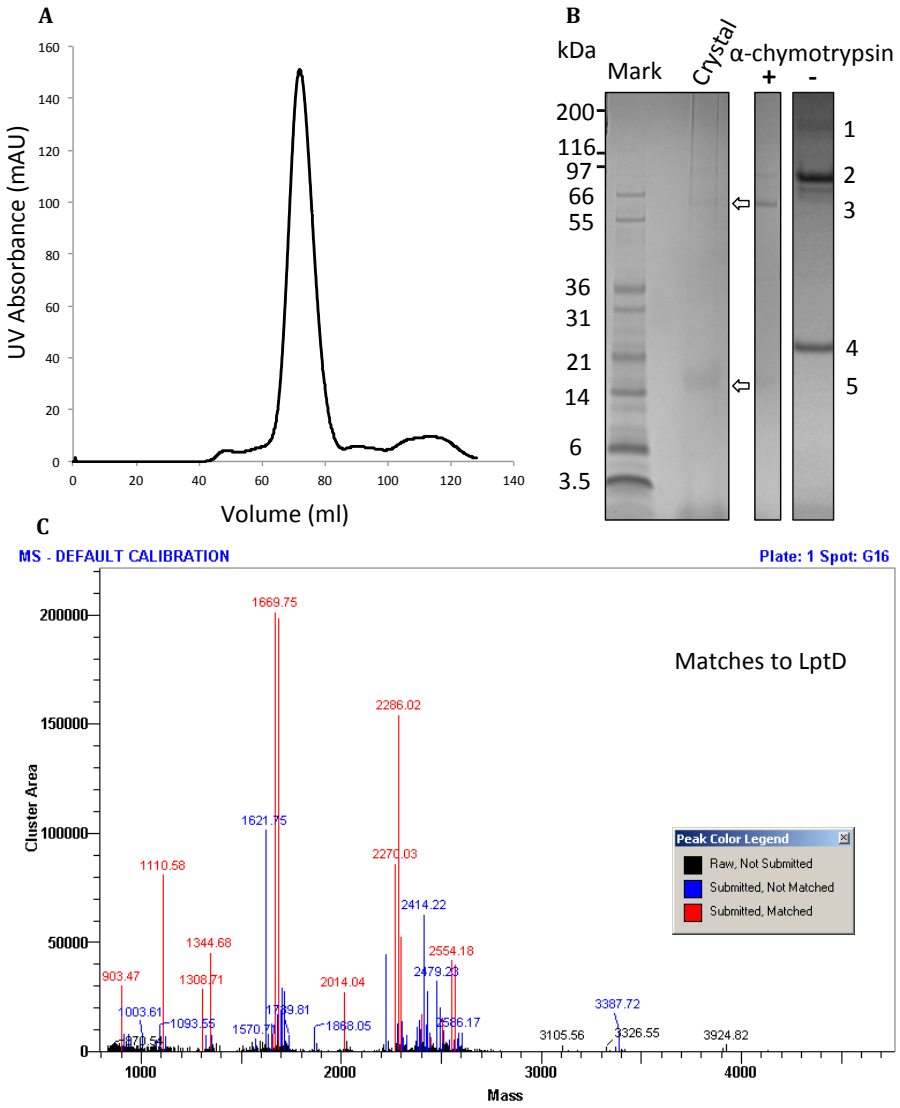


Figure 3.11 LptD/E proteolysis, purification and mass spectrometry. A, Gel-filtration of LptD/E. The main peak represents the stable LptD/E complex after proteolysis. B, Limited protease digestion of the LptD/E complex was carried out for 180 minutes at room temperature using a-chymotrypsin. The N-terminal domain residues 25–211 of LptD and C-terminal residues 170–194 of LptE were removed by this protease. Band 1, the oxidized LptD (130 kDa). Band 2, reduced and un-cleaved LptD (87 kDa). Band 3, cleaved LptD (62 kDa). Band 4, un-cleaved LptE (21.5 kDa). Band 5, cleaved LptE (18.7 kDa). This result is according to that of the LptD/E complex of *E. coli* by trypsin digestion (Chng et al. 2010). The crystal was washed in crystallization buffer three times to avoid protein contamination.. **+,** indicates that protein incubated with a-chymotrypsin. **-,** as control . C, Mass spectrometry analysis confirmed that the N-terminal residues 25-211 of LptD was removed**.**

The α-Chymotrypsin resulting LptD/E crystals were obtained and optimized by varying the precipitant and then frozen in liquid nitrogen and diffracted to 3.9-angstrom resolution. The best crystals were obtained in 0.15 M zinc acetate, 0.08 M sodium cacodylate pH 6.5 and 16% (w/v) PEG 8000 (Figure 3.11)

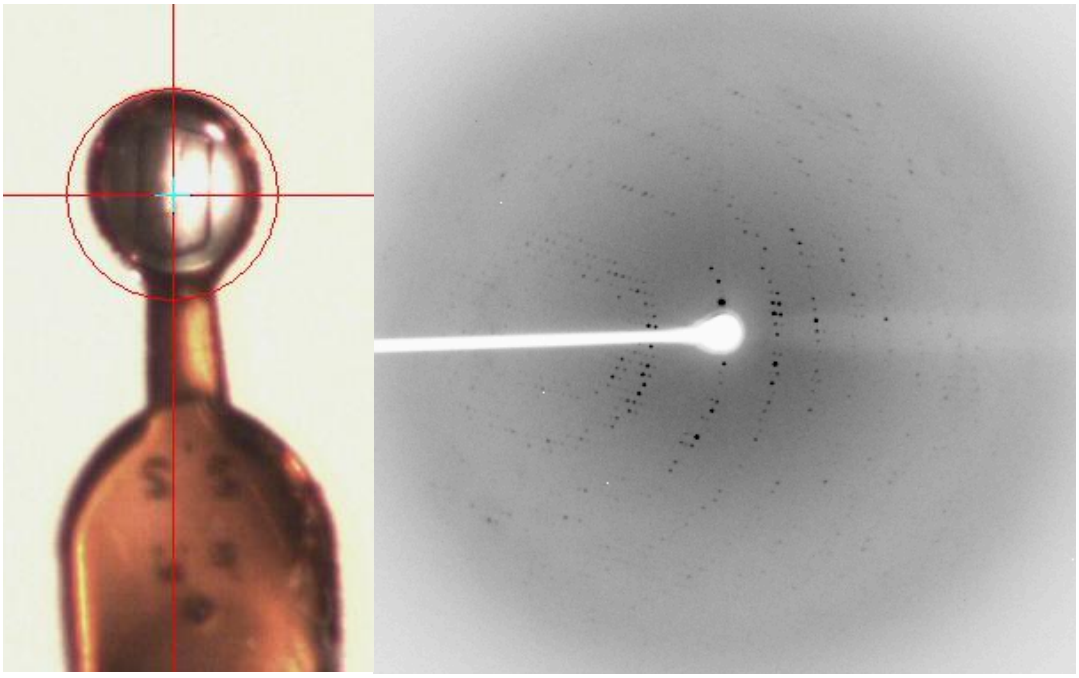


Figure 3.12 Crystal of LptD/E and X-ray diffraction pattern. The left image is protein crystal of LptD/E mounted with 0.2 μm litho-loops (Molecular Dimensions), and the right image is the crystal diffracted to 3.9 angstrom resolution using in-house x-ray source.

3.3.2 Crystallization of selenomethionine labeled LptD/E

As there is no LptD structure available, LptD/E complex structure could not be determined by molecular replacement. Heavy atom soaking experiments were attempted but they destroyed the crystals. I then decided to make selenomethionine (SeMet) labeled LptD/E protein. SeMet-labeled protein can be expressed by substituting methionine with selenomethionine (Bakke et al. 2010; Boles et al. 1991). The SeMet labelled crystals can be used to determine the protein structure using single or multi wavelength anomalous dispersion (MAD) (Hendrickson, Horton, and LeMaster 1990; Hendrickson 1991).

The purified SeMet labeled LptD/E fractions were collected, and concentrated to 7 mg ml^{-1}. The crystallization trials were set up using both sitting-drop and hanging drop vapour diffusion, based on the native crystallization conditions as mentioned above. The SeMet-LptD/E complex crystals were produced, which showed better quality than its native crystals. Crystals were further optimized to attain higher resolution. The best crystallization condition is 0.15 M zinc acetate, 0.08 M sodium cacodylate pH 6.2 and 14% (w/v) PEG 8000 (Sigma-Aldrich). All the crystals were harvested after 16 days and cryoprotected by supplementing the crystallization solution with 20% glycerol in the crystal well before being flash frozen in liquid nitrogen. The SeMet LptD/E complex crystals were screened in house and were diffracted to 3.2-3.6 Angstrom (Figure 3.12).

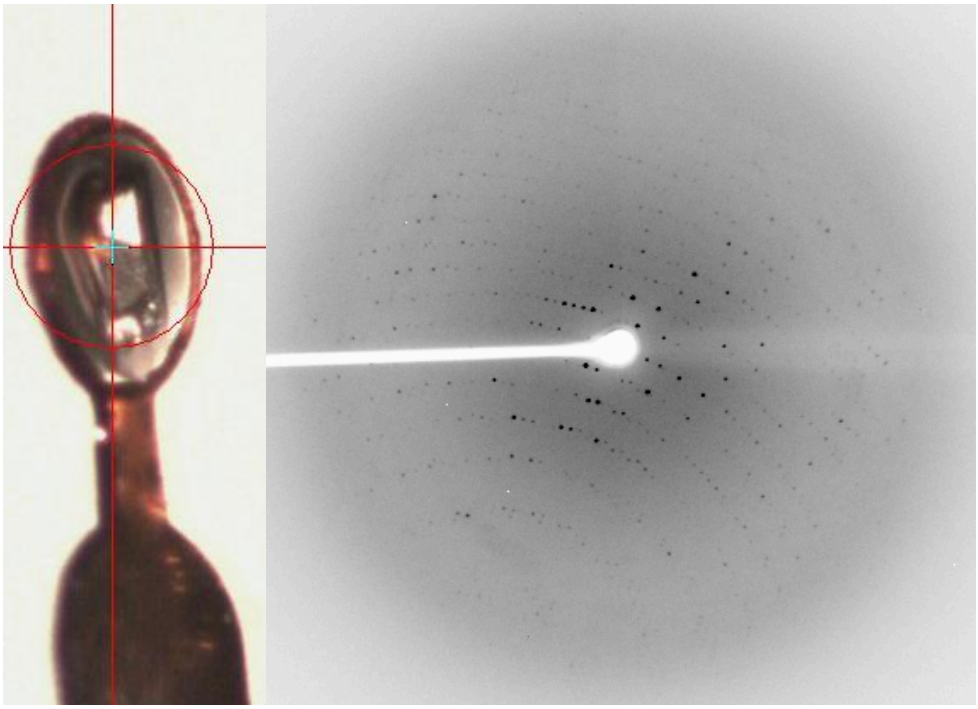


Figure 3.13 Crystal of SeMet labelled LptD/E and X-ray diffraction pattern in house. The left image is protein crystal of SeMet labeled LptD/E. The crystal is thick and shines which was mounted with 0.3 μm litho-loops (Molecular Dimensions). The right image is the crystal diffracted to 3.2 angstrom resolution in house.

3.3.2.1 Data collection of selenomethionine labeled LptD/E

The MAD datasets were collected at Diamond Light Source, UK with a 2.86-angstrom resolution (Figure 3. 13).

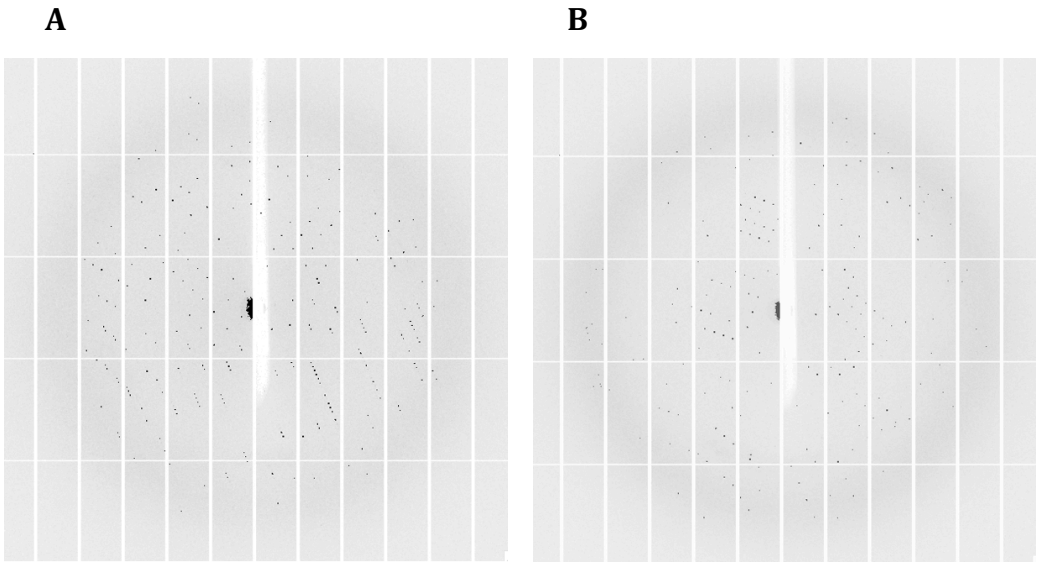


Figure 3.14 X-ray diffraction pattern. The data were collected at Diamond Light Source, UK and the diffractions were isotropic. A, Image was collected at 0 degrees to 2.8 Å. B, Diffraction image collected at 90 degrees to 3.2 Å.

The MAD collection strategies were described as in Chapter 1.1.5.6. The crystals belong to space group I2 with unit-cell dimensions: a = 173.430 Å, b = 76.082 Å, c = 213.596 Å, $\alpha = \gamma = 90°$ and $\beta = 111.519°$. The data collection statistics are listed below in table 1.1.

Data collection	**Peak**	**Low**	**High**	**Inflection**
Wavelength (Å)	0.9784	0.9818	0.9775	0.9788
Resolution (Å)	43.92-3.00	43.90-3.00	43.94-3.00	43.95-3.00
Space group	I2	I2	I2	I2
Cell dimensions (Å/)	a =173.430 b= 76.1 c= 213.6 α = γ = 90° β=111.519°	a = 173.430 b = 76.0 c= 213.7 α= γ = 90° β=111.519°	a = 173.430 b = 76.3 c = 213.6 α = γ = 90° β = 111.6°	a = 173.430 b= 76.082 c = 213.596 α = γ = 90° β = 111.519°
Unique	52322 (4472)	52295	52427	52430
I/σ (I)	15.4 (1.8)	14.7 (1.9)	15.6 (2.0)	17.7 (2.5)
Anomalous completeness (%)	99.9 (100)	100 (100)	100 (100)	100 (100)
Anomalous redundancy	15.9 (16.3)	16.1 (16.3)	16.0 (16.3)	16.0 (16.3)
Refinement				
Resolution				43.92-2.86 (2.96-2.86)
R_{factor} / R_{free}				0.27/0.31
Ligand atoms				12
Solvent atoms				10
R.M.S.D.				
Bond (Å) / Angle				0.009/1.36
B-factors (Å^2)				
Protein				110.5
Ligand				100.0
Solvent				102.1
PDB code				4N4R

Table 1.1 SeMet of LptD/E data collection statistics. Values in parentheses are represents for the highest-resolution shell. RMSD, root mean square deviation. $\mathbf{R_{factor} = \Sigma \, || F_{obs}| - | F_{cal}|| / \Sigma \, |F_{obs}|}$, where F_{obs} and F_{cal} are observed all reflection measured and calculated currently model as structure factors, respectively. R_{free} is calculated using 5% of total reflections, which is randomly selected not used in refinement.

3.3.3 Structure determination of LptD/E

All data were processed using XDS (Kabsch 2010). Unmerged data were used to determine the phases problem using SHELX suite (Sheldrick 2007). The anomalous signal was calculated using SHELXC, and four wavelength of data resolution were also displayed (Figure 1.13 A). The best correlation coefficient was 33.6 using SHELXD, indicating that initial phases were an unreliable solution for the structure. Based on the SHELXD manual, the best correlation coefficient should be over 40% for a reliable solution (Figure 1.13 B). The figure of the occupancy of 44 Se sites did not show a clear solution (Figure 1.13 C) as well. The separation of the contrast and connectivity of the maps from the SHELXE is small between the original and the inverted hands, suggesting that phases were not determined (Figure 1.13 D).

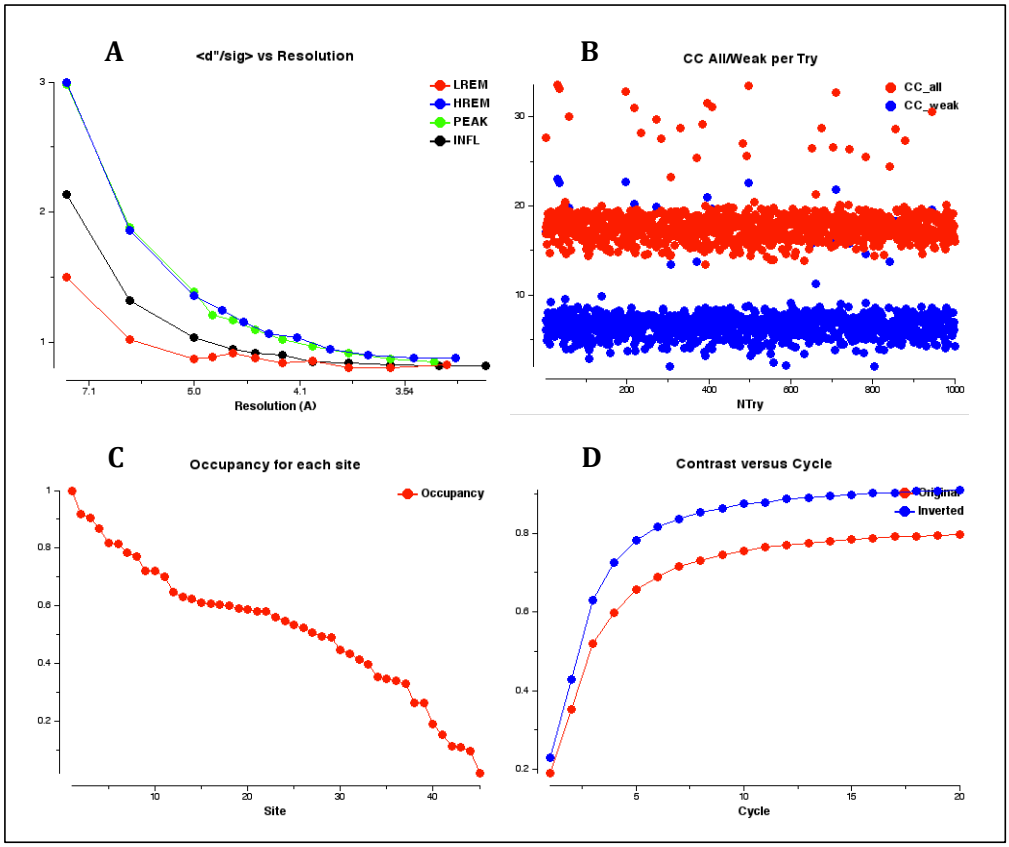


Figure 3.15 The first trial of determination of LptD/E structure using SHELX. A, Analysis of the data set using SHELXC, <d"/sig> is means of anomalous signal for peak wavelength as a function of resolution. B, The correlation coefficient (CC_ all, CC_ weak) was about 33.6% by SHELXD. C, The heavy atoms of Se were detected using SHELXD. D, Electron density maps were modified by SHELXE.

Despite that no solution was obtained, we tried to build the model automatically using the program Buccaneer (Cowtan 2006) with the phases obtained from the SHELX. As expected, after 100 building cycles, the initial model was not successfully built with an R_{free} of 0.5017 and overall figure of merit of 0.7439. The structure built by Buccaneer was checked using Rasmol (Goodsell 2005), which did not show a predominantly secondary structure (Figure 3.15 A).

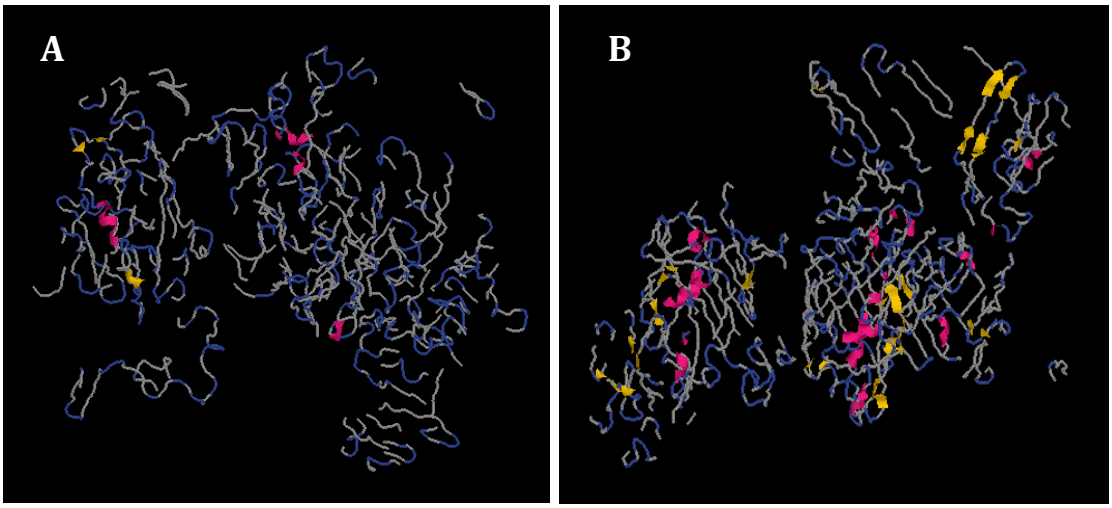


Figure 3.16 The initial model of LptD/E. A, The lptD/E model was built by 100 cycles of Buccaneer. Rasmol was used to check the structure. B, The LptD/E complex structure was built using Buccaneer for 5 cycles.

In order to enhance the anomalous signal, we combined the MAD datasets from 4 different crystals using the Aimless program from CCP4 (Winn et al. 2011). The combined peak, inflection, high remote and low remote data were used for the structural determination.

The model was built in 5 cycles using Buccaneer. The structure was checked using Rasmol (Goodsell 2005) (Figure 3.15 B). The LptD/E complex was built using Buccaneer for 100 cycles. Although the structure was not completely built, the model of second structure was observed using program PyMOL (Figure 3.16) (DeLano 2002).

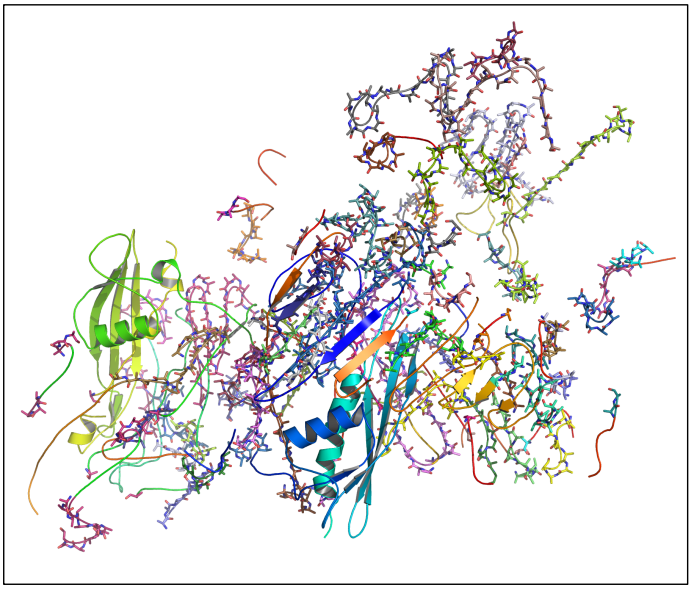


Figure 3.17 Crystal structure of LptD/E. The LptD/E structure was built automatically using Buccaneer for 100 cycles.

The structure was determined by AutoSharp (Vonrhein et al. 2007). The model was partially built by Buccaneer (Cowtan 2006) and finished using manual model building in Coot (Emsley and Cowtan 2004). The structure was refined with REFMAC5 (Murshudov et al. 2011).

3.3.4 The crystal structure of LptD/E complex

The crystal structure of LptD contains the residues from 226 to 786, which included 26 anti-parallel β-barrel strands, and 13 extracellular loops (Lp1 to Lp13). LptE consist of residues 19 to 169, which formed 2 alpha and 4 parallel β-strands (Figure 3.18, 3.19). LptD forms β-barrel with dimensions approximately 70 Å in length, 50 Å in width and 50 Å in height (Figure 3.18), while LptE forms a roll like structure with three quarters of it inserted into LptD and the remaining quarter extended into the periplasm (Figure 3.19 C, D). The LptD barrel is enclosed by strands β1 and β26. The N terminal domain is located in the periplasm (Chng et al. 2010), while the C-terminal residues are covered inside the LptD barrel on the periplasmic. The lipid modified N-terminus of LptE is located in a perfect positi on for insertion into the OM (Figure 3.18).

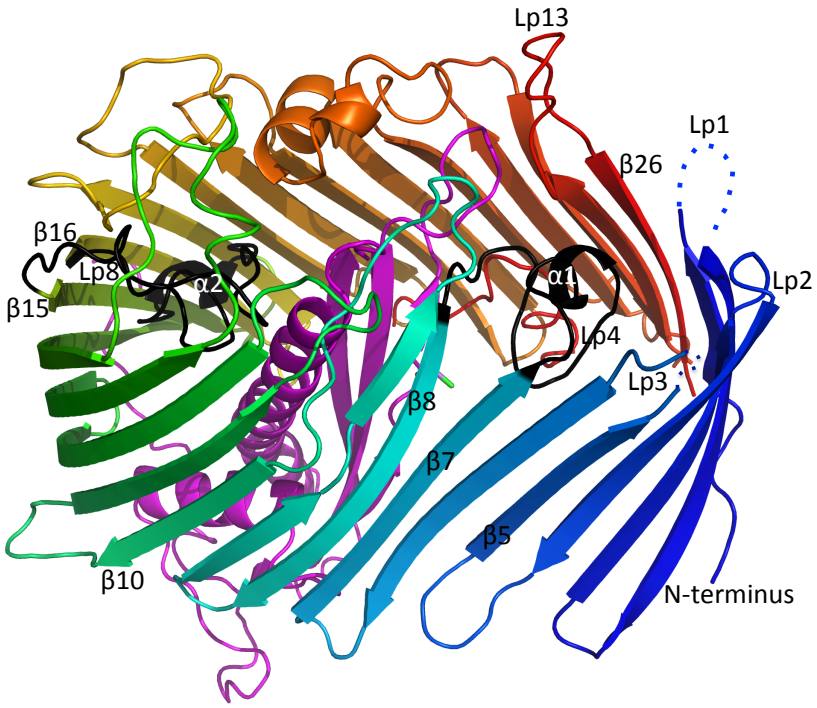


Figure 3.18 Crystal structure of LptD/E complex. The LptD barrel containing 26 anti-parallel strands and 13 extracellular loops, shown in rainbow colour, Lp4 and Lp8 are coloured in dark. The N terminus in blue, the C terminus in red, and LptE in purple. Disordered parts are represented as dashed

lines.

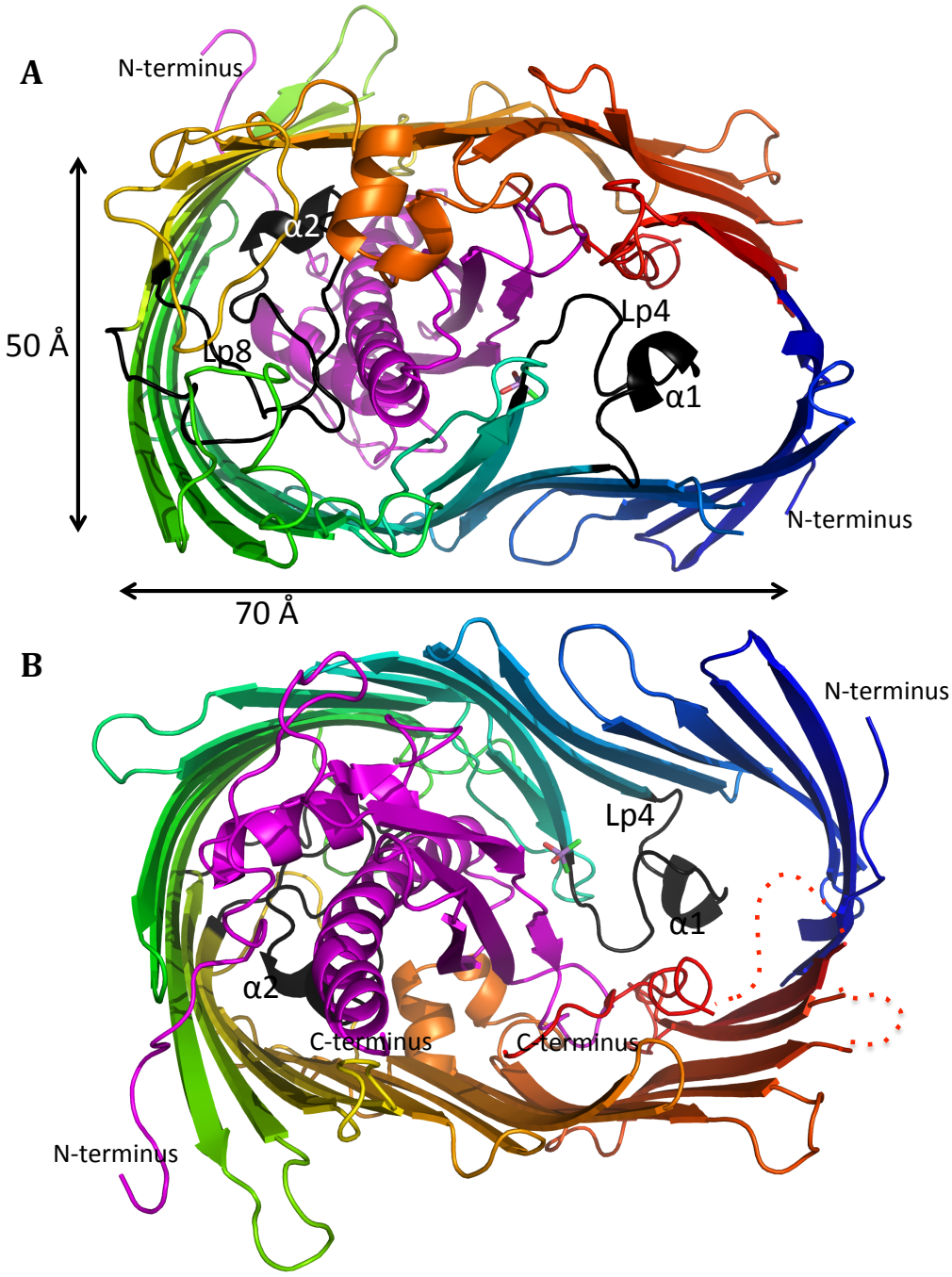


Figure 3.19 Top and bottom view of the LptD/E complex structure. The left image is top view of the structure. Extracellular loops Lp4 and Lp8 are located in the interior of the barrel, while other loops are at the surface of the barrel pore. The right image is bottom view the structure of LptD/E complex.

There are aromatic residues located on the outer sidewall of the LptD barrel, which play a role in helping the barrel to insert into the outer membrane (Figure 3.19 A, B). The previously reported largest single outer membrane usher protein barrels, PapC and FimD (Remaut et al. 2008; Phan et al. 2011), are responsible for translocation of P pilus and type 1 pili subunits respectively. They comprise 24 anti-parallel β-strands and the barrels are completely blocked by a plug formed through the middle domain of PapC or FimD (Remaut et al. 2008; Phan et al. 2011). In contrast, LptD forms a 26-stranded β barrel and the "plug" is from another protein LptE. To our knowledge, LptD is the largest single bacterial outer membrane β-barrel protein reported to date, and also the only barrel structure that uses another protein as a "plug" (Chng et al. 2010; Grabowicz et al. 2013; Fairman et al. 2011).

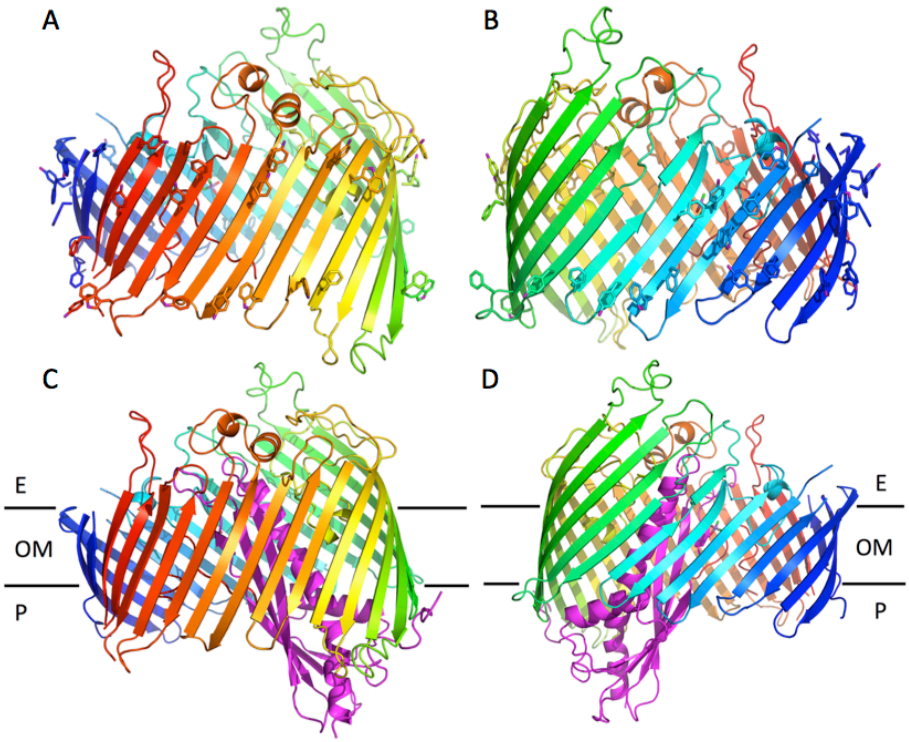


Figure 3.20 From side view structure of the LptD barrel and LptD/E complex. A, B, Figure A rotates 180° along the y axis to Figure B, the aromatic residues located in the barrel outer wall are shown as sticks. C, D, Figure C rotates 180° along the y axis to Figure D. The LptD/E complex in the outer membrane. E, OM and P represent extracellular space, the outer membrane and periplasm, respectively (Figure adapted from H. Dong et al. 2014).

To determine whether *S. typhimurium* LptE had undergone any changes in its structural conformational in its structure upon LptD/E complex formation, the LptE structure of *S. typhimurium* was superimposed and compared with four available LptE structures (X-ray crystal and NMR) in protein data bank (PDB). It is remarkable that the LptE was strikingly superimposable with the available structures (2R76, 3BF2, 4KWY and 2JXP), even though the sequence identities are as low as 13%. The obvious structural differences between LptE *S. typhimurium* and other LptEs were observed in the loops (Figure 1.22). The structure of LptE located inside LptD indicated that it play an essential role in LptD's folding and assembly. We predict that most of the LptD/E complexes of Gram-negative bacteria possess a similar structure to that of the *S. typhimurium* protein. Therefore, the structure of the *S. typhimurium* LptD/E translocon may provide a common model for studying LPS translocation for most Gram-negative bacteria.

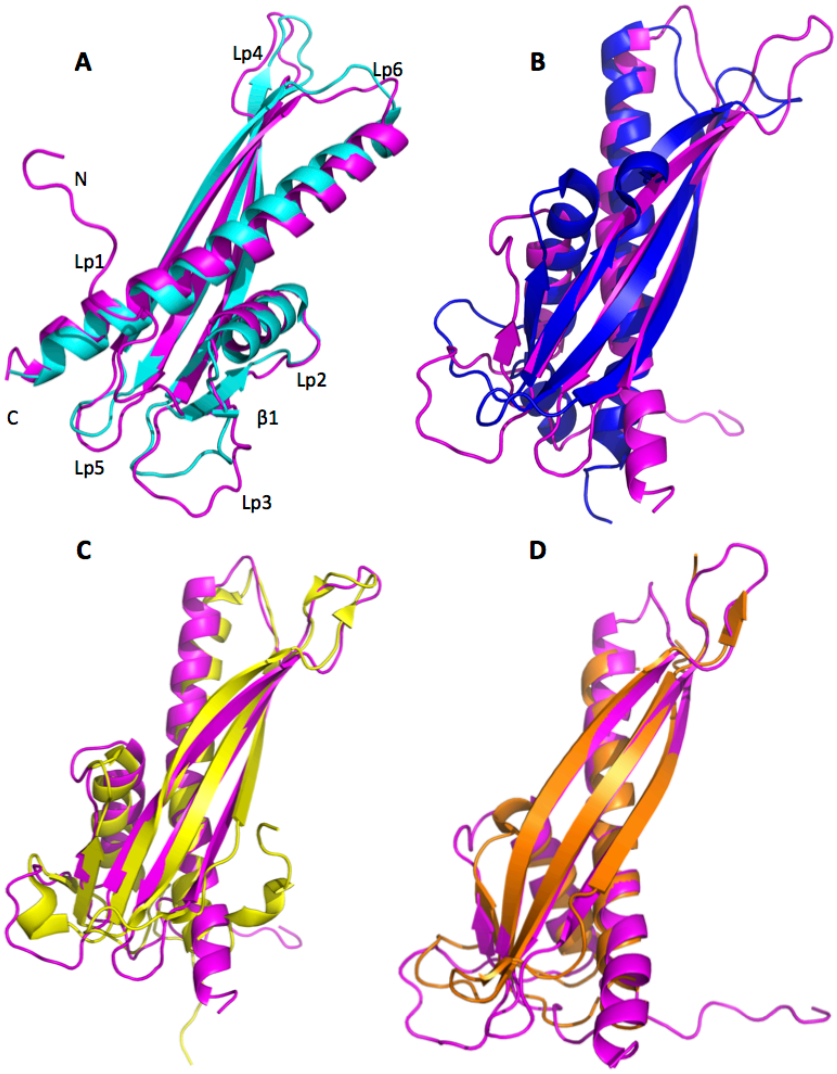


Figure 3.21 Crystal structure of LptE from *S. typhimurium* (purple) superimposed with other homologues. The *S. typhimurium* LptE structure from the LptD/E complex is very similar to other structures of LptE. A, the cartoon representation of the LptE structure superimpose with the LptE from *Shewanella oneidensis* (rmsd of 1.68 over 131 Cα). LptE of *S. oneidensis* (2R76) is shown in cyan. B, LptE superimposition with LptE of *C. crescentus* CB15 (4KWY) coloured in blue with r.m.s.d. of 3.046 over 115 Cα. C, LptE superimposition with LptE of *N. europaea* (2JXP), coloured in yellow with r.m.s.d. of 2.3177 over 128 Cα. D, LptE superimposition with LptE of *N. meningitidis* (3BF2) coloured in orange with root-mean-square deviation (RMSD) of 1.6533 over 115 Cα (Figure adapted from H. Dong et al. 2014).

3.3.5 Extensive interaction between LptD and LptE

The β-strands in LptD are linked by 13 extracellular loops (Lp1 to Lp13) and most of the loops are located at the surface of the structure and these loops on the extracellular side are longer than the loops on the periplasmic face (Figure 3.17). It is noteworthy that Lp4 forms helix α1 and Lp8 forms helix α2 located in the interior of the barrel. The Lp4 located between strands β7 and β8, and another is Lp8 located between strands β15 and β16 (Figure 3.17- 3.18). The two loops are made from residues V334-Y354 and V519-S556 respectively (Figure 3.18). The LptD/E complex structure indicates that Lp4 and Lp8 play an important role in LptD and LptE interaction.

The structure is consistent with a previous report that LptD/E from *E. coli* formed a very stable 1:1 complex (Wu et al. 2006; Chimalakonda et al. 2011; Chng et al. 2010), even under SDS-PAGE conditions. LptE is essential for the overexpression of LptD and LptD can protect LptE from proteolytic digestion *in vivo* (Chimalakonda et al. 2011; Chng et al. 2010; Freinkman et al. 2010). LptE is mainly located inside the barrel LptD and interacts extensively. The surface area of monomeric of LptE is 9860.8 $Å^2$, but the interface of LptD/E complex is 3195.4 $Å^2$, suggesting that almost one-third of the LptE is involved in interacting with LptD to stabilize the β-barrel.

It is worth noting that a large part of the A87-T95 region of LptE interacts with residues T351 andD352 on Lp4 and Y680 of LptD. The loop residues A87-T95 of LptE stabilize the LptD barrel by interaction with C-terminal residues T771-M786 of LptD (Figure 3.21 A). Furthermore, the LptD/E interactions were enhanced by hydrophobic interaction between residues W21-L23 of LptE with the outer surface of the LptD barrel. The side chain R497 of LptD interacted with residue W21, while residues L23 and side chains of F495 and W444 of LptD form sandwiches

respectively (Figure 3.21 B). LptE was anchored to LptD by a zinc-binding site, which consists of two LptE residues, H22 and S25 and two LptD residues, E496 and D498. The zinc ion was confirmed through crystal fluorescence scanning during data collection at Diamond Light Source, UK. The extensive interaction between LptD and LptE, especially the two interior loops Lp4 and Lp8, and the C-terminal residues of LptD, indicate that LptE plays an essential role in stabilizing and assembling the largest β-barrel of LptD reported to date (H. Dong et al. 2014).

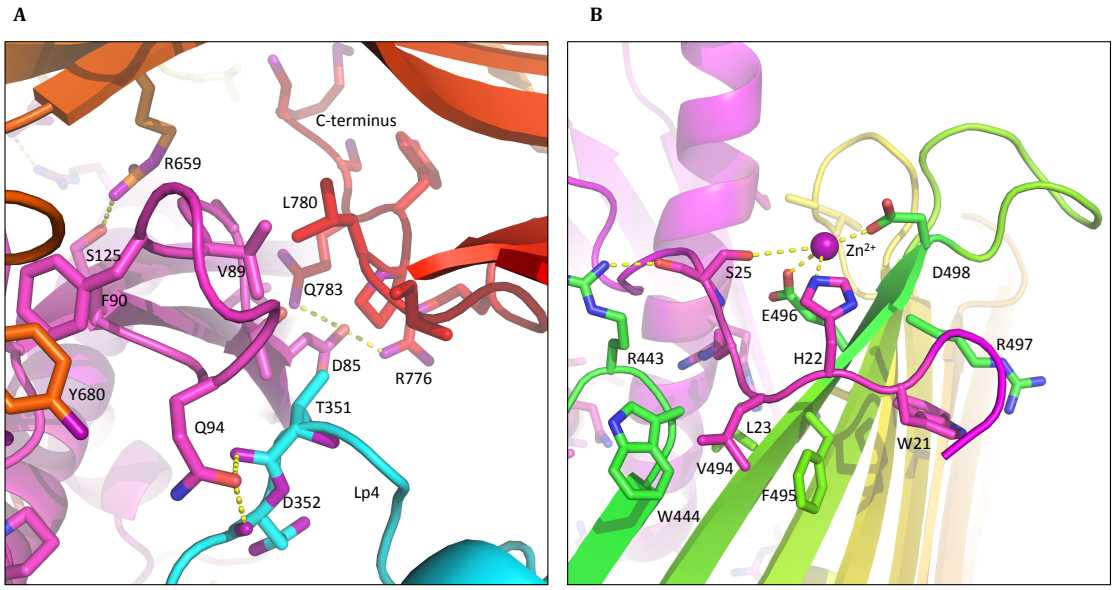


Figure 3.22 LptE interaction with LptD to stabilize the β-barrel. LptE in purple and LptD in rainbow according to the structure of LptD/E described above. A, LptE interacts with Y680, and residues on the Lp4, and C-terminal residues T771-M786 of LptD. B, LptE residues W21-L23 interact with hydrophobic residues at the outer surface of LptD barrel, and zinc ion co-ordinates LptD and LptE interaction (Figure adapted from H. Dong et al. 2014).

3.3.6 Mutagenesis of LptE

In order to investigate further the role of LptE in LptD/E complex formation, single alanine or glycine amino acid substitution and deletion mutations of *lptE* were generated. Depletion of *lptE* in chromosome causes death of *E. coli* cells. Introduction of pBAD plasmid containing LptE with the addition of L-arabinose as a inducer has shown to rescue the *lptE* form depleted *E. coli* cells (AM689) (Sperandeo et al. 2008). Similarly introducing plasmid containing *lptD* into the depleted *E. coli ltpD* cells (AM661) also prevented cell death (Sperandeo et al. 2008).

To confirm whether *lptE* and *lptD* from *salmonella Typhimurium* can replace the function of *E. coli lptE* and *lptD* in the depleted strains, we transformed individual plasmid containing the *salmonella Typhimurium lptE* or *lptD* gene into the *E. coli lptE* depleted strain AM689 or *lptD* depleted strain AM661 respectively. The transformed strain exhibited similar growth to the wild type *E. coli* cells (data not shown), suggesting that *lptE* and *lptD* from *salmonella Typhimurium* can replace the function of *E. coli* proteins.

The mutants with deletions LptE (ΔW21-L23) and LptE (ΔA87-T95) in the LptE depleted strain AM689 and LptD (ΔT771-M786) in the LptD depleted strain AM661 demonstrated greatly impaired cell growth of *E. coli* in LB medium supplemented with 0.5 % SDS and 1mM EDTA (Figure 3.22 A), suggesting that the residues W21-L23 and A87-T95 of LptE and residues T771-M786 of LptD play an important role for LptD/E interactions. In contrast, LptE (ΔT170-N196) does not slow the cell growth (Figure 3.22 A), suggesting that LptE residues T170-N196 is not involved in the interaction with LptD. LPS is constantly transported to the outer membrane to protect the bacterial cell membrane integrity from being compromised. SDS was used

as a detergent in the functional assay, which could potentially break lipid membrane to enter the bacterial cell. EDTA was used to decrease divalent cations in the OM, which is able to help SDS to enter into the bacteria. These results suggest the mutants may cause poor plugging of LptE into the LptD barrel or poor LptD assembly, which result in increasing the outer membrane permeability to SDS.

We also tested the expression level of the proteins (both wild-type and mutants) in *E. coli lptD* or *lptE* depleted cells (AM689 and AM661) by western blots. The results showed that the protein expression levels of the mutants were at similar levels to the wild-type proteins (Figure 3.22 B, C). These results are consistent with that previously reported (Chimalakonda et al. 2011; Freinkman, Chng, and Kahne 2011; Grabowicz et al. 2013; Ruiz et al. 2005).

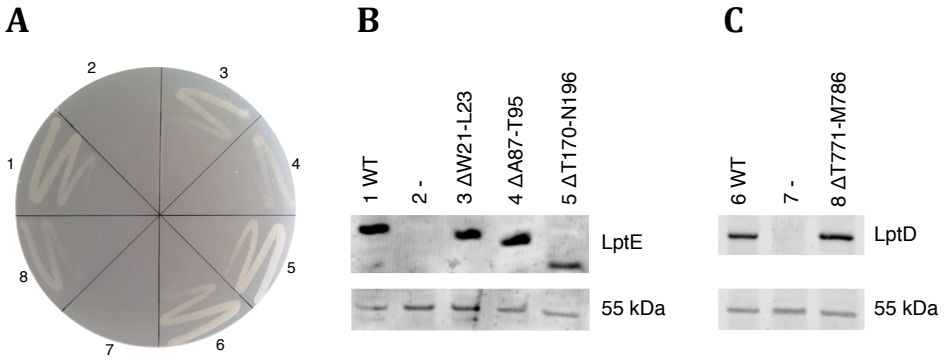


Figure 3.23 Deletion mutation of *lptE* and *lptD* in depleted strain AM689 and AM661 respectively. **A,** Functional assay on the these mutations greatly impair cell growth in LB medium supplemented with 0.5% SDS and 1mM EDTA, with the exception of LptE(ΔT170-N196) which was not involved in interaction with LptD. Segments 1-5 are AM689 cell with wild-type (WT) LptE, the empty plasmid pACYCDuet-1 as control, LptE (ΔW21-L23), LptE (ΔA87-T95) and LptE (ΔT170-N196) respectively, while segments 6-8 are AM661 cell with wild-type *lptD*, the empty plasmid pACYCDuet-1 and LptD (ΔT771-M786). B, Wild-type and differnet mutants were expressed in *E. coli lptD* or *lptE* depleted cells and detected by anti-His-tag antibody. The western blots results showed that the protein expression levels of wild-type LptE and the mutants are comparable on the cell membrane except the LptE (ΔT170-N196). Unidentified protein 55 KDa was used in the experiment as a loading control (Ruiz et al. 2010), 2- is empty plasmid pACYCDuet-1 as a control. C, The deletion mutation LptD (ΔT771-M786) protein expression level are the same as the wild-type LptD, 7- is empty plasmid pACYCDuet-1 as a control.

3.3.7 Structure of LptD indicates that lateral opening is required for LPS insertion

LptD forms a 26-stranded β-barrel with a kidney shape. It has a pore with dimensions of 70 Å by 50 Å (out side) and 50 Å by 30 Å (inner side) at its widest point (Figure 3.23 A, B). However, most of the pore of LptD was sealed by the extracellular loops, just leaving a hole at one side of approximately 15 Å by 10 Å in diameter. The hole is occluded by LptE to close the channel completely (Figure 3.23 A, B).

The lumen of the LptD barrel is very hydrophilic as are those of Wza and AlgE proteins, which transport polysaccharide to the OM (Dong et al. 2006; Whitney et al. 2011). It is therefore a challenge for LptD/E complex to transport the hydrophobic lipid A of the LPS molecule across the hydrophilic lumen in the barrel.

LptE specifically binds LPS and help with its delivery across the LptD barrel (Chng et al. 2010). The highly hydrophobic structure of LptA and LptC were reported to bind lipid A of LPS (Suits et al. 2008; Tran et al. 2010; Okuda et al. 2012) and LPS is transported along the hydrophobic cores. The N-terminal domain of LptD is predicted to have similar structures to that of LptA and LptC. Therefore, the LptC, LptA and the N-terminal domain of LptD form a bridge to transport LPS. The N-terminal domain transport LPS along the hydrophobic core to the outer membrane. When observed from the extracellular face, the β-barrel is twisted in an anticlockwise fashion way to the periplasmic side (Figure 3.18). In the structure of LptD, the β1- β4 strands tilt at an angle of around 30° to the plane of the membrane, while the angle of tilt of strand β20- β26 is about 67° showing that β1 and β26 are separated at the periplasmic side (Figure 3.23 C, E). This phenomenon is the similar to OM barrel proteins FadL,

OmpW, OprG and PagP (van den Berg et al. 2004; Hong et al. 2006; Touw et al. 2010; Cuesta-Seijo et al. 2010).

This feature laterally opens the barrel wall for hydrophobic substrate diffusion. It is worth noting that five hydrogen bonds are located on the strands between β1 and β26 of LptD (Figure 3.23 D), suggesting lateral opening of the barrel between the two strands is possible. The recently reported the structure of BamA is composed of 16 β-stranded barrel and the BamA can undergo lateral opening between strands β1 and β16 for insertion of OM protein into the OM (Noinaj et al. 2013). There are eight hydrogen bonds located on the two strands β1 and β16 of BamA.

Functional LptD is in the oxidised form and disulphide bond formation is required to transport and insert LPS into the OM (Ruiz et al. 2010). LptD is oxidised *in vivo* and has at least one correct disulphide bond formed, which helps LPS transport in correct pathway (Ruiz et al. 2010; Chng et al. 2012). There are four cysteine (Cys) residues in the *Salmonella typhimurium* LptD, which is similar to LptD from *E. coli*, with two (Cys31, Cys173) in the N-terminal domain and two (Cys726, Cys727) in the C-terminal domain (Figure 3.25). These cysteine residues form two non-consecutive disulphide bonds.

Residues Cys173 and Cys727 are conserved in Gram-negative bacteria (Narita and Tokuda 2009), suggesting that the second disulphide is essential. It is notable that Cys726 and Cys727 are individually located in the very flexible loop between β-strands β24 and β25 at the periplasmic side (Figure 3.23 C). Residues Cys31 and Cys173 located in the N-terminal domain are connected to strand β1 of LptD (not shown in the solved structure due to cleaved by α-chymotrypsin), which are at perfect positions for disulphide formations.

The LptD/E complex exists in two conformations, the oxidized and reduced form. The oxidized form migrates to a molecular weight of around 130 kDa, and the reduced form migrates to molecular weight of 100 kDa on SDS-PAGE. The protein conformation was significantly changed from reduced form to the oxidised form, which is evident from the protein migration in SDS-PAGE (Ruiz et al. 2010; Chng et al. 2012). The structural conformational changes of LptD may relate to the LPS transport.

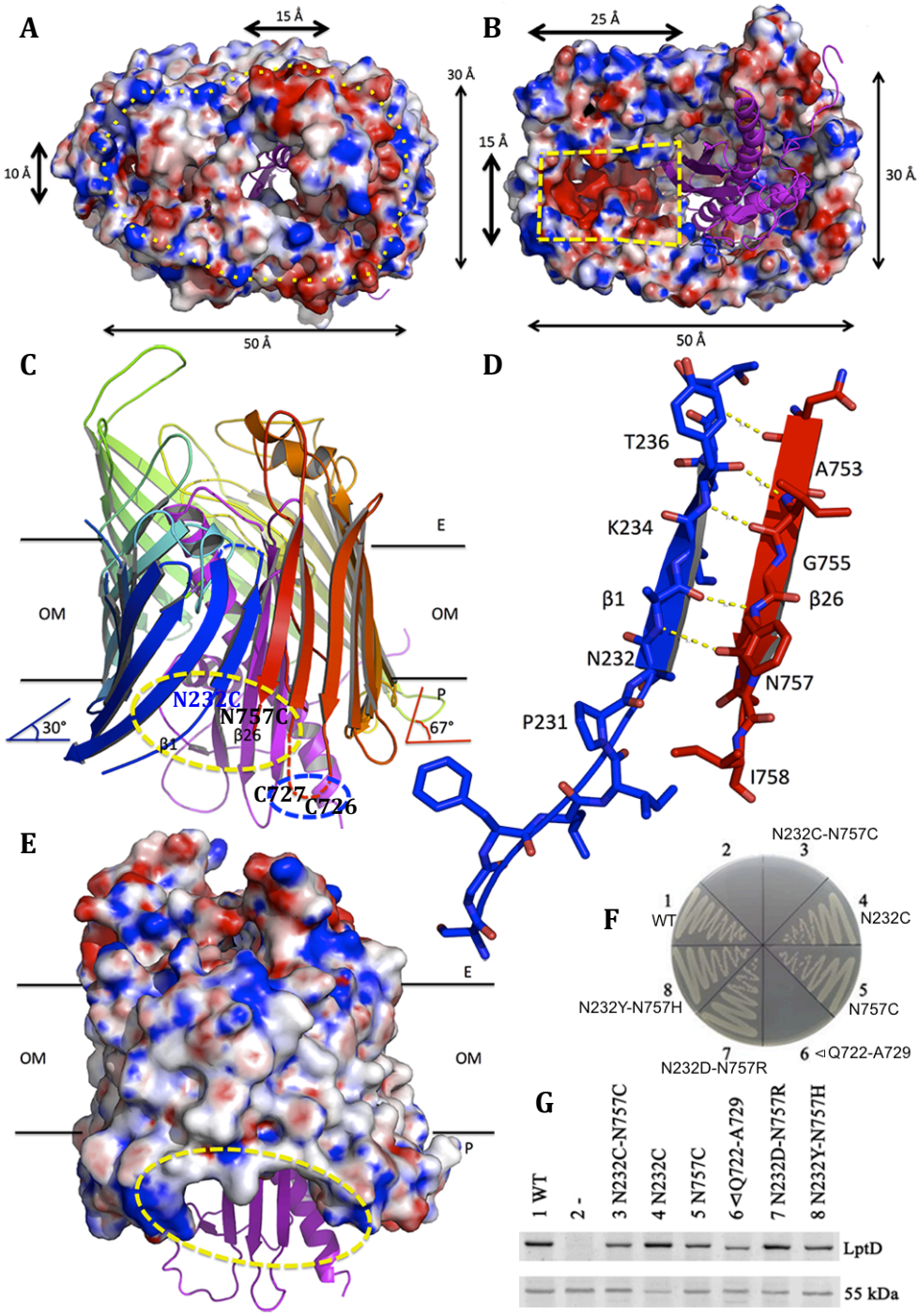


Figure 3.24 The largest barrel and pore of the LptD/E complex. The colour of the cartoon representation is according to the structure of LptD/E described above, where LptD is in rainbow and LptE is in purple. Electrostatic surface potential map of LptD/E, negatively charged residues are shown in red and positively charged residues in blue. A, Top view of the electrostatic surface potential map of the LptD barrel. The pore was shown by yellow dotted line, which is covered by extracellular loops. LptE mainly blocked the pore at one side. B, Bottom view of electrostatic surface potential map of the barrel. A free cavity of LptD is shown in yellow box with a diameter of 25 Å by 15 Å. C, In the structure of LptD, the β1- β4 tilt angle is around 30° to the plane of the membrane, and the tilt angle of strand β20- β26 is about 67°. Strands β1 and β26 are separated at the periplasmic side, which is shown

by a yellow dotted circle. The N232C located in β1 and N757C located in β26 strand may form a disulphide bond in the oxidized environment and lock strands β1 and β26, which are shown in the yellow dotted circle. The residues C726 and C727 are shown in a blue dotted circle. E, P and OM indicate extracellular side, periplasmic side and outer membrane, respectively. D, Five hydrogen bonds are located on the strands between β1 and β26 of LptD. E, Side view of the LptD barrel, which shown the hydrophobic belt of the barrel exterior. The yellow dotted circle shows the separation between strands β1 and β26. F, Function assay analysis showed that strain AM661 shown that two-cysteine mutation N232C/N757C, which potentially locks β1 and β26, resulted the death, indicating that the double mutant N232C/N757C may form a disulphide bond and prevent the lateral opening. Deletion LptD (**Δ**Q722-A729) is lethal of *E. coli* cell. Segment 1-8 represent AM661 cells with the wild-type lptD, the empty pACYCDuet-1, lptD double mutation N232C/N757C, N232C, N757C, deletion Q722–A729, N232D/N757R and N232Y/N757H, respectively. G, LptD and its mutant were expressed in the *E. coli lptD* deletion strain were analyzed by western blot using anti-His-tag antibody. The protein expression of N232C/N757C is similar to that of LptD mutants N757C and N232Y/N757H on the cell membrane. Both single mutations LptD N232C and N757C can grow as efficiently as the wild-type. The unidentified 55-kDa-membrane protein was used in the experiment as loading control 2, Negative control (pACYCDuet-1 plasmid without the lptD gene) (Figure adapted from H. Dong et al. 2014).

3.3.8 Mutagenesis of LptD

To examine whether the lateral opening of LptD is required for LPS transportation, single and double amino acid substitutions and deletions mutants of LptD were generated, then functional assay was carried for these mutations by a colleague Dr. Yinghong Gu (University of East Anglia). The deletion mutation LptD (ΔQ722-A729) removed Cys726 and Cys727, resulting in the death of *E. coli* cells, suggesting that disulphide bond formation (C31-C726, and C173-C727) is important for LPS translocation. The result is in accordance with previously reported (Ruiz et al. 2010; Chng et al. 2012).

The residues N232 and N757 of LptD are located in the strand β1 and β26 separately (Figure 3.23 C, D). The structure suggests two residues may form a disulphide bond in the oxidized condition. The double N232C/N757C mutant is able to form the disulphide bond, locking the two strands β1 and β26 of LptD to prevent any lateral opening. The double mutation N232C/N757C was lethal. In contrast, the single mutation N232C, N757C, and double mutations N232D/N757R and N232Y/N757H of LptD retained the same viability as the wild type (Figure 3.23 F). The protein expression level of the double mutation N232C/N757C is similar to mutations N757C and N232Y/N757H in the membrane, strongly indicating that LPS translocation is required for the lateral opening between strands β1 and β 26 of LptD (Figure 3.23 G).

3.3.9 Molecular dynamic simulations in LptD/E

Molecular dynamics (MD) simulations were performed by our collaborator Dr. Phill Stansfeld (University of Oxford) to further study the stability of the LptD/E complex and LPS translocation paths. MD simulations revealed that the β-barrel of LptD/E may undergo a lateral opening between the strands β1 and β26. Opening of LptD channel was observed when pressures was below the -65 bars (Figure 3.24 A-C). In particular, the molecular dynamics simulations showed that the channel opening and lateral opening occur simultaneously for LPS translocation through the LptD/E machinery.

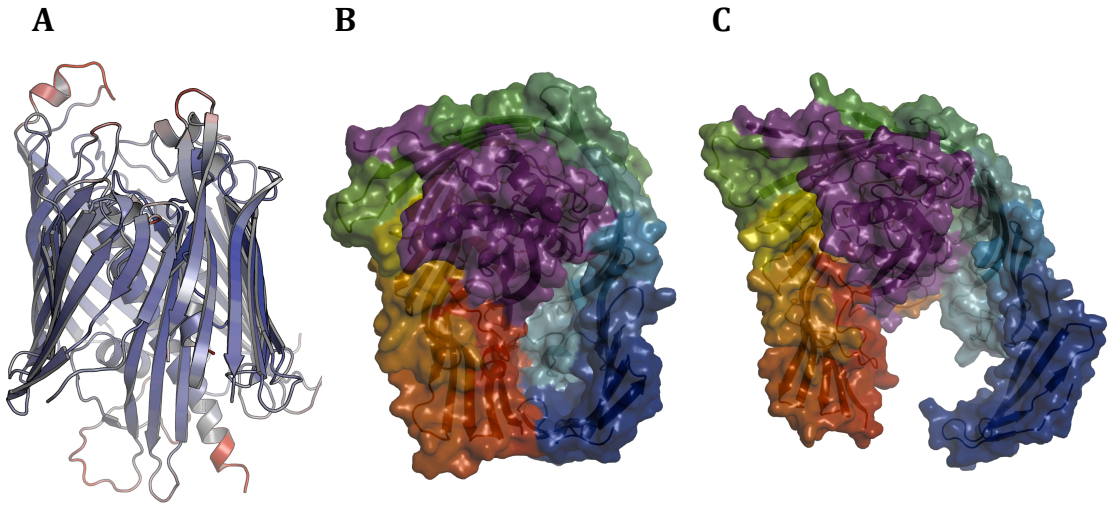


Figure 3.25 Molecular dynamics simulations reveal that a lateral opening in LptD/E. A, Structure of LptD/E complex, the coloured blue represents stability of the LptD/E, and the coloured red represents mobility of domain. B, The LptD barrel is in closed form. The simulations would perform by applying a negative constant pressure to the membrane plane. C, Simulations of LptD/E to translocate LPS. Lateral opening between strands β1 and β26 and the pore opening have been revealed for LPS translocation and insertion (Figure adapted from H. Dong et al. 2014).

3.4 Discussion and conclusion

The lumen of the LptD barrel is very hydrophilic, which is similar to translocases involved in hydrophilic polymer translocation. Wza and AlgE are responsible for translocation of polysaccharides across the outer membrane. Although their structures are different, both of them have hydrophilic lumens for the polysaccharide translocation (Dong et al. 2006; Whitney et al. 2011). The O-antigen of LPS from *Salmonella typhimurium* LT2 is composed of trisaccharide repeat units with short branches of single sugars, and its diameter of the linear polymer is about 13 Å (Figure 3.3). It is worth noting that although LptE occupies part of the LptD barrel, there is free cavity inside LptD with diameter of 25 Å by 15 Å, which is easy to accommodate the O-antigen. The O-antigen is composed of hundreds of saccharide units, and has to pass-through the LptD barrel first, which may use a similar mechanism to AlgE (Whitney et al. 2011). The hydrophilic O-antigen and core oligosaccharide pass-through the inside of the barrel of LptD. How does LptD/E transport the hydrophobic Lipid A of LPS and insert it into the outer leaflet of OM? It is a great challenge to transport the large hydrophobic substrates across the water-filled barrel LptD spanning the OM.

Several other OM proteins use a lateral opening mechanism to transfer hydrophobic molecules through OM (van den Berg et al. 2004; Hong et al. 2006; Van Den Berg 2010; Khan et al. 2009). More recently, the outer membrane protein assembly protein BamA was reported to use the lateral opening between strands β1 and β16 to insert OM proteins into the OM (Noinaj et al. 2013). The structure of LptD/E shows that strands β1 and β26 were separated at the periplasmic side, and between the two strands only five hydrogen bonds exist, which provide strong structural evidence that

the lateral opening is possible. Mutagenesis and function assay revealed that the lateral opening between strands β1 and β26 of LptD is required for the LPS translocation and insertion (Figure 3.23 G). The MD simulations reveal that channel opening and the lateral opening occur simultaneously for LPS translocation and insertion. LPS normally consists of six fatty acyl chains, which are 25 Å in length and 5 Å in width (Figure 3.3). Therefore, there is enough space for the 5-Å sides Lipid A of LPS to pass through an open gate around 7-9 Å between strands β1 and β26. The diameter of the free cavity in LptD is 25 Å by 15 Å (Figure 3.23 B), which may also be involved in the transfer of Lipid A to the gate between the strands β1 and β26. Our new studies have revealed that the LptD/E protein complex forms an intramembrane hole, where lipid A is inserted into the OM (Gu et al. 2015).

Once the non-consecutive disulphide bonds are formed, the LptD N-terminal domain links to LptA and LptC to form a bridge from the IM to the OM in a head to tail fashion (Figure 3.25). The structures of LptC, LptA, and the N-terminal of LptD share a jelly-roll structure (Ruiz et al. 2010). The LptC, LptA and the N-terminal domain of LptD mediate the transport of LPS across the periplasm from the IM to the OM through binding of LPS to the hydrophobic residues inside of the β-jellyroll structures (Suits et al. 2008; Tran et al. 2010). We proposed that the LPS is extracted by LptBFG complex in the IM, whereby LptB provides the energy (Okuda et al. 2012). LPS was then transfered to LptC, LptA, and finally to the N-terminal domain of LptD/E (Sperandeo et al. 2011). The width of *E. coli* periplasm is about 220 ± 20 angstroms, indicating that one LptC, four LptAs and one N-terminal domain of LptD form a rotational slice to transport LPS from the inner membrane to the OM. Our recent studies showed that LptD forms the intramembrane hole for LPS insertion into the OM, while the hydrophilic O-antigen and core oligosaccharide of LPS pass easily

through the barrel of LptD, which trigger the lateral opening between strands β1 and β26, and finally promote the lateral insertion of LPS into the OM (Figure 3.25). As LPS is highly negative charged, it uses divalent cations to form bridges with neighboring LPS molecules. These bridges form part of the OM permeability barrier. We speculate that once the LPS reached the positively charge cation-rich outer leaflet of the OM, the divalent cations mediate electrostatic interaction between LPS core oligosaccharide, which draws LPS to be inserted into the outer leaflet of the OM. However, the mechanism of LptD/E lateral translocation of the LPS is different from that of other lateral transport barrels (van den Berg et al. 2004; Hong et al. 2006; Van Den Berg 2010; Khan and Bishop 2009; Noinaj et al. 2013).

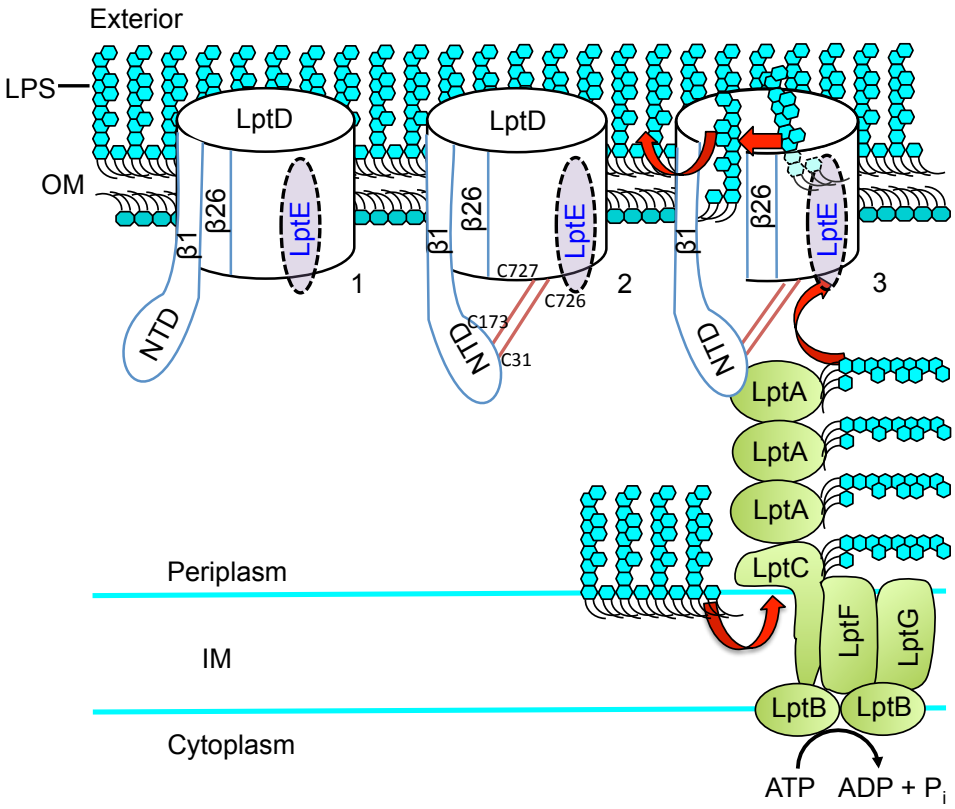


Figure 3.26 The proposed mechanism of LptD/E transport of LPS. 1, The newly synthesis LptD has a closed gate between strands β1 and β26 and in the reduced form, the N-terminal domain (NTD) of LptD is flexible. 2, LptD forms the disulphide bonds (red lines), resulting in N-terminal domain of LptD conformational change. 3, The N-terminal domain of LptD interacts with LptA to form seven protein trans-envelope complex for LPS transport. The LPS molecules are extracted from the inner membrane by LptBCFG and passed to LptC, LptA and the LptD/E complex, while triggering the lateral opening of the LptD for LPS insertion.

LPS is a substantial polymer, which requires two proteins, the LptD/E complex, to transport and insert into the outer leaflet of the OM (Wu et al. 2006; Chimalakonda et al. 2011; Chng et al. 2010; Ruiz et al. 2008). The disulphide bond is required for the function of LptD. It is notable that Cys726 and Cys727 are located in the very flexible loop between strands β24 and β25 at the periplasmic side (Figure 3.23 C), and are at the perfect position to form non-consecutive disulphide bonds (C31-C726, C173-

C727). The lateral opening between the strands β1 and β26 and the non-consecutive disulphide bonds is required to help LPS enter the barrel and insert correctly into the outer leaflet of the OM, rather than the inner leaflet of the OM.

In summary, the crystal structure of the LptD/E translocon described here has numerous unique features. It is the largest known β-barrel with 26-strandeds β-barrel, and has a unique two-protein "barrel and plug" architecture. The structure reveals that LptD and LptE have extensive interactions. The structure, function assay and MD simulations suggest that the LptD β-barrel is responsible for the O-antigen's translocation using the large hydrophilic molecule translocation mechanism, while lipid A is inserted into the outer leaflet of the outer membrane through a lateral opening between the strands β1 and β26 of LptD. This opening is possibly induced by conformational changes in the N-terminal domain and non-consecutive disulphide bond formation within LptD. LPS enters the barrel and inserts correctly into the outer leaflet of the OM, rather than the inner leaflet of the OM. The findings provide a new platform for the study of outer membrane biogenesis and the development of the novel drugs to combat multi-drug resistant Gram-negative bacteria.

The future direction of the research will be the investigation of how the LptBCFG complex extracts LPS from the inner membrane, and whether the LPS insertion into the OM requires energy and development of the novel compounds based on the LptD/E structure.

Chapter 4

Crystal structure of Schmallenberg virus nucleoprotein

4.1 Introduction

Bunyaviridae is a large family of enveloped, single-stranded negative-sense RNA viruses, with more than 350 members across five genera: *Orthobunyavirus*, *Hantavirus*, *Nairovirus*, *Phlebovirus* and *Tospovirus* (Elliott et al. 2013). Several of these viruses are highly pathogenic bunyavirus family members such as Rift Valley fever phlebovirus (RVFV), Sin Nombre Hantavirus, La Crosse, Crimean-Congo hemorrhagic fever viruses (CCHFV) and reassortant Garissa (Elliott 1990; Soldan et al. 2005). *Orthobunyaviruses* cause lethal hemorrhagic fever in humans or animals (Blitvich et al. 2012; B. Hoffmann et al. 2012), while new bunyaviruses cause severe fever with thrombocytopenia syndrome virus in China (Yu et al. 2011) and Schmallenberg virus (SBV) in Europe (Bernd Hoffmann 2012; Beer et al. 2013). Some of the above mentioned viruses have been categorized by the Centre for Disease Control and Prevention as high priority, category A pathogens due to their ability to cause lethal hemorrhagic fever (Barr et al. 2004).

Of these five genera, the *Orthobunyavirus* genus is the largest, currently comprising about 170 membranes. SBV is a negative-sense single-stranded RNA virus. Phylogenetic analyses suggest that SBV belongs to the species *Sathuperi virus* (Garigliany et al. 2012; Goller et al. 2012), which is an *Orthobunyavirus* of the Simbu serotype. SBV is a newly emerging virus, which infects cattle, sheep and goats in farms of Germany, and was first identified in November 2011. Since then the SBV have widely spread across Europe and more than 5000 farms have reported severe symptoms in newly born sheep, goats and cows. This suggested that the virus can pass from an infected mother during pregnancy and affect her offspring and cause

malformations in the new born baby (Beer et al. 2013; Tarlinton et al. 2012; Bernd Hoffmann 2012).

SBV virus is a newly emerging Orthobunyavirus, which was initially found in Germany in November 2011. Since then, it has spread across Europe and caused severe disease in cattle, sheep and goats. Currently, there are no vaccines or effective therapies to combat this viral infection (Bernd Hoffmann 2012). The clinical signs of infected cattle and sheep include fever, reduced milk production, and diarrhoea. SBV can cause stillbirth, as well as birth defects, resulting in considerable economic losses (Garigliany et al. 2012; Tarlinton et al. 2012). There is strong evidence to suggest that they use intermediate midge or mosquito hosts to mediate its transmission (De Regge et al. 2012; Rasmussen et al. 2012). Genetic analysis showed that SBV has gene sequences that are similar to other three Simbu serogroup viruses: Shamonda, Akabane and Aino, which have been isolated from cattle in Japan (Bernd Hoffmann 2012; Goller et al. 2012).

4.1.1 Bunyavirus genome organization

The Bunyamwera virus (BUNV) is the prototypic member of both the Orthobunyavirus genus and the family *Bunyaviridea*. The Bunyaviruses are spherical, and the genome comprises three segments of negative sense RNA, the large segment (L), the medium segment (M) and the small segment (S) (Figure 4.1) (Elliott et al. 2013; Walter et al. 2011; Soldan et al. 2005). The S RNA segment encodes nucleoprotein (NP) and a non-structural protein (NSs) from alternative open reading frames (ORFS). The NP is the most abundant protein in the virus and infected cell and encapsulates viral genomic and antigenomic RNA to form ribonucleoprotein complexes (RNPs), which act as templates for genomic RNA replication and

transcription (Elliott et al. 2013; Walter et al. 2011). The NSs protein acts as an interferon (IFN) antagonist (Bridgen et al. 2001). The M RNA segment encodes a polyprotein that is cleaved to generate two glycoproteins Gn and Gc and a non-structural protein (NSm). Finally, the L RNA segment encodes RNA-dependent RNA polymerase or L protein, which is responsible for all viral RNA and mRNA synthesis. Each genomic segment is complexed within ribonucleocapsid. The three ribonucleocapsids are further enclosed by an envelope composed of a lipid bilayer and are anchored by two glycoprotein (Gn and Gc) (Walter et al. 2011; Elliott et al, 2013; Shi et al. 2010) (Figure 4.1).

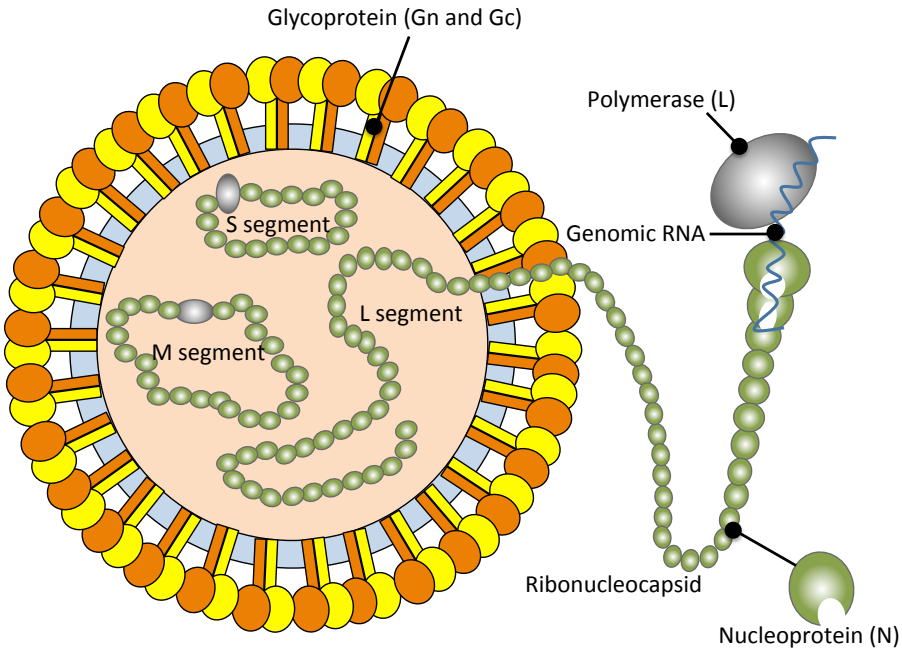


Figure 4.1 Schematic representation of the Bunyavirus virion. The three-genome segments of negative sense RNA (S, M, L) are encapsidated by the nucleoprotein with viral genomic and antiviral genomic RNAs to form ribonucleoprotein complexes.

4.1.2 Crystal structures of RVFV NP

Among buyaviruses, the structure of NP from RVFV has been reported (Raymond et al. 2010; Ferron et al. 2011). RVFV is a prototypic member of the *phlebovirus* genera infects both livestock and humans and is a causal agent of Rift Valley fever (Pepin et al. 2010). The structure of RVFV NP purified under denaturation and refolding condition was monomeric (Figure 4.2 A) (Raymond et al. 2010). The monomeric RVFV NP is a novel compact all-helical folded structure, and the structure lacks a positively charged crevice for RNA binding and has no protruding terminal domain or loops for NP oligomer or RNP formation. Another group, Ferron *et al.* purified the RVFV NP under non-denaturation conditions and its crystal structure was solved as a hexamer with a highly positively charged region in the inner part of the ring for the RNA accommodation (Figure 4.2 B) (Ferron et al. 2011).

Although both of them are RVFV NP structures, the conformations are significantly different with regard to the position of the N-terminal arm. In the monomeric structure of RVFV NP, the N-terminal arm packs closely against the core domain (Figure 4.2 A) (Raymond et al. 2010), while the N-terminal arm extends away and interacts with adjacent subunit in the hexamer (Figure 4.2 B), suggesting that the N-terminal arm is essential for the oligomerization of NP (Ferron et al. 2011).

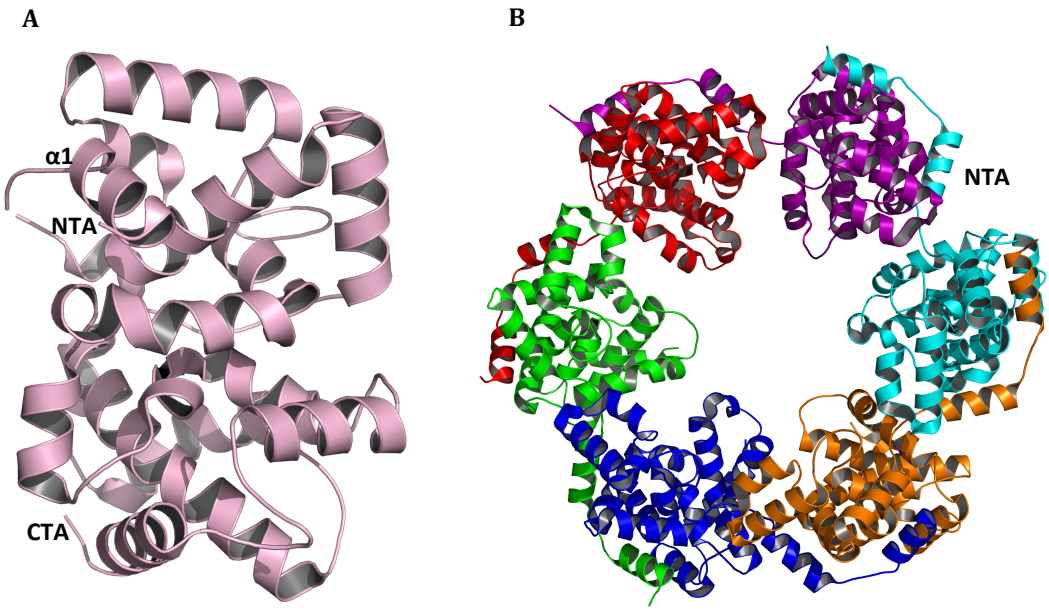


Figure 4.2 Structures of RVFV NP. A, Crystal structure of monomeric RNFV NP. In the monomeric structure, the N-terminal arm packs closely against the core domain. The NTA and CTA represent N-terminal arm and C-terminal arm, respectively (PDB code: 3LYF). B, Structure of the RVFV NP as a hexameric ring-shape structure. The N-terminal arm is essential for the oligomerization of the NP and extends out to interact with the adjacent subunit in the crystallographic hexamer. Hexameric subunits are coloured in red, green, blue, orange, cyan and purple, respectively (PDB code: 3OV9).

4.1.3 Crystal structure of CCHFV NP

Following crystal structure studies on the RVFV NP within *phlebovirus*, the structure of CCHFV NP from stains YL04057 was determined (Guo et al. 2012). CCHFV NP belongs to the *Nairovirus* genera within the Bunyaviridea family and is the causative agent of severe hemorrhagic fever with high mortality in humans (Vorou et al. 2007). The monomeric structure of CCHFV NP possesses a racket-shape with two parts: a "head" domain and "stalk" domain. The structure has unexpected endonuclease activity and the positive charge residues in the "head" and "stalk" were suggested to be responsible for RNA binding (Guo et al. 2012) (Figure 4.3).

Subsequently, Carter *et al*. reported the structure of CCHFV NP from strain Baghdad-12. This structure was superimposed on that of CCHFV NP from strain YL04057, and showed that "head" domain is very similar, but there were differences in the transposition of the "stalk" domain through a rotation of 180 degrees and a translation of 40 Å, suggesting the structural flexibility of switching between alternative NP conformations during RNA binding and oligomerization (Carter et al. 2012).

Furthermore, Wang *et al.* solved the structure of the CCHFV NP form strain IbAr 10200. The structure was shown to have two conformations. When incubated with single stranded RNAs, the structure conformation is monomeric and is identical to the CCHFV NP from strain YL04057. In contrast, when the protein was purified under native condition (that using second peak devoid of nucleic acids), the structural conformation was oligomeric and comprised double antiparallel superhelices (Y. Wang et al. 2012).

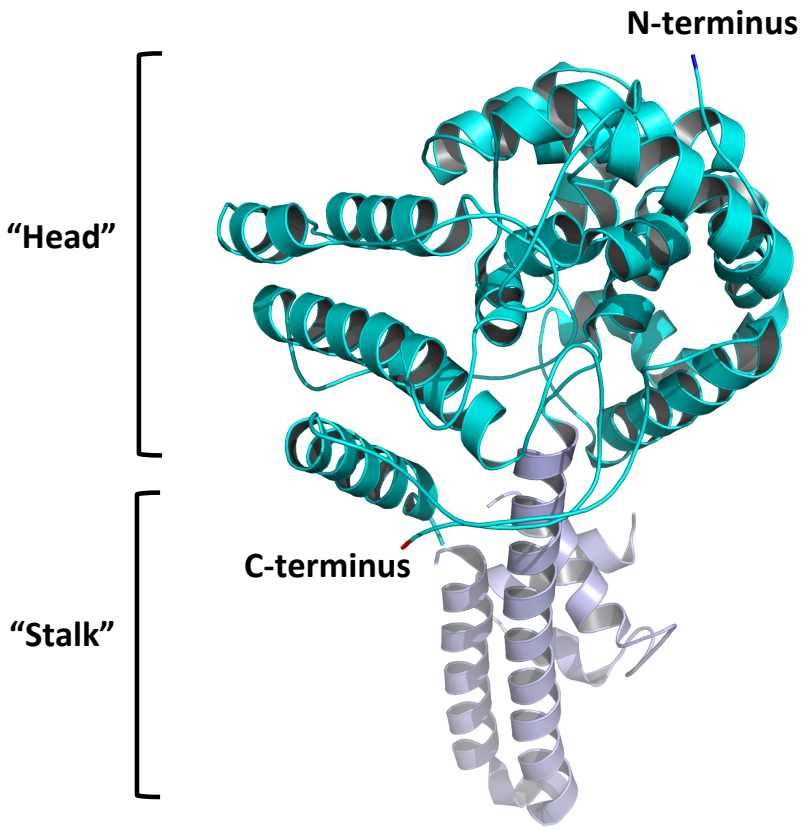


Figure 4.3 Crystal structure of CCHFV NP. The NP from strain YL04057 in cartoon representation. Head and stalk domain are colored in cyan and light blue, respectively (PDB code: 3U3I).

4.1.4 Newly emerging bunyavirus

Although the structure of NP (RVFV, CCHFV) from *Phlebovirus* and *Nairovirus* genera within Bunyaviridae family were reported by different research groups independently, the mechanism by which they bind genomic RNAs are significantly different (Raymond et al. 2010; Ferron et al. 2011; Guo et al. 2012; Carter et al. 2012; Y. Wang et al. 2012). SBV is a newly emerging bunyavirus, which has spread rapidly across European, causing congenital abnormalities in the offspring of cattles, sheeps and goats. SBV belongs to Simbu serogroup of Orthobunyavirus genera within Bunyaviridea family. The SBV NP may have a novel fold and a unique mechanism for RNA recognition and encapsidation, and therefore we decided to determine the nucleoprotein structure (Bernd Hoffmann 2012; Beer et al. 2013).

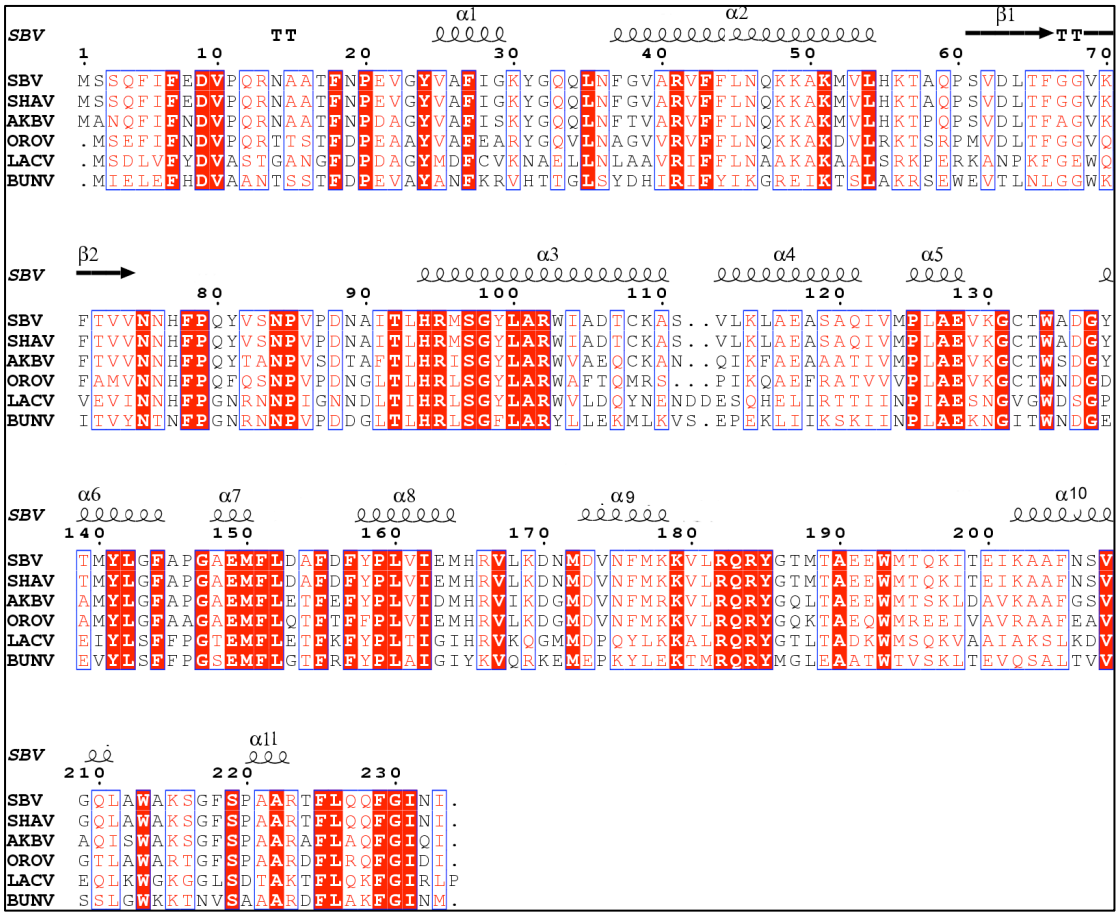


Figure 4.4 Amino acid seqence aligment of orthobunyavirus NPs. The amino acid sequence of NPs from selected orthobunyaviruses genus share high level of amino acid similarity. The SBV, SHAV, AKBV, OROV, LACV and BUNV represent Schmallenberg virus (sequence access number CCF55031), Shamonda virus (YP_006590077), Akabane virus (YP_001497161), Oropouche virus (NP_982305), La Crosse virus L78 (Q8JPR0) and Bunyamwera virus (NP_047213), respectively. The predicted secondary structures above the sequencs were based on the SBV N structure. Conserved residues are colored in red.

4.2 Aims

The SBV NP shares a high level of amino acid identity with other Orthobunyavirus NP (Figure 4.4), representing a novel model for a subfamily of the diverse bunyaviruses and can serve as a model for all members of the *Orthobunyavirus* genus. In order to study of the function and mechanism of SBV NP, the crystal structure was sought. The protein will purify under native condition, as well as denaturation and refolding condition to remove the host RNAs from the SBV NP.

4.3 Results

4.3.1 Crystallization of native SBV NP

The protein was successfully expressed in Rosetta cells. We can routinely obtain 12 mg of protein from 10 litre cell cultures, which is sufficient for crystallization and optimisation.

The native SBV NP was further purified using size exclusion chromatography with an Äkta Xpress and shows that N protein in oligomeric state in solution (Figure 4.5 A). Two peaks appeared and the NP purity was checked using SDS-PAGE. Only the purest fractions were pooled and concentrated to 10.4 mg ml^{-1} (Peak 1) and 12.6 mg ml^{-1} (Peak 2) (Figure 4.5 B) (Vivaspin 20, cut off 30,000 kDa MWCO, Sartorius stedim biotech), which were measured by a Nanodrop. The protein oligomers were further examined by cross-linking under reduced (Figure 4.5 C) and non-reducing conditions (Figure 4.5 D), which indicated that protein exists as a tetramer. This cross-linking experiments were performed by Dr Ping Li (University of Glasgow). The crystallization trails were described as in Chapter 1.2.6.1. The protein was flash frozen in liquid nitrogen and stored at -80°C.

After one week, protein crystals grew from the peak 2, but not from the Peak 1, due to the Peak 1 presents as an oligomerises in gel filtration. The crystals were obtained in different crystallizations conditions, with the best crystals growing from 1.4 M Sodium/Potassium phosphate pH 5.6 (Figure 4.6 A).

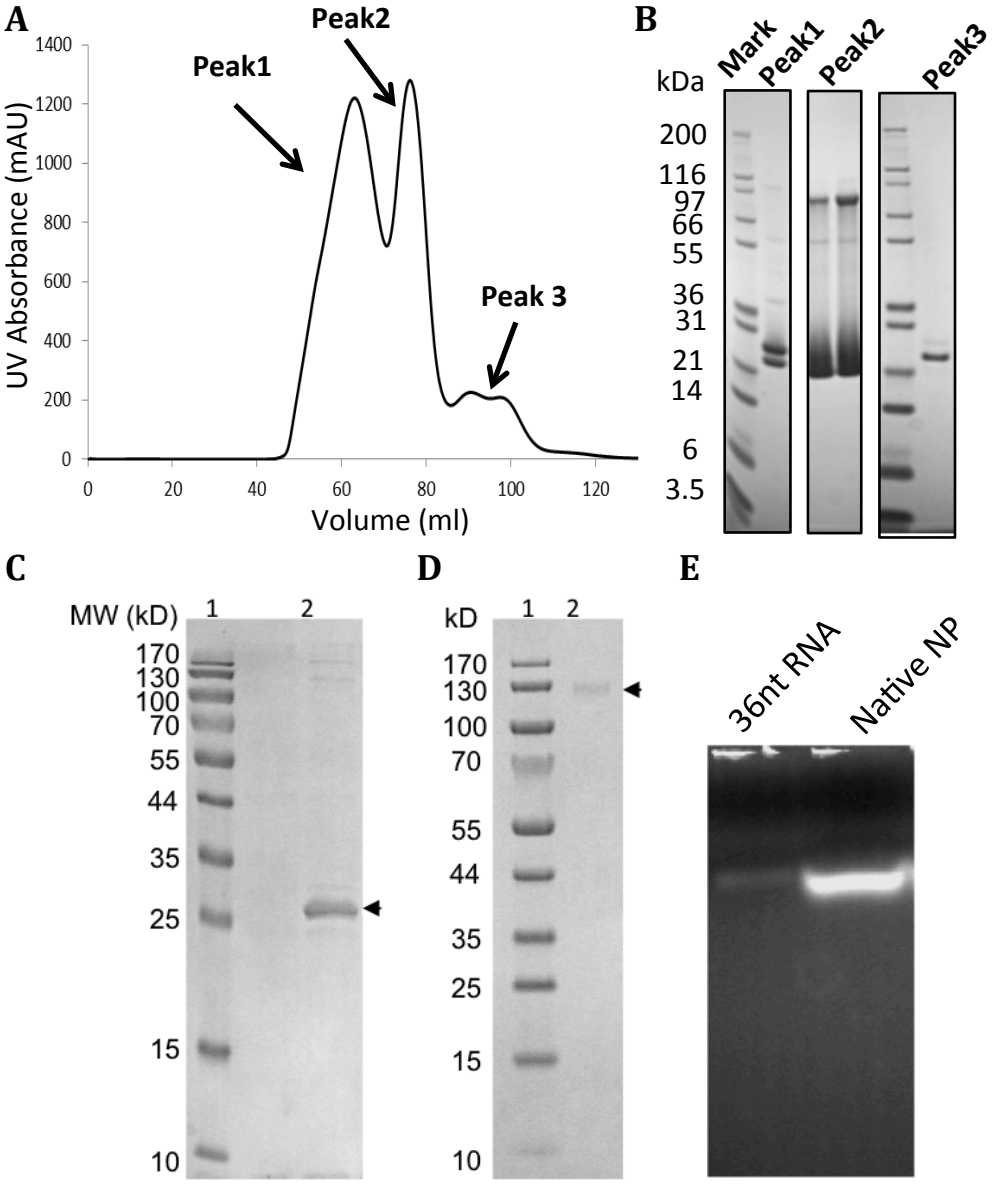


Figure 4.5 Native SBV NP purification. A, Gel-filtration of SBV NP. B, The first peak contains similar amounts of 6His-tag intact (upper band) and cleaved SBV NP (lower band). The second peak contains mostly 6His-tag cleaved SBV NP. The upper bands are oligomeric state of the SBV NP. The band of SBV NP near the 97 kDa molecular weight marker represents the tetrameric form, which has been verified by cross-linked (D) and mass spectrometry (University of St Andrews). The third peak contains small amounts of 6His-tag intact and cleaved SBV NP. C, D, The protein was cross-linked, under reducing (C) and non-reducing (D) conditions and NP shows by arrows. E, Native SBV NP was complex with bacteria RNA, 36nt RNAs was used as control, indicates protein bound RNA, about 30-40 nucleotides (nt) in length.

Crystal optimisation proceeded by varying the pH from 5.2 to 6.2 and the sodium/potassium concentration from 1.0 to 2.0 M. The crystal optimisation was performed using 96 well crystal clear sitting–drop plates by mixing 2 µl of protein and 2 µl of crystallization precipitant with 100 µl reservoirs solution (Figure 2.5 B). The optimised crystallization plates were incubated at room temperature (20°C).

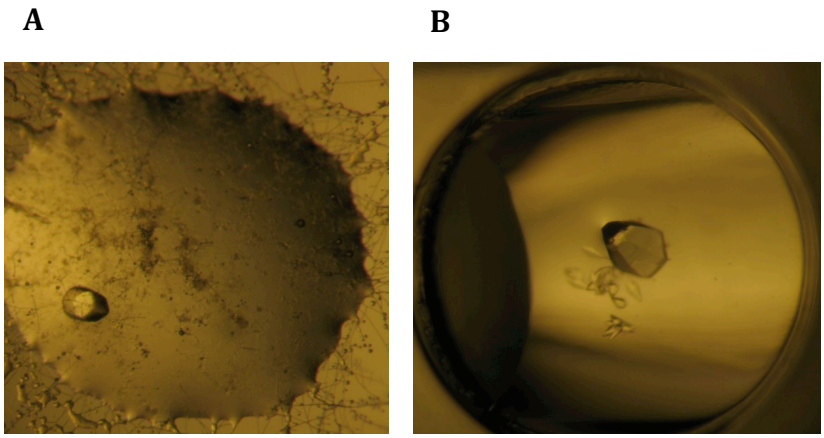


Figure 4.6 Photographs of nucleoprotein crystals of SBV. A, The original crystal from screening. B, The crystal from optimisation. The crystals grew in 1.4 M Sodium/potassium phosphate pH 5.8 at 20°C after 7 days.

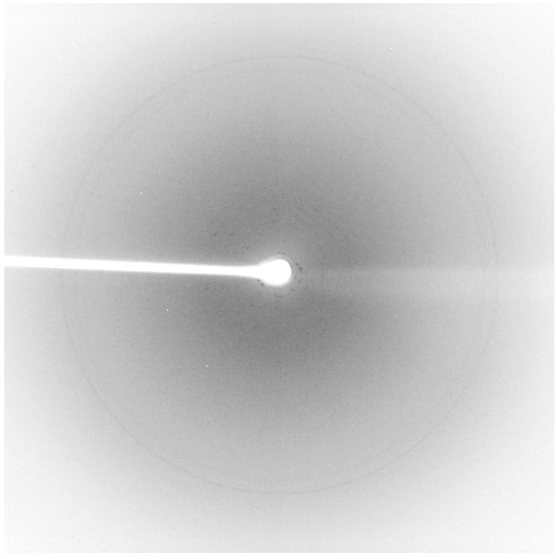


Figure 4.7 The native SBV NP crystal diffraction pattern. The image on the left shows the spots to be very weak with 5.7-Å resolution from our in-house X-ray source.

The crystals were large and fragile, but the resulting diffraction was poor. We assumed that the cryoprotectant (containing 20% glycerol) might damage the crystals. Therefore, we tried to change crystallization conditions by introducing (2 - 26%) ethylene glycol and also varying pH (5.2 - 6.2). Ethylene glycol is an odourless and colourless liquid and is commonly referred to as an antifreeze. The crystals grew well with up to 20% ethylene glycol in the crystallization solution, but no crystals were obtained with more than 20% ethylene glycol. Crystals grown in more than 10% ethylene glycol do not need additional cryoprotectant. Unfortunately, crystal diffraction quality did not improve (Figure 4.8), even using the Diamond Synchrotron Light source.

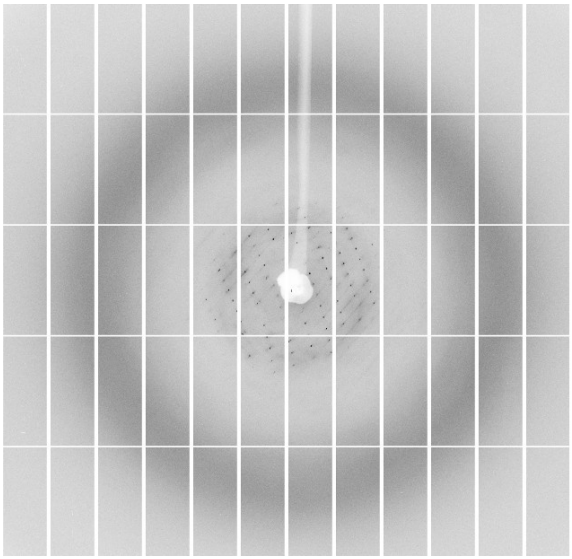


Figure 4.8 The crystal diffraction image was collected using I02 beamline at Diamond Light Source, UK. There was no significant improvement in the diffraction using 2-20% ethylene glycol, rather than 20% glycerol as cryopretectant. This indicates that ethylene glycol did not help in improving the crystal X-ray diffraction.

As the nucleoprotein's function is to form a ribonucleoprotein complex and because many purified viral nucleoproteins have been shown to contain RNAs from expression hosts (Raymond et al. 2010; Ferron et al. 2011; Guo et al. 2012; Carter et al. 2012; Wang et al. 2012), pertinent questions one must ask are: 'Does purified SBV nucleoprotein contain *E. coli* RNAs?' and 'Are the SBV nucleoprotein crystals complexed with RNA?' These are serious questions, which are well worth considering.

To test this, we first examined the absorbance at A260/280 nm and derived an absorbance of 1.41 using a Nanodrop (Thermo Scientific), thus indicating that the protein is complexed with *E. coli* RNAs.

Then we performed electrophoresis of the crystals using a 1% agarose gel. To prevent RNA contamination from the crystallization drop, crystals were picked up before being washed four times using the crystallization solution. The running buffer was 1X TAE and the electrophoresis was carried out at 100 V for 30 mins. The gel was stained with 0.01% ethidium bromide and the RNAs were visualised using a UV transilluminator. Indeed, both purified SBV protein and crystals contained RNA (Figure 4.9). RNAs transcribed from the *E. coli* genome are essentially random, varying in sequence and length. Therefore, this may explain why crystal diffraction quality was consistently not good enough for data collection because the oligomer contained *E. coli* RNA in their heterogeneous nature.

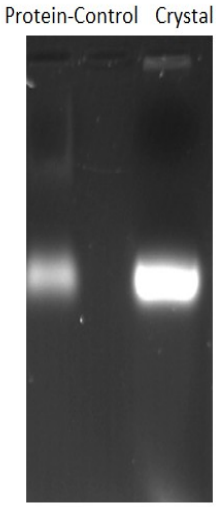


Figure 4.9 Both purified SBV protein and crystals are complexed with RNAs. The agarose gel shows that the RNA from crystal is much stronger than that from protein. The crystal was washed in 1.4 M Sodium/potassium phosphate crystallization solution four times to avoid contamination of protein from the crystallization drop, before the crystal was dissolved in water.

Although the native SBV NP crystals were large from second peak, the crystal diffraction was poor. The reason of the poor crystals is that the SBV NP bound heterogonous RNA from the *E. coli* (Figure 4.9), and further experiments have identified that native SBV NP can bind about 30-40 nucleotides in length (Figure 4,5 E).

In order to remove host RNAs from the purified SBV protein, RNase A was added to the protein purification process. We hoped that the RNase could degrade host RNAs from the NP and improve the crystal diffraction quality to allow us to collect high-resolution data for structural determination.

4.3.2 Crystallization of RNase A treated native SBV NP

The protein purification was described in section 1.2.5.1, and the RNAse A and contaminating proteins were removed by further purification with the gel-filtration column (Figure 4.10 A), showed two peaks. The NP purity was checked using SDS-PAGE (Figure 4.10 B). Only the purest fractions were collected and concentrated to 7 mg ml^{-1} (Peak1) and 12.2 mg ml^{-1} (Peak2) respectively. The crystallization trails were described as in Chapter 1.2.6.2.

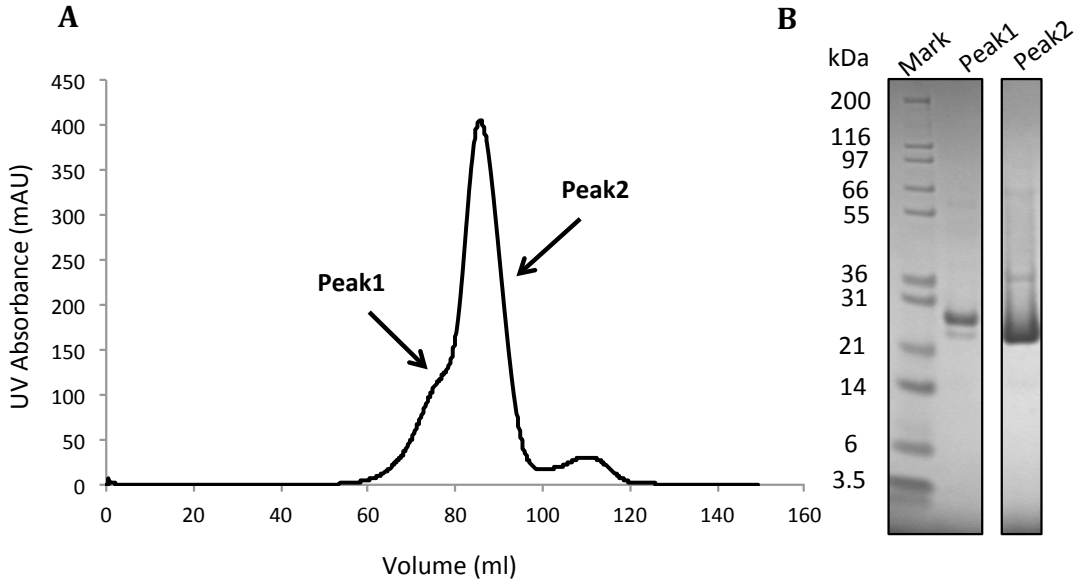


Figure 4.10 Purification of SBV NP treated with RNase A. A, Gel-filtration of SBV NP treated with RNase A. B, SDS-PAGE analysis of the protein purity, the first peak contains similar amounts of 6His-tag intact (upper band) and cleaved SBV NP (lower band). The second main peak contains mostly 6His-tag cleaved SBV NP.

Crystals from peak2 were obtained from 0.2 M sodium formate and 20% PEG 3350. The crystals were optimized by varying precipitant: (0.05- 0.4 M) sodium formate, and (16- 24 %) PEG 3350. Optimisation was carried out by building crystallization drop of 1 µl protein and 1 µl crystallization solution using the sitting- drop vapour diffusion method (Figure 4.11). The crystals were harvested after 10 days, and protected in a cryoprotectant containing 20% glycerol in liquid nitrogen for data collection.

4.3.2.1 Data collection of RNase A treated SBV NP crystals

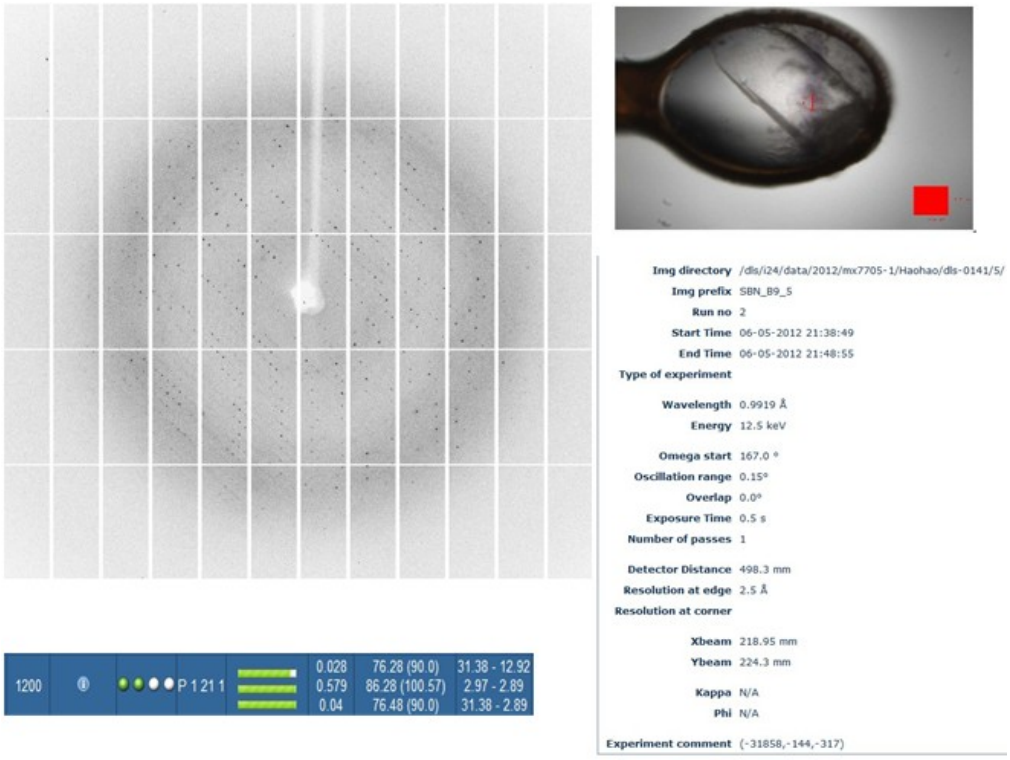


Figure 4.11 X-ray diffraction of SBV NP crystal using beamline I24 of Diamond Light Source, UK. The left image is a diffraction pattern, the right upper image is of the crystal used for data collection and the right bottom table summarises the data collection statistics.

Data collection was described as in chapter 1.2.6.6. The data were indexed and integrated using iMosflm (Battye et al. 2011). Pointless (Winn et al. 2011) suggested the space group is $P2_1$. The data was scaled with Scala (CCP4 suite) (Evans 2006; Winn 2003). A summary of the data collection statistics is listed below, in table 4.1.

Resolution (Å)	75.09 - 2.95 (3.05 - 2.95)
Wavelength (Å)	0.9919
Completeness (%)	98.00 (98.62)
I/σ	16.16 (2.56)
Multiplicity	3.2 (1.7)
Unit Cell (Å)	a = 76.2, b = 86.2, c = 76.4 α = γ = 90°, β = 100.57°
Unique reflections	20602 (2069)
R_{merge} (%)	4 (57.9)
Space Group	$P2_1$

Table 4.1 Native SBV NP data collection statistics

4.3.2.2 Molecular replacement failed

We tried to determine the SBV NP structure of the above dataset using molecular replacement using Phaser in CCP4 (McCoy et al. 2007) with Rift Valley Fever virus NP (PDB code 3OV9) as a search model. However, we could not solve the structure with molecular replacement using either Phaser or Molrep in CCP4 or Phenix (Zwart et al. 2008). This is because of only 12% sequence identity between the SBV NP and the model (Figure 4.12).

```
RVFV.          MDNYQELAIQFAAQAVDRNEIEQWVREFAYQGFDARRVIELLKQYGGADWEKDAKKMIVL
SBV.           ------MSSQFIFEDVPQRNAATFNPEVGYVAFIGKYGQQLN--FGVARVFFLNQKKAKM
                     :: **  : * :.:   :  *..* .* .:   :*   :* *      :*   :

RVFV.          ALTRGNKPRRMMMKMSKEGKATVEALINKYKLKEGNPSRD-ELTLSRVAAALAGRTCQAL
SBV.           VLHKTAQPSVDLTFGG----VKFTVVNNHFPQYVSNPVPDNAITLHRMSGYLARWIADTC
               .* :  :*   :   .    ... .: *::    .**  *  :** *::. **   .::

RVFV.          VVLSEWLPVTGTTMDGLSPAYPRHMMHPSFAGMVDPSLPGDYLRAILDAHSLYLLQFSRV
SBV.           KANVLKLAEASAQIVMPLAEVKGCTWADGYT-MYLGFAPGAEMFLDAFDFYPLVIEMHRV
                .    *. :.: :    .         .:: *     **  :      .   :::: **

RVFV.          INPNLRGRTKEEVAATFTQPMNAAVNSNFISHEKRREFLKAFGLVDSNGKPSAAVMAAAQ
SBV.           LKDNMDVNFMKKVLRQRYGTMTAEEWMTQKITEIKAAFNSVGQLAWAKSGFSPAARTFLQ
               :: *:  .  ::*      .*.*    .    * :  * ..  *. ::.  *.*. :  *

RVFV.          AYKTAA
SBV.           QFGINI
                :
```

Figure 4.12 Protein sequence alignment of RVFV and SBV NP. The NP conserved amino acids from RNFV and SBV are shown with asterisks (*).

As we could not determine the SBV NP structure by molecular replacement, we decided to make Selenomethionine (SeMet) labelled protein (Bakke et al. 2010; Boles et al. 1991) and determine the structure by single-wavelength anomalous dispersion (SAD) and multi-wavelength anomalous dispersion (MAD) (Hendrickson, Horton, and LeMaster 1990; Hendrickson 1991).

4.3.3 Crystallization of RNase A treated SeMet SBV NP

The size exclusion chromatography pattern of the SeMet labelled SBV NP behaved similarly as the native SBV NP treated with RNase A. Therefore, I harvested the second peak of the pure protein, which was concentrated to 12.6 mg ml^{-1}. The crystallization trials were setup using sitting-drop vapour diffusion, based on the original SBV NP crystallization condition: 0.2 M sodium formate, 20% PEG3350. To obtain larger crystals, extensive optimisations were carried out.

4.3.3.1 Data collection of RNase A treated SeMet labelled SBV NP crystals and structure determination

Data collection was described as in chapter 1.2.6.7. The data were indexed and integrated using iMosflm (Battye et al. 2011) and scaled using Scala (Evans 2006). The crystal belongs to the space group of $P2_1$. The data collection statistics are listed below in Table 4.2.

Data collection	Peak	Inflection	Remote
Wavelength (Å)	0.9797	0.9799	0.9218
Resolution (Å)	29.82-3.1 (3.19-3.11)	29.61-3.22 (3.30-3.22)	29.71-3.08 (3.16-3.08)
Space group	-	-	$P2_1$
Completeness (%)	98.4 (80.3)	99.1 (89.7)	99.4 (93.8)
Anomalous completeness (%)	97.30 (67.80)	98.70 (84.40)	99.00 (89.00)
I/ **σ**	24.0 (2.7)	17.0 (2.9)	18.6 (2.6)
Unit cell a (Å)	76.5	76.2	76.3
Unit cell b (Å)	86.7	86.4	86.4
Unit cell c (Å)	77.7	77.5	77.4
Unit cell α (°)	90	90	90
Unit cell β (°)	101.27	101.22	101.26
Unit cell γ (°)	90	90	90
Unique reflection	17,901 (1070)	16,103 (1067)	18,405 (1260)
Anomalous completeness	97.30 (67.8)	98.7 (84.4)	99.0 (89.0)
R_{merge} (%)	11.20 (86.20)	8.80 (65.20)	9.30 (77.30)
Average redundancy	20.50 (13.10)	7.50 (6.60)	7.50 (6.90)
Refinement			
R_{factor}	-	-	0.32
R_{free}	-	-	0.36
RMSD bonds (Å)/ angles (°)	-	-	0.007/1.257
PDB code	-	-	4IDU

Table 4.2 SBV NP MAD data collection statistics. Values in parentheses are represents for the highest-resolution shell. $\mathbf{R_{factor}= \Sigma \; || \; F_{obs}| - | \; F_{cal}||/ \; \Sigma \; |F_{obs}|}$, where F_{obs} and F_{cal} are observed all reflection measured and calculated currently model as structure factors, respectively. R_{free} is calculated using 5% of total reflections, which is randomly selected not used in refinement.

Unmerged data from SCALA was used to solve the phase problem using the SHELX suite (Sheldrick 2007; Sheldrick 2010). The best correlation coefficient was 58.87 using SHELXD (Schneider and Sheldrick 2002), indicating that an initial set of phases for the structure had been found (Figure 4.13 A). The positions of 29 Se sites were determined using SHELXD (Figure 4.13 B), and the contrast and connectivity suggested that the inverted, rather than the original, provided the correct phases using SHELXE (Figure 4.14). The connection contrast of the inverted and original line was separated very well, suggesting that the crystal structure of SBV NP was determined

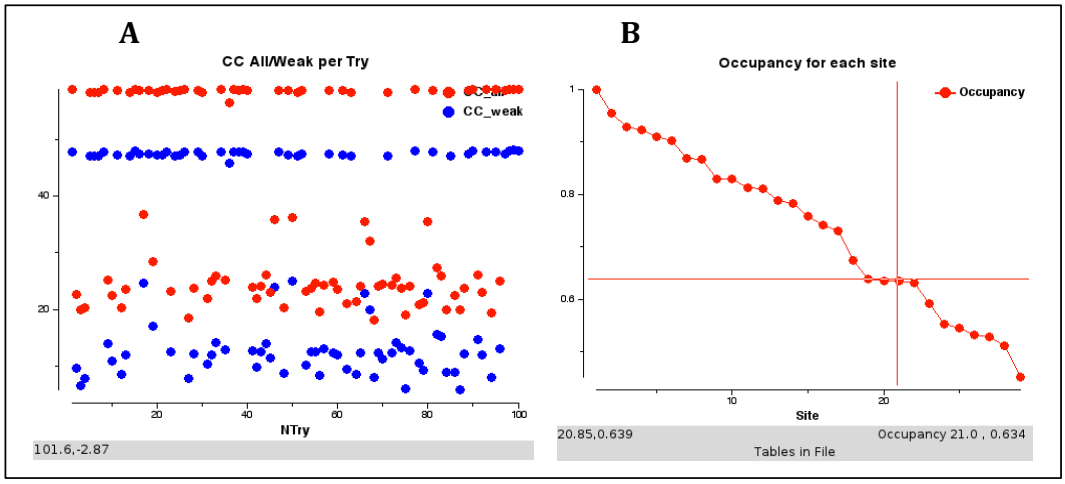


Figure 4.13 Analysis of the heavy atom substructure of SBV NP using SHELX. A, The SBV NP SeMet correlation-coefficient (CC) values at 58.87 from SHELXD, showing phase of correct solution. B, The heavy atom of SeMet labeled sites in SBV NP were determined using SHELXD.

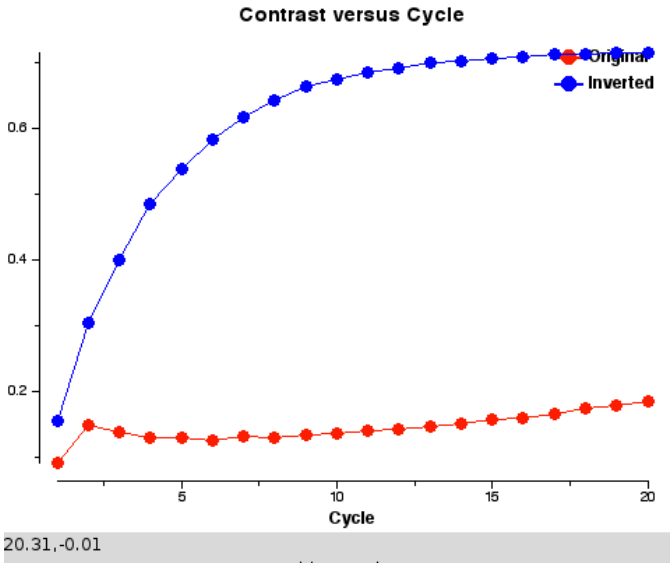


Figure 4.14 SHELXE suggested the phases to be inverted hand.

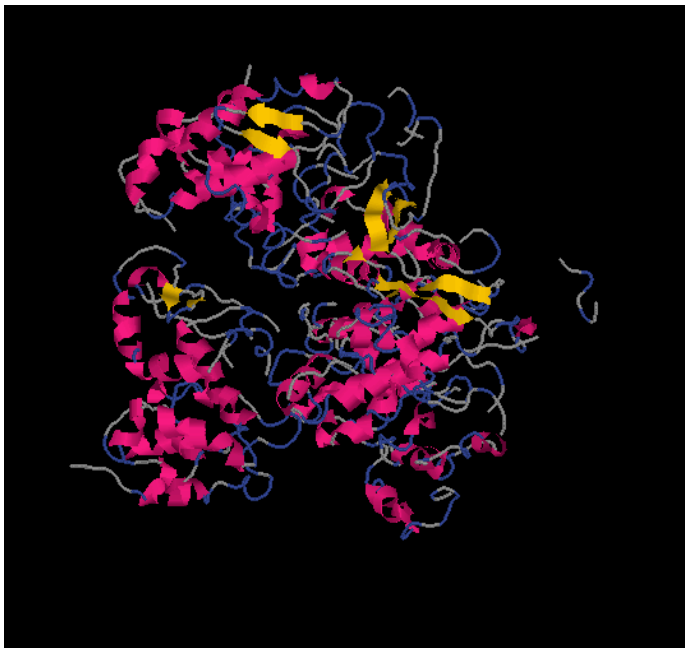


Figure 4.15 An initial model of the SBV NP. The initial model was built by 100 cycles of Buccaneer, and structure is shown in Rasmol.

The initial model of the SBV NP was built automatically using the program Buccaneer (Cowtan 2006) with the phases obtained from the SHELX. After 100 cycles of building, the initial model was obtained with an R_{free} of 0.4007 and overall figure of merit of 0.7558. The structure built by Buccaneer was checked using Rasmol (Goodsell 2005), which showed a predominantly alpha-helical structure (Figure 4.15).

The rest of the SBV NP structure was built using sequential rounds of manual model building in COOT (Emsley et al. 2004) followed by structure refinement with REFMAC5 (Murshudov et al. 2011).

4.3.4 The SBV nucleoprotein structure

Once the SBV NP structure finished building, the structure was compared to other protein structures published in the protein data bank (PDB) using Dali server (Holm et al. 2010). No similar structures were found, suggesting that the SBV NP had a novel fold. SBV NP protomer contains an N-terminal and C-terminal domains in addition to flexible N-terminal and C-terminal arms (Figure 4.16). Residues 1 to 19 form the N-terminal arm, whilst residues 214 to 230 form the C-terminal arm. The N-terminal domain spans residues 20 to 127 and is formed by five α helices (α 1- 5) and two anti-parallel β sheets (β 1- 2), whilst the C-terminal domain spans residues 128 to 213 and forms six α helices (α 6 -11). There are highly positively charged residues K48, K51, H77, R95, R184, R182, K178, K179, and R166 between these domains, which could potentially be the RNA binding cleft (Figure 4.17).

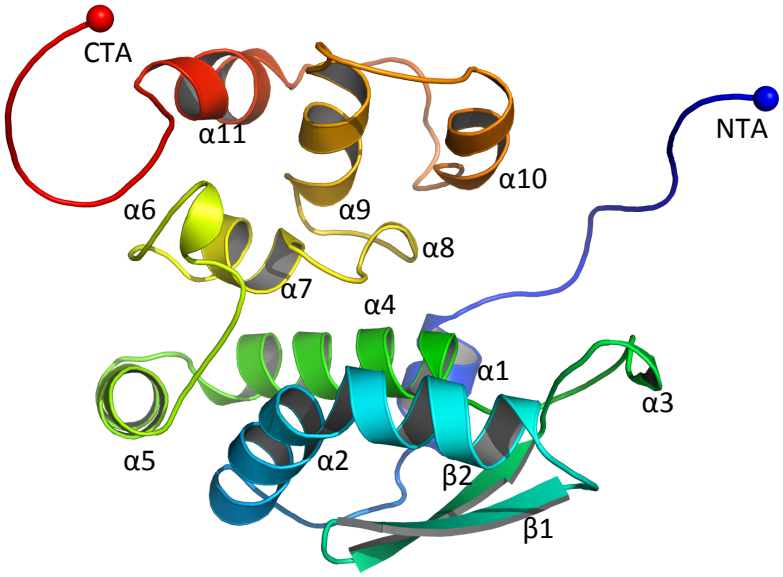


Figure 4.16 Cartoon representation of the monomeric SBV NP structure. The C-terminal arm (CTA) is shown as a red sphere and the N-terminal arm (NTA) is shown as a blue sphere.

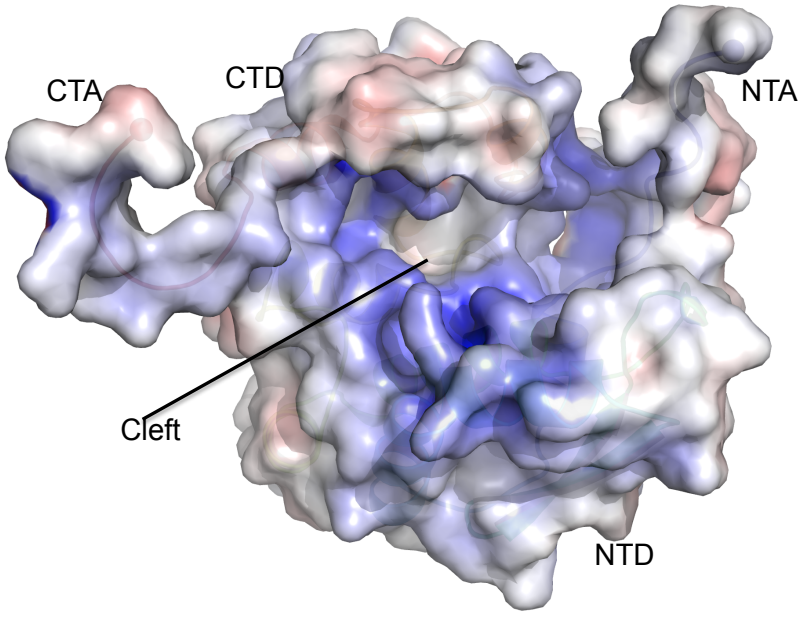


Figure 4.17 Electrostatic potential map of monomeric SBV NP. Positively charged residues are shown in blue and negatively charged residues in red. The cleft that is highly positively charged and is the potential RNA binding site. The NP composes of a C-terminal arm (CTA) and a C-terminal domain (CTD), an N-terminal arm (NTA) and an N-terminal domain (NTD).

4.3.5 The tetrameric structure

The SBV NP oligomerises in solution, predominantly as tetramers, formed by two dimers at a twofold rotation axis. The N-terminal arm is in contact with an adjacent protomer in the C-terminal edge of RNA binding cleft within tetramer, whereas the C-terminal arm is in contact with an adjacent protomer in hydrophobic pocket. Both the C-terminal and N-terminal arms play a role in oligomerisation (Figure 4.18 A, B). It is worth noting that one protomer has a free C-terminal are, which extends out in the dimer. The RNA-binding cleft of the promoter without a free C-terminal arm is exposed to the outside, but the RNA binding cleft of the promoter with a free C-terminal arm is largely covered at the interface (Figure 4.18 B).

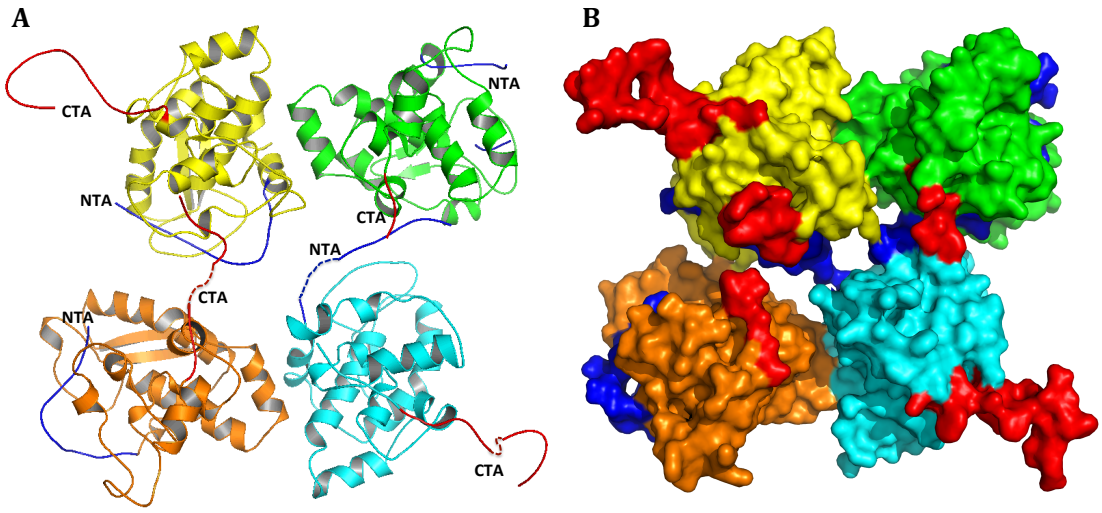


Figure 4.18 Tetrameric SBV NP structure. A, SBV NP tetramer representation as a cartoon. The C-terminal arm is in contact with an adjacent protomer in hydrophobic pocket (left dimer), whereas the N-terminal arm is in contact with an adjacent protomer in the C-terminal edge of RNA binding cleft within tetramer (right dimer). B, Surface representation of the tetrameric structure. The four different subunits are coloured in yellow, green, cyan and orange, respectively. The N-terminal arm (NTA) is in blue and the C-terminal arm (CTA) is in red.

The hydrophobic residues F5, I6 and F7 of N-terminal arm are located near a neighboring subunits hydrophobic binding site, composed of residues M124, L126, V129, F44 and L45 of the adjacent protomer (Figure 4.19 A). The side chain of K48 forms a hydrogen bond with side chain of S3, which may be involved in anchoring the N-terminal arm. Also the two dimers interact with each other to form a tetramer, whereby one dimer containing helix α5 interacts with helix α1 and α4 from the other dimer. The side chains of E117 and Q121 on helix α5 form hydrogen bonds with the side chain of Y24 on helix α1, while L113 on helix α5 located in hydrophobic pocket interact with W103, F27, I28 and the side chain of residue R102 on helix α1 and α4 (Figure 4.19 B). The tetrameric structure is stabilised by hydrophobic residues and pockets.

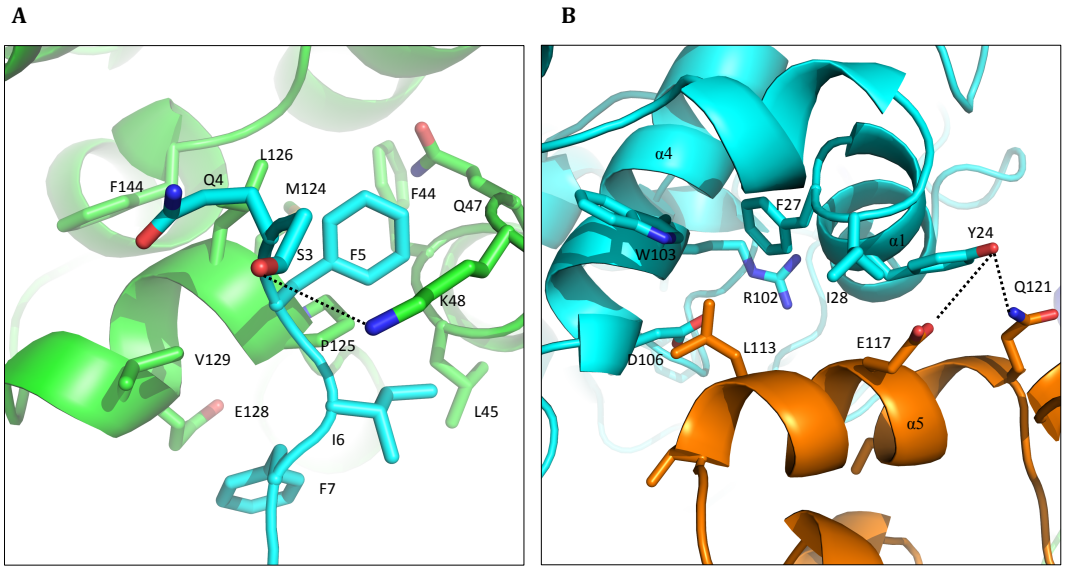


Figure 4.19 Tetrameric structure of interactions. A, The N-terminal (colored in cyan) arm interacts with a hydrophobic region of neighboring subunit (colored in green). The color according to SBV NP tetramer represented as a cartoon (Figure 4.18). B, Two dimers interact within tetramer. The side chains L113 located into a hydrophic pocket on the adjacent dimer. The dotted lines is indicate hydrogen bonds (Figure adapted from H. Dong et al. 2013).

4.3.6 Crystallization of SBV NP purified under denaturing and refolding conditions

To better understand the RNA-free SBV NP structure, we decided to purify the protein under denaturing and refolding condition using 8 M urea. As the native crystal diffracted poorly in-house (Figure 4.7), we measured the proteins 260/280 nm UV absorbance ratio and found it was 1.41. This indicated that it contained *E. coli* RNAs (Figure 4.9). As these RNAs are heterogeneous this could explain somewhat why the crystals only diffracted to low resolution.

The purification was described as in chapter 1.2.5.3. TEV protease was added to the purified SBV NP to remove the N-terminal 6His-tag. The protein was further purified by a second round of affinity chromatography, followed by gel filtration. This showed two peaks (Figure 4.20 A), and protein purity was checked using SDS-PAGE (Figure 4.20 B), which were similar to those from the native SBV NP. This is different from the RVFV-NP, which became monomeric after refolding (Raymond et al. 2010). The two peaks were pooled and concentrated to 9 mg ml^{-1} (Peak 1) and 8 mg ml^{-1} (Peak 2), respectively. The crystallization trails were described as in Chapter 1.2.6.4.

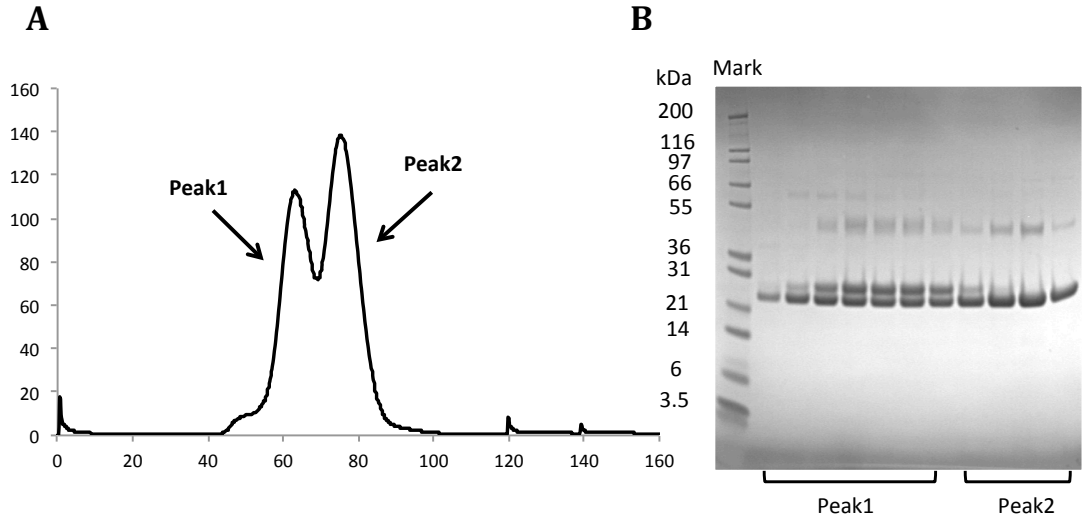


Figure 4.20 Purification of SBV NP denaturation and refolding. A, Gel-filtration of SBV NP denaturation and refolding. B, SDS-PAGE analysis of the protein's purity, the first peak contains similar amounts of 6His-tag intact (upper band) and cleaved SBV NP (lower band). The second main peak contains mostly 6His-tag cleaved SBV NP.

4.3.6.1 Data collection and structure determination of denatured and refolded SBV NP

Data collection was described as in chapter 1.2.6.8. The data were processed and scaled by HKL2000, and the cubic crystals were found to belong to space group *I*422 with cell dimensions a = b = 159.2 Å, c = 157.8 Å, and α = γ = β = 90° (Figure 4.21). The data collection statistics are listed below in Table 4.3.

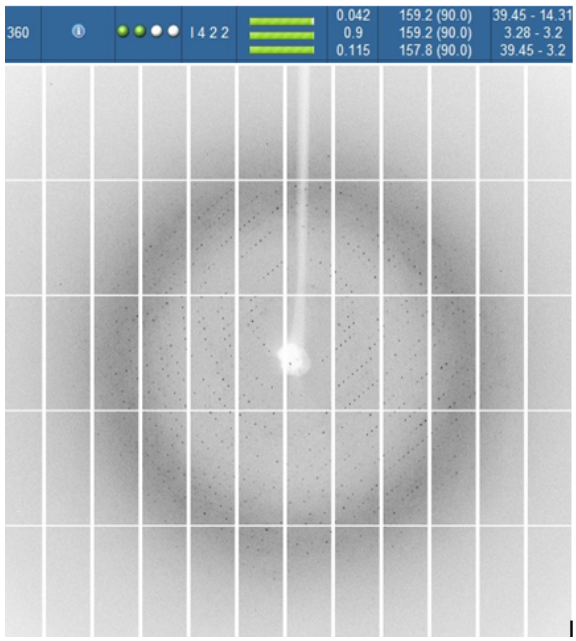


Figure 4.21 Diffraction pattern of the denatured and refolded SBV NP. The diffraction extend to 3.2 Å.

Resolution (Å)	49.94 – 3.21 (3.28 – 3.21)
Wavelength (Å)	0.9919
Space Group	*I*422
Completeness (%)	99.7 (99.9)
I/σ	16.4 (2.2)
Average redundancy	5.6 (5.5)
Unit Cell (Å)	a =b = 159.2, c = 157.8 α = γ = β= 90°
Unique reflections	16505 (1080)
Wilson B-factor	94.77
R_{merge} (%)	7.3 (68.30)
Ramachandran favoured (%)	81
Protein residues	640
Water	18
Refiment	
R_{factor}	0.30
R_{free}	0.37
R.M.S. deviation	
Bonds (Å)	0.006
Angles (°)	1.162
PDB code	4IDX

Table 4.3 Denatured and refolded SBV NP data collection statistics. Values in parentheses are represents for the highest-resolution shell. RMSD, root mean square deviation.

$\mathbf{R_{factor}= \Sigma || F_{obs}| - | F_{cal}||/ \Sigma |F_{obs}|}$, where F_{obs} and F_{cal} are observed all reflection measured and calculated currently model as structure factors, respectively. R_{free} is calculated using 5% of total reflections, which is randomly selected not used in refinement.

4.3.6.2 Cell content calculation

The unit cell content was estimated using the Matthews co-efficient to contain three copies of SBV NP within the unit cell within CCP4 suite (Winn et al. 2011). The structure of the denatured and refolded SBV NP was determined by molecular replacement using Phaser within CCP4 (McCoy et al. 2007) using structure of native SBV NP's monomeric form as a search model. The model was built in COOT (Emsley et al. 2004) and structure refinement was carried out using REFMAC5 (Murshudov et al. 2011), in an iterative process.

4.3.7 The structure of denatured and refolded SBV NP

The protomer structure of the refolded SBV NP is very similar to the native SBV NP. The most notable change is that the N-terminal arm has becomes disordered, while the C-terminal arm becomes ordered (Figure 4.22).

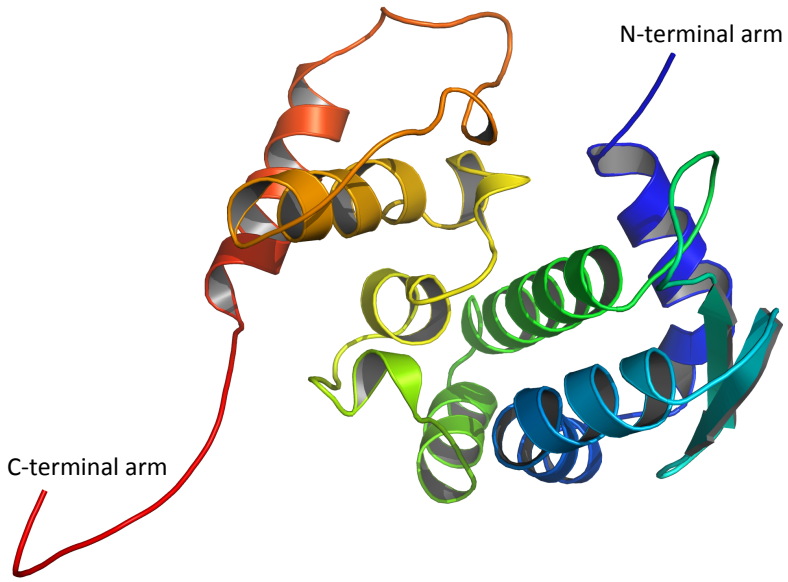


Figure 4.22 The protomer SBV NP structure from denatured and refolded protein. The N-terminal arm (blue) is disordered. The C-terminal arm colored red.

There are three SBV NP subunits within the asymmetric unit, which do not form a circular structure (Figure 4.23 A). Instead, their potential RNA binding clefts are exposed and accessible to solvent. (Figure 4.23 B). Unlike monomeric RVFV-NP, oligomerisation occurs within the denatured and refolded SBV NP crystal structure.

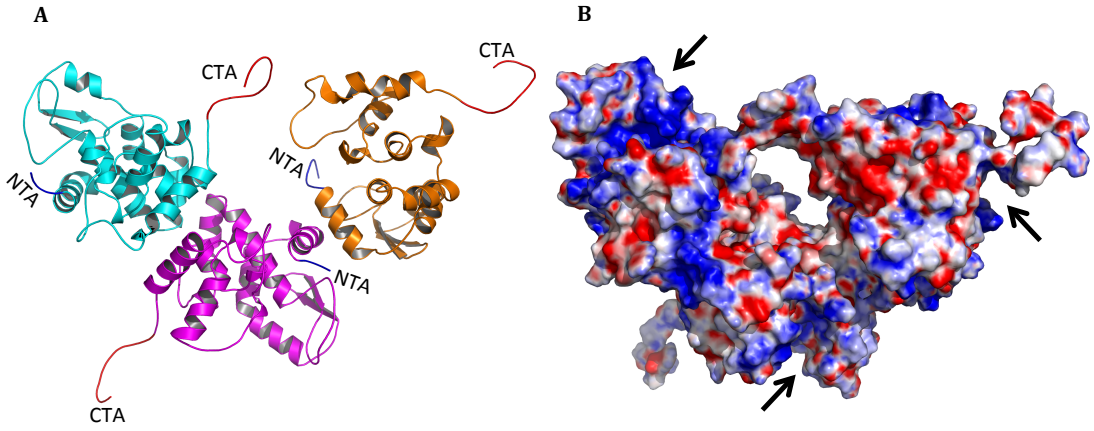


Figure 4.23 Trimeric SBV NP denatured and refolded structure. A, The refolded SBV NP structure shows in trimer, and the three subunits are coloured in cyan, purple and orange respectively. NTA and CTA represent N-terminal arm and C-terminal arm, respectively. B, Electrostatic surface potential map of trimeric SBV NP. The arrows point to the exposed RNA binding clefts within each subunit. The positive charges are in blue and negative charges are in red.

4.3.8 Mutagenesis and mutant protein purification

The N-terminal arm is in contact with an adjacent protomer in the C-terminal edge of the RNA binding cleft within the tetramer. The N-terminal arm is also very flexible, suggesting that it may play a role in protecting and binding RNA. To test this hypothesis, we generated N-terminal (Δ1-19) and C-terminal (Δ217-233) arm truncates, along with single (K48A, R41A, K51Q), double (R41G/K51Q) and triple (R41G/K51Q/W95Q) SBV NP mutants. These truncated and mutants were overexpressed and purified as the native SBV NP (Chapter 1.2).

4.3.9 Mutation in proteins impaired RNA Binding activity

Equal amounts of purified native and mutant proteins were loaded onto a 2% agarose gel and ran at 100 V for 30 min. The gel was stained with GelRed. SBV NP with truncated N-terminal arm and the R41G/K51Q/W95Q triple mutant completely lost their RNA binding ability suggesting that these residues located in potential RNA binding cleft and are essential for RNA binding. However, the single and double mutants partially lost their RNA binding activity (Figure 2.24). It is worth nothing that single mutant K48E, and double mutant R41G/K51Q were less bound with RNA than SBV NP native which served as a control. The experiment data was provided by Dr. Ping Li (University of Glasgow).

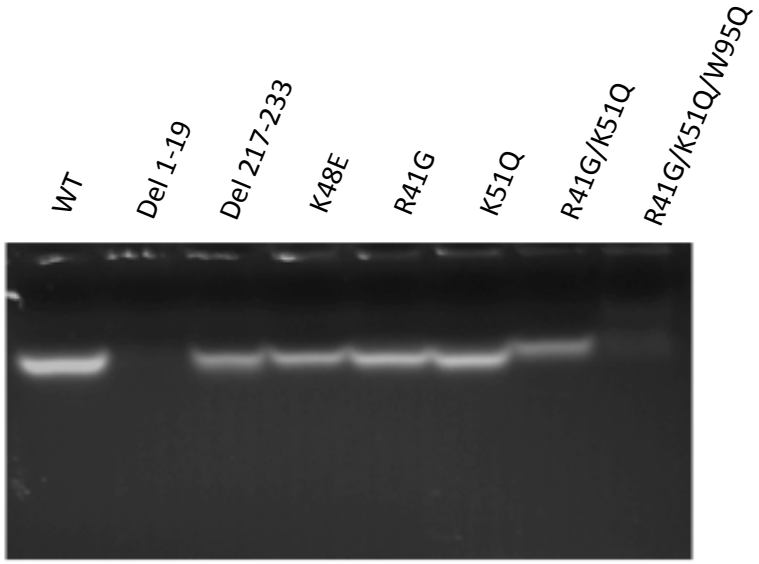


Figure 4.24 The RNA binding activities of native SBV NP and single, double and triple SBV NP mutants. The SBV NP native protein contained host RNA as a wild type (WT) control, the N-terminal arm deletion (Del 1-19) and triple mutant (R41G/K51Q/W95Q) completely lost their RNA binding ability, while the single (K48E, R41G, K51Q) and double (R41G/K51Q) mutants partially lost their RNA binding activity.

4.4 Discussion and conclusion

The first Orthobunyavirus nucleoprotein structure of SBV NP was determined, which has a novel fold. A few viral nucleoprotein structures (RVFV, CCHFV) of negative single stranded RNA virus has been solved (Raymond et al. 2010; Ferron et al. 2011; Guo et al. 2012; Carter et al. 2012; Wang et al. 2012), and they share the feature of having a highly positive charge cleft for binding genomic RNA. The SBV NP also has a putative RNA binding cleft between the N-terminal and C-terminal domains, and the N-terminal arm is crucial for RNA binding which according to our mutagenesis studies involved several positively charged residues. These are likely to play an important role in genomic replication. The tetrameric oligomerisation states, which are formed from dimers and are mediated by N and C-terminal arms interactions, indicate that the N and C-terminal arms play an essential role in oligomerization.

With RVFV, the protein purified from denaturation and refolding is monomeric (Raymond et al. 2010). In contrast, the SBV NP from denaturation and refolding is trimeric and the N-terminal arm becomes disordered, while C-terminal arm becomes ordered. The purified denatured and refolded SBV NP should provide us with a great opportunity to obtain a complex with a synthetic RNA. The SBV NP/RNA complex will provide valuable information on how the SBV NP recognized and encapsidates the genomic viral RNA. This would reveal, in exquisite molecular detail, how the RNA binding residues interact the RNA, which will provide a model for understanding the mechanisms of RNA encapsulation, replication and transcription of the largest Orthobunyavirus family. This might provide an opportunity for novel vaccine and therapy development to control infections caused not only by SBV, but other pathogenic human and animal Orthobunyviruses.

Chapter 5

Crystal structure of Schmallenberg virus nucleoprotein in complex with RNA

5.1 Introduction

Bunyaviridae is a large family of enveloped, single-stranded negative-sense RNA viruses, with more than 350 members across five genera: *Orthobunyavirus*, *Hantavirus*, *Nairovirus*, *Phlebovirus* and *Tospovirus* (Elliott et al. 2013; Walter et al. 2011). Of these five genera, the *Orthobunyavirus* genus is the largest, currently comprising about 170 members.

All Orthobunyaviruses contain three genome segments of single-strand negative-sense RNA, which are encapsidated within the virus encoded nucleocapsid protein to form a ribonucleoprotein (RNP) complex. The RNP facilitates virus replication, transcription and assembly. SBV is a negative-sense single-stranded RNA virus and phylogenetic analyses suggest that SBV belongs to the species *Sathuperi virus* (Garigliany et al. 2012; Goller et al. 2012), which is an *Orthobunyavirus* of the Simbu serotype.

We have recently solved the crystal structure of SBV NP in tetrameric and trimeric forms, with the purified protein under native condition and under denaturation and refolding condition respectively, and revealed a novel fold (H. Dong et al. 2013a). Although a potential RNA-binding cleft has been identified, details of the mechanism how SBV NP recognizes and encapsidates RNAs to form an RNP complex is unknown. Here, I report the crystal structure of SBV NP in complex with 42 nucleotide (nt) long RNA, which will help us to understand viral RNA encapsidation, replication and transcription, and this could be an important target for developing vaccines and drugs.

5.2 Aims

In Chapter 4, the SBV NP was purified under both native and denature and refold conditions, and the structures were determined. However, how the SBV NP recognizes and encapsidates RNAs to form an RNP complex is unknown. In this chapter, I will purify the protein using same method that protein without RNA (*E. coli*), then I will crystallize protein in complex with different lengths of synthesized RNAs to obtain crystals. The SBV NP/RNA complex will provides valuable information on how the SBV NP recognized and encapsidates the genomic viral RNA.

5.3 Results

5.3.1 Crystallization of SBV NP/ RNA complexes

The purification of SBV NP was described as in chapter 1.3. The purification in low salt concentrations (300-0 mM) during gel-filtration was attempted, in order to obtain SBV NP and nucleic acid complex. The protein was concentrated, incubated on ice with different length RNAs, 21-, 28-, 42-nt RNAs individually for 90 minutes. The crystallization trials were set up and screened (chapter 1.3.2). The crystals of SBV NP protein in complex with 21 nt and 28 nt RNA were obtained, but the crystals were diffracted poorly. I then increased the salt concentration to 300 mM in the final gel-filtration step. After one week, protein-RNA complex crystals were obtained with protein associated with 21-, 28-, 42-nt RNA complex

5.3.1.1 Crystallization of complex SBV NP with 21-nt RNA

The crystallization trails were described as in Chapter 1.3.2. Protein-21-nt RNA complex crystals were formed in varying crystallizations conditions with the best crystals growing from 0.1 M Bis-Tris pH5.5, 0.2 M Ammonium acetate, 25% PEG 3350. Crystal optimisation was carried out by varying the 0.1M Bis-Tris pH from 5.1 to 6.0 and the PEG 3350 concentration from 21 to 29%. The crystal optimisation was performed using 96 well crystal clear sitting–drop plates by mixing 0.5 µl of protein and 0.5 µl of crystallization precipitant with 100 µl reservoirs solution (Figure 3.1). The crystals were sensitive when the drop size was increased to 1 µl of protein and 1 µl of crystallization precipitant. The optimised plates were incubated at room temperature (20°C).

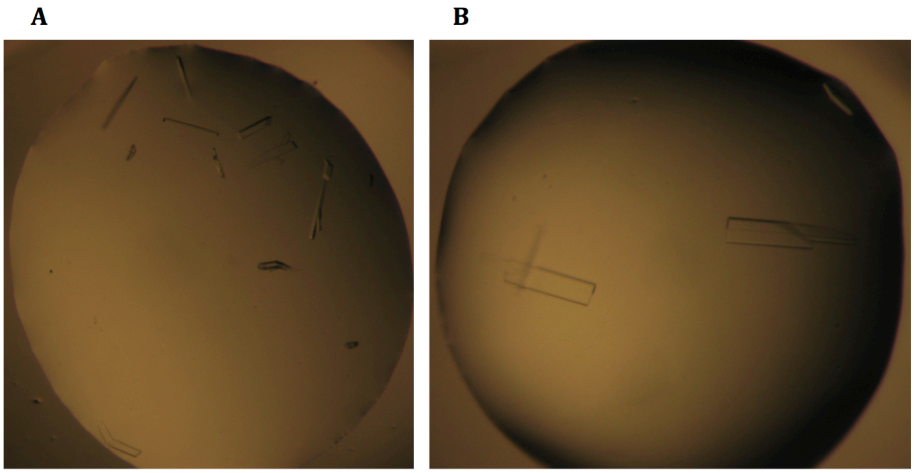


Figure 5.1 Crystal of SBV NP in complex with 21-nt of RNA complex crystals. A, The initial crystals from screening. B, The crystals from optimization; the crystal grew in 0.1 M Bis-Tris pH 5.7, 0.2 M ammonium acetate, 23% PEG 3350.

5.3.1.2 Crystallization of complex SBV NP with 28-nt RNA

The best crystals of the protein SBV NP complexed with 28-nt RNA grew in crystallization conditions of 0.1 M Bis-Tris pH 5.5, 0.2 M lithium sulphate and 25% PEG 3350. Crystal optimisation was carried out by varying the 0.1M Bis-Tris pH from 5.1 to 6.0 and the PEG 3350 concentration from 21 to 29%.

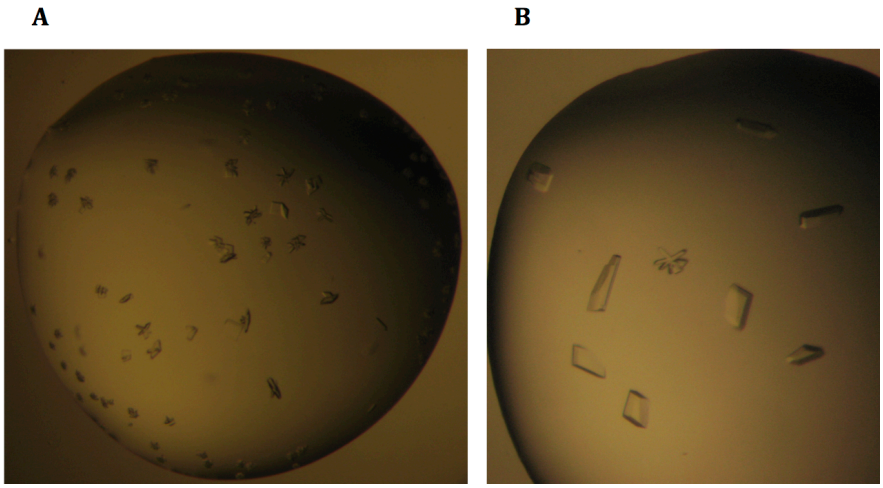


Figure 5.2 Crystals of the NP in complex with 28-nt of RNA complex. A, Crystals from original screening condition. B, The crystals from optimizations; the crystals grew in 0.1 M Bis-Tris pH 5.7, 0.2 M lithium sulphate and 25% PEG 3350.

5.3.1.3 Crystallization of complex SBV NP with 42-nt RNA

The protein complexed with 42-nt RNA crystals were grew in crystallization condition of 0.1 M Bis-Tris pH 5.7, 0.2 M sodium chloride and 25% PEG 3350. Crystallization optimisation was carried out by varying the 0.1M Bis-Tris pH from 5.1 to 6.0 and the PEG 3350 concentration from 21 to 29%.

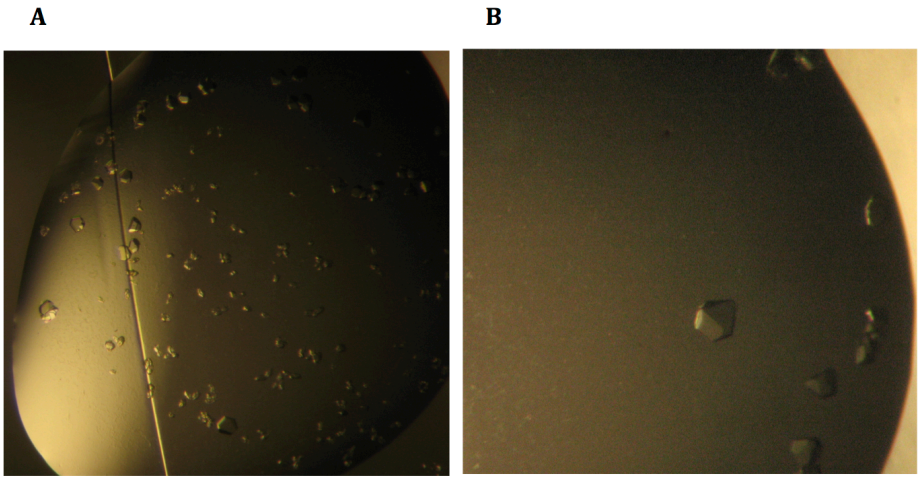


Figure 5.3 Crystals of SBV NP complexed with 42-nt RNA. A, Crystals from original screening condition. B, The crystals from optimizations. The crystals grew in 0.1 M Bis-Tris pH 6.0, 0.25- 0.3 M lithium sulphate and 25% PEG 3350.

5.3.2 Determination of structure of protein complexed with 21-, 28-, 42-nt RNA

The crystals of SBV NP in complex with 21-, 28-, 42-nt RNA were screened using in-house X-ray source. All crystals were diffracted poorly, except the crystals of 42-nt RNA complex, and data set was collected using in-house X-ray facility (Figure 5.4). Data collection was described as in chapter 1.3.2.2.

The data was indexed and integrated using iMosflm (Battye et al. 2011), and the CCP4 program Pointless (Winn et al. 2011) suggested the space group is $P2_1$. The data was scaled using Scala (Evans 2006). A summary of the data collection statistics is listed below in table 3.1.

The structure of the SBV NP-42-nt RNA complex was solved by Molecular replacement with Phase (McCoy et al. 2007) using the native SBV NP's chain B of PDB (4IDU) as a search model (H. Dong et al. 2013a) and a single solution was found with Z-score of 21.6. The unit cell content was estimated using the Matthews co-efficient, which suggested four copies of SBV NP per unit cell using CCP4 suite (Winn et al. 2011). The model was built in COOT (Emsley et al. 2004) and structure refinement was carried out using REFMAC5 (Murshudov et al. 2011).

Four monomers were identified in the structure, and RNA density was observed at the positively charged cleft (Figure 5.4 C). Due to a low resolution of 2.7 angstroms, the RNA density was not very good. Therefore it was important to collect a higher resolution dataset at Diamond, and therefore the structure was not further built and refined (Table 5.1).

Resolution (Å)	**85.62 – 2.77 (2.86 – 2.77)**
Wavelength (Å)	**1.54178**
Completeness (%)	98.71 (86.72)
I/σ	14.23 (3.37)
Multiplicity	3.0 (2.8)
Unit Cell (Å)	a = 76.33, b = 85.68, c = 77.48 α = γ = 90°, β = 101.34°
Unique reflections	24441 (2135)
R_{merge} (%)	13.2 (57.4)
Space Group	$P2_1$
Wilson B-factor	50.74
Refinement	
R_{factor}	0.37
R_{free}	0.42
Number of atoms	6692
Protein residues	844
RMS (bonds)	0.01
RMS (angle)	1.57
Ramachandran favoured (%)	69

Table 5.1 SBV NP-42-nt RNA complex data collection statistics. Values in parentheses are represents for the highest-resolution shell. $\mathbf{R_{factor} = \Sigma \, || F_{obs}| - | F_{cal}|| / \Sigma \, |F_{obs}|}$, where F_{obs} and F_{cal} are observed all reflection measured and calculated currently model as structure factors, respectively. R_{free} is calculated using 5% of total reflections, which is randomly selected not used in refinement.

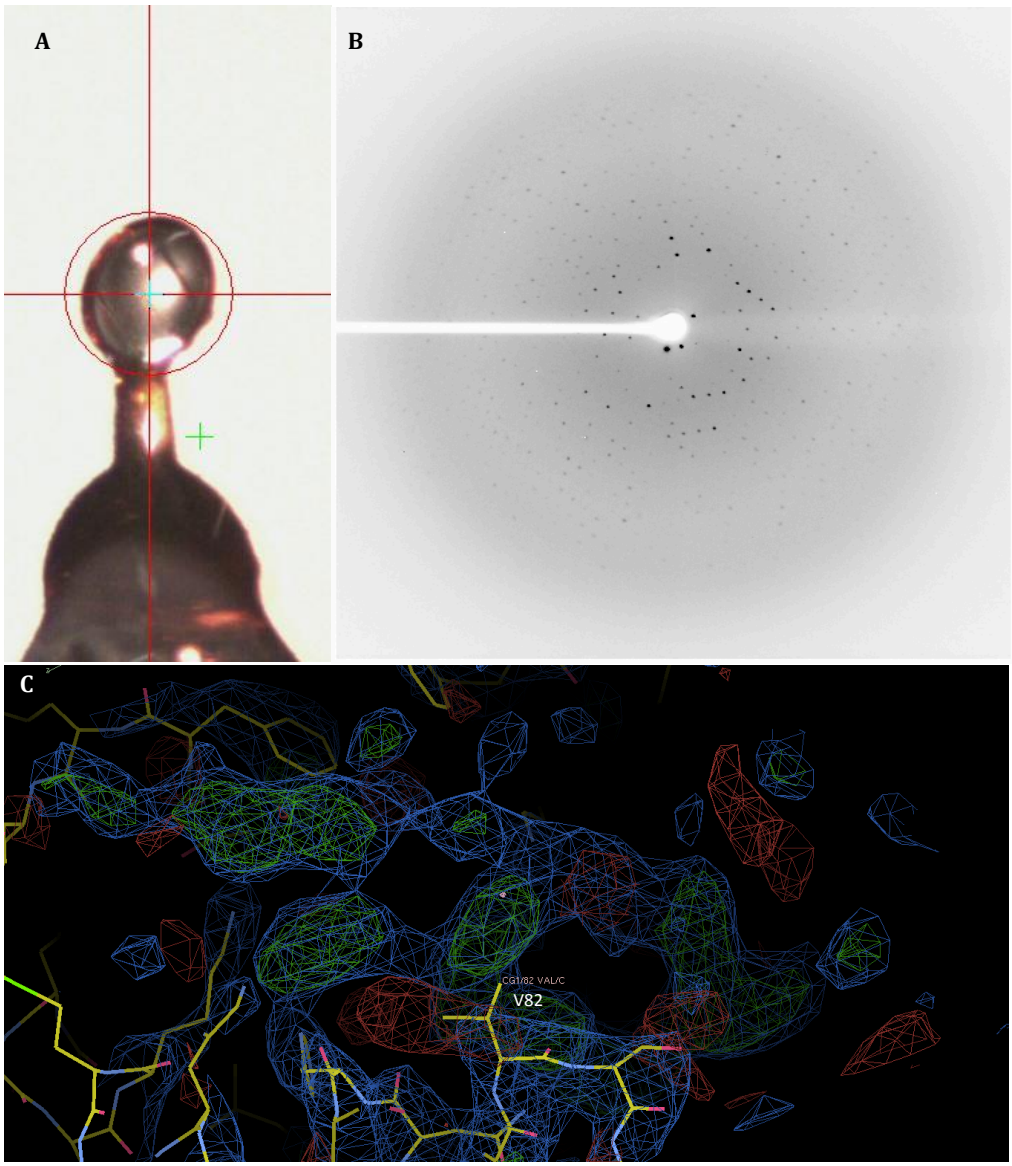


Figure 5. 4 Data collection of SBV NP in complex with 42-nt RNA in-house. A, Crystal mounted in 0.2 μm litho-loops (Molecular Dimensions). B, The crystal diffraction pattern to 3.0-Å resolution from in-house X-ray source. C, RNA density maps coloured green, which close to residue V82 was shown as sticks and the 2Fo-Fc map contoured at 1.0 sigma.

5.3.3 Data collection of 42-nt RNA-protein complex

The dataset was collected at Diamond beamline I24 (Figure 5.5). The data was indexed and integrated using iMosflm (Battye et al. 2011), and the CCP4 program Pointless (Winn et al. 2011) suggested the space group is $P2_1$. The data was scaled with Scala (Evans 2006). A summary of the data collection statistics is listed below in table 5.2.

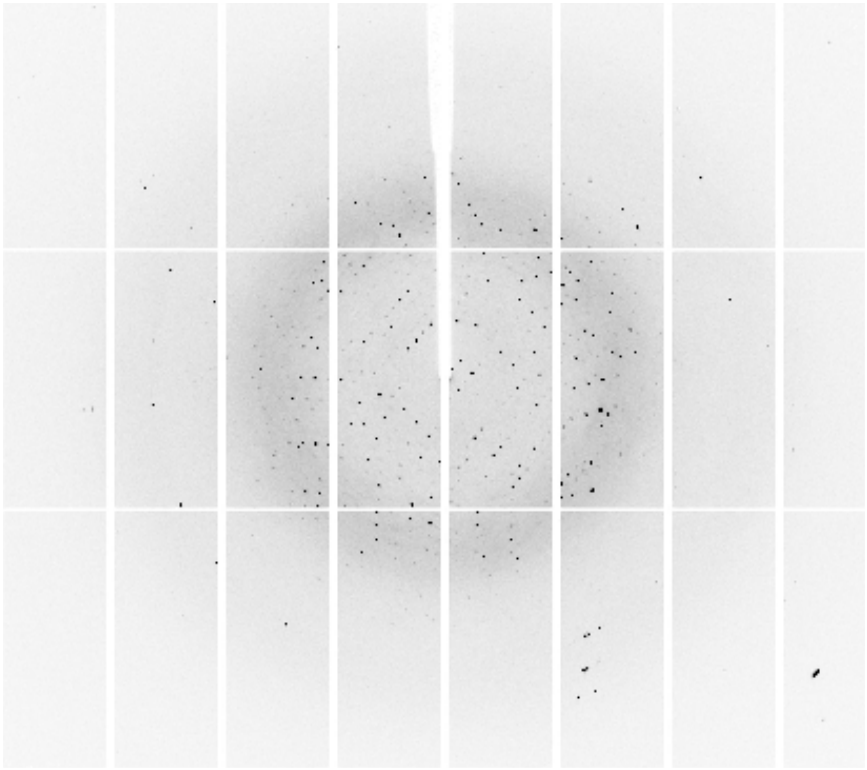


Figure 5.5 Data collection of SBV NP in complex with 42-nt RNA. The crystal dataset was collected to 2.16 Å at Diamond beamline I24 in UK.

Resolution (Å)	**37.82 – 2.12 (2.23 – 2.16)**
Wavelength (Å)	**0.92**
Completeness	95.9% (89.5%)
I/σ	11.8 (2.6)
Unit Cell (Å)	a = 76.50, b = 86.05, c = 77.46 α = γ = 90°, β = 101.98°
Unique reflections	54303 (7984)
Average redundancy	4.1 (4.3)
R_{merge} (%)	61 (57.2)
Wilson B-factor	33.14
Space Group	$P2_1$
Refiment	-
Number of atoms	-
Protein	7026
RNA	840
Solvent	152
Mean B-factor (Å^2)	-
Protein	45
RNA	48
Solvent	52
R.M.S. derivation	-
Bonds (Å)	0.01
Angles (°)	1.50
R_{free} **R_{work}**	0.2766 0.2594
PDB code	4JNG

Table 5.2 SBV NP-42-nt RNA complex data collection statistics. Values in parentheses are represents for the highest-resolution shell. $\mathbf{R_{factor} = \Sigma \, || F_{obs}| - | F_{cal}|| / \Sigma \, |F_{obs}|}$, where F_{obs} and F_{cal} are observed all reflection measured and calculated currently model as structure factors, respectively. R_{free} is calculated using 5% of total reflections, which is randomly selected not used in refinement.

5.3.4 SBV NP complexed with 42-nt RNA forms tetramer

The oligomeric NP with the 42-nt RNA was confirmed by chemical cross-linking analysis (Figure 5.6), which indicated that this complex exists as a tetramer (lane 3). This is similar to protein purified under native condition containing *E coli* RNA (lane 1). The denatured and refolded protein (RNA-free) exists in trimer on SDS-gel (lane 2). Crosslinking data was provided by Dr Ping Li (University of Glasgow).

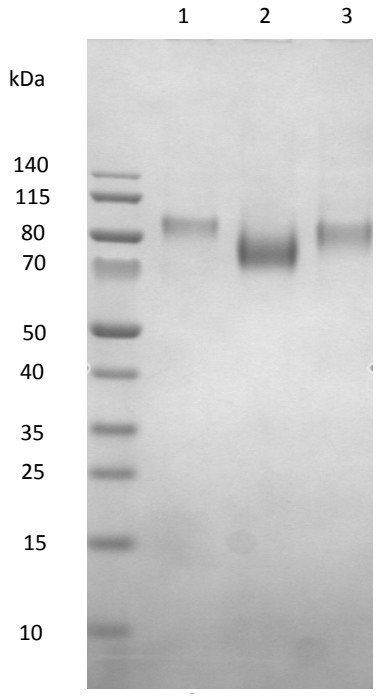


Figure 5.6 Chemical cross-linking shows that SBV NP in complex with 42 nt RNA form tetramer in solution. SDS-PAGE under non-reducing conditions. Lane 1, protein purified under natively condition contained host RNA; Lane 2, protein purified under denaturation and refolded RNA-free; Lane 3, refolded protein incubated with 42-nt RNA complex. Molecular weight size markers are shown on the left.

5.3.5 Crystal structure of tetrameric SBV NP-42-nt RNA complex

Crystal structure of SBV NP in complex with 42-nt RNA was determined using molecular replacement to 2.16 Å resolution by Phaser. The structure formed a tetrameric ring structure (Figure 5.7 A, B), where C-terminal arm of each protomer interacted with adjacent C-terminal domain in hydrophobic region of the protein (Figure 5.7 C). This result is consistent with that in Chapter 2. The structure suggested that the C- terminal arm plays an important role to mediate oligomerization. It is worth nothing that the N-terminal arms of three protomers were interacted with each adjacent protomer to further stabilized the tetrameric structure. The N-terminal arm of fourth protomer was disordered (Figure 5.7 D). The RNA strand was located inside the tetramer ring bound to the RNA-binding site of each protomer (Figure 5.7 A, B).

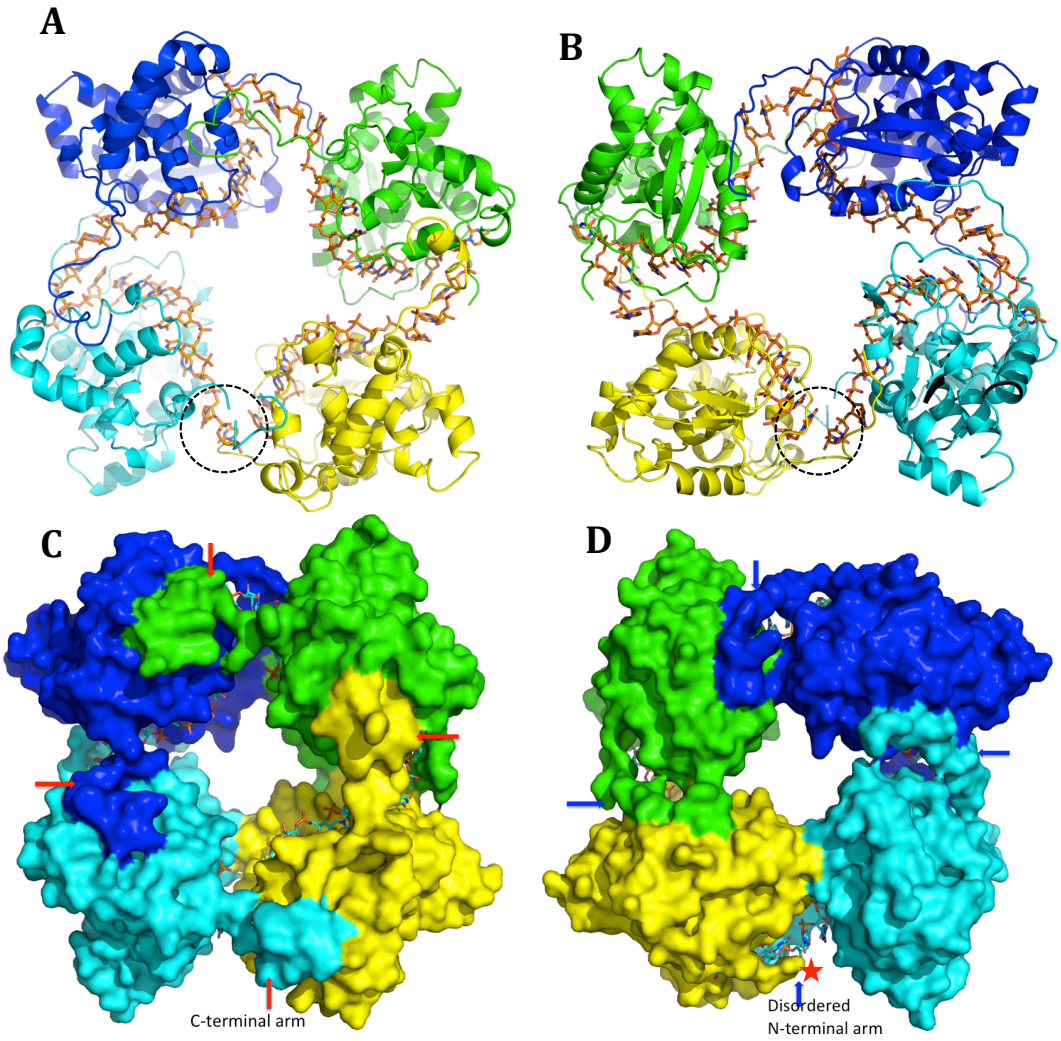


Figure 5.7 Crystal Structure of the SBV NP in complex with 42-nt RNA. A, The tetrameric ring of the SBV NP-RNA complex in cartoon representation bound to RNA. RNA is represented as orange stick inside the tetrameric ring. Four protomers are shown in blue, green, yellow and cyan respectively. The black dotted line shows the gap in the RNA. B, From A to B is the 180 ° rotation of complex along the y-axis. C, Electrostatic surface of the tetrameric structure. The C-terminal arms are depicted by red narrows, and in contact with protomers of adjacent C-terminal domains in the hydrophobic region. D, From A to B is rotation about along y-axis at 180 °. The N-terminal arms are depicted by blue narrows with N-terminal arms bound to RNAs and interacting with neighboring protomers. An N-terminal arm of the protomer shown as yellow cartoon is disordered, and labeled by a red star.

5.3.6 SBV NP binds to RNA mainly at the positively charged groove

The 42-nt RNA electron density is clearly observed along the inner edge of tetramer ring-structure (Figure 5.8 A). This is consistent with previously suggestion that a positively charged cleft is necessary for potential RNA-binding in the cleft between the N- and C- domains of SBV NP (H. Dong et al. 2013a). In the tetramer-RNA complex structure, it clearly shows that each protomer binds 11 nucleotides, of which 8 RNA bases (U1- U8) bind at the positively charged cleft, and 3 RNA bases (U9-U11) bind at the N-terminal arm. The residues K48, K51, H77, R95, R184, R182, K178, K179, and R166 interact with RNA and play an essential role in binding the RNA (Figure 5.8 B- D). Single mutations on residues K48 and K51 impaired RNA-binding activity as described in section chapter 2.4.7.1 previously. Two of the hydrophobic residues, F18 and F176 affect the orientation of the RNA bases to an "S" shape for the RNA chain in the cleft (Fig. 5.8 B). Nucleotides U1, U2 and U3 face the protein inwardly, and interact with residues K48, K51, L126, R166 and F176 (Fig. 5.8 B-D), while nucleotides U4, U5, U6 and U7 were exposed to the solvent, stacked together at the outer surface of the protein, and were in direct contact with positive charged residues R95, K178, K179, and R182. The nucleotide U8 was buried deeply within the cleft, and was sandwiched by residues N19 and R184 (Fig. 5.8 B- D).

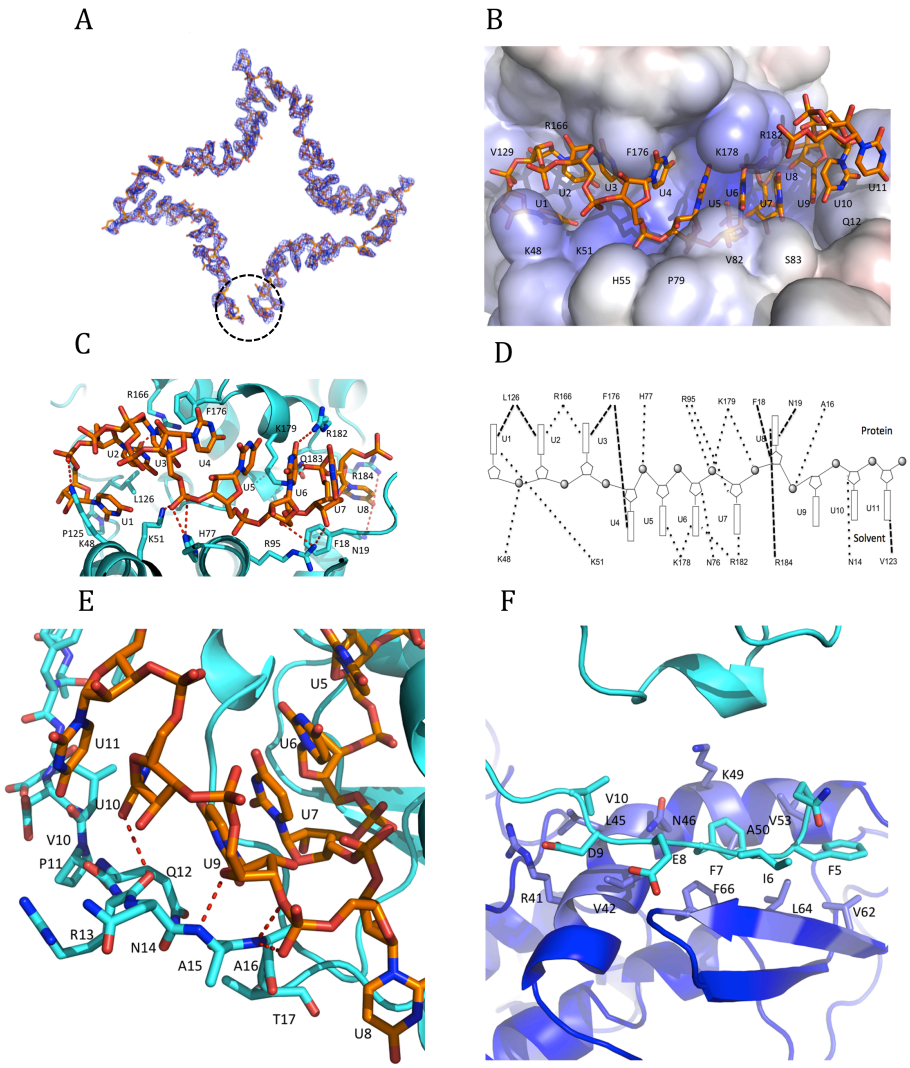


Figure 5.8 Orthobunyavirus SBV NP interactions with RNA. A, Electron density map (Fo-Fc) of 42-nt RNA (polyU) was contoured at 3 σ. The RNA density is blue, and RNA is shows as orange stick. The black dotted line shows the gap in the RNA. B, Electrostatic potential surface shows a protomer bound to 11-nt RNA of which 8-nt RNA bound in the positive charged cleft and 3-nt RNA bound in the N- terminal arm. C, SBV NP interacts with RNA at positive charged cleft, and RNA form "s" shape architecture. D, Schematic diagram shows that protein residues interact with RNA. Nucleotides oriented to the top faced the protein in the RNA-binding cleft, and nucleotides oriented to the bottom were exposed to the solvent. The dotted dark line shows residues interacting with RNA. E, Interaction between the N-terminal arm and the RNA. Nucleotides (U9-U11) interact with residues of main chain, which is located in the N-terminal arm. F, The N-terminal arm is in contact with adjacent protomer in the hydrophobic site. The residues F7 and F66 form π–π interaction. The N-terminal arm is colored cyan and the hydrophobic site of the adjacent protomer is blue (Figure adapted from H. Dong et al. 2013).

5.3.7 The N-terminal arm binds RNA

To examine the interaction between the N-terminal arm and RNA, 3 nucleotides (U9-U11) are found to bind at N-terminal arm and interact with neighboring protomers (Figure 5.7 D and 5.8 B- E). The residues Q12, A15, A16 on the N-terminal am interacted with bases U9, U10, U11 to exposed RNA to solvent. The result is consistent with that section 4.3.7, whereby the protein structure of SBV NP purified under denaturation and refolding showed that the N-terminal arms were disordered (Figure 4.22). In chapter 4.3.9 it was described that deletion of the N-terminal arm (1-19) resulted in completely loss of RNA-binding ability. The N-terminal arm of each protomer with adjacent protomer to form π–π interactions in the hydrophobic site consisting of residures V42, F66, L64, V62, V53. In particular the N-terminal residue F7 on a protomer interact with residues F66 on adjacent protomer (Figure 5.8 F).

The structure of the SBV NP-RNA complex showed that each SBV NP protomer binds to 11 nucleotides. Therefore, the tetramer is able to bind 44-nt RNA. Since the protein (RNA-free) was incubated with a 42-base-length RNA for the crystallization, this resulted in a 2 nucleotides gap in the RNA between two adjacent protomers within the tetrameric ring structure (Figure 5.7 A, B). The electron density map indicates clearly the RNA density and the gap (Figure 5.8 A).

5.3.8 Oligomerization of SBV NP using electron microscopy

In order to examine the oligomeric states of SBV NP in solution, the purified native protein (Containing *E. coli RNA*) and the protein (RNA-free) in complex with 42-nt RNA were examined by electron microscopy (EM). The EM studies indicated that the natively purified protein and the refolded protein in complex with 42-nt RNA were mainly observed as tetramers, and minor amounts of oligomerization states as trimers or pentamers (Figure 5.9 A, C). While the oligomerization states of denatured and refolded protein were heterogeneous and irregular, the predominant oligomeric states were trimeric that were observed (Figure 5.9 B). The EM projection maps match greatly with the crystal structures, and the maps were selected class averages to show in overlays of crystal structure (5.9 D).

The result is consistent with that of section 5.3.4, chemical cross-linking analysis of SBV NP revealed oligomeric species in solution (Figure 5.6). The EM analysis of SBV NP oligomerization states in solution was largely in agreement with chemical cross-linking and crystal structures. Our collaborator Dr. Bettina Böttcher (University of Edinburgh) provided the EM data.

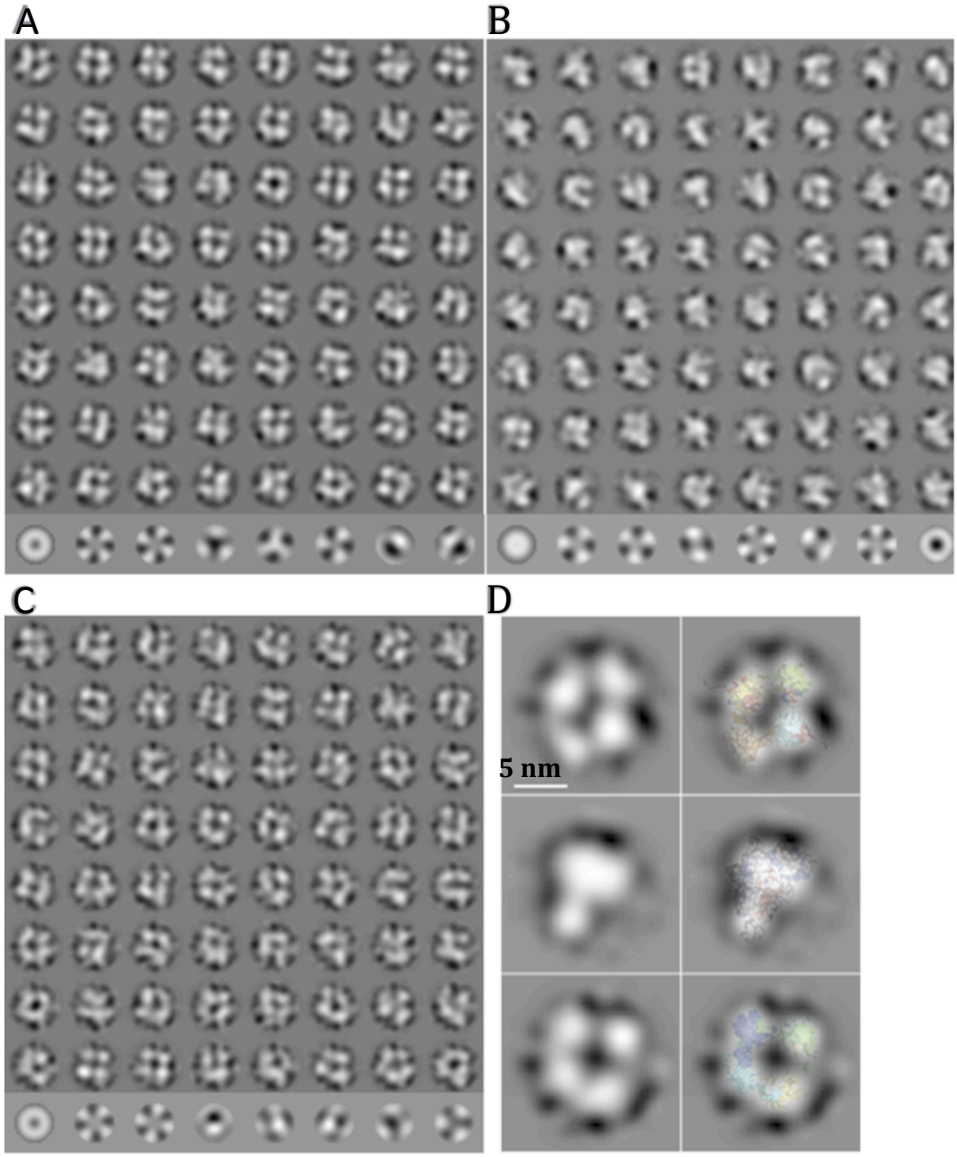


Figure 5.9 EM of negatively strained SBV NP. A, Micrograph of native purified protein, class averages (row 1- 8) and eigen-images (row 9). Most of the class averages observed as tetramer, and there are minor other oligomeric states as trimer (row 8, column 3) or pentamer (row 8, column 6). B, Class averages of refolded protein. The class averages show heterogeneous particle, and the predominant oligemeric state as trimer (row 1, column 3; row 2, column 8; row 3, column 6; row 7, column 1) and some tetramers were also observed (row 1, column 4; row 6, column 3). C, Micrograph refold protein in complex with 42-nt RNA. The class averages mainly show tetramers. D, Selected class averages (left) and class average EM projection map match well with crystal structure overlaid (right). The native SBV NP (top), refolded SBV NP (middle), and refold SBV NP in complex with 42-nt RNA (bottom) respectively (Figure adapted from H. Dong et al. 2013).

5.3.9 Refolded SBV NP undergoes conformational changes upon RNA binding

After SBV NP was purified under denaturation and refolding condition, the protein was in trimeric form and SBV NP requires conformational change for RNA binding. The EM analysis of the refolded SBV NP showed that the refolded protein in solution is heterogeneous with trimer as the predominant species (Figure 5.9 B). The refolded protein's crystal structure also shows as trimer, and the structure is deposited in PDB (code: 4IDX). In this structure all the positively charged residues in the RNA-binding cleft were exposed to the solution and that is ready accessible to RNA in solution (Figure 5.10 A). When refolded protein was incubated with 42-nt RNA, the refolded protein was converted from trimer to tetramer and most of the positively charged residues in the RNA-binding cleft were orientated inwardly, facing ring within tetramer instead of being exposed to solvent on the outside. Although the overall structures of refolded SBV NP and SBV NP in complex with RNA are similar with root mean square deviation (RMSD) of 1.05 over 201 residues (Cα backbones) (Figure 5.10 B). It is noteworthy that three of the N-terminal arms were disordered in the trimeric structure and three of N-terminal arms were ordered in the tetramer ring with RNA complex structure (Figure 5.7 D). Furthermore, the side chains of residues L45, R41, K48, K51, H77, V82, F18, N19, P20, R182, R184, K178, K179, F176, and R166 facilitate significantly conformational changes for binding the RNA (Figure 5.10 C).

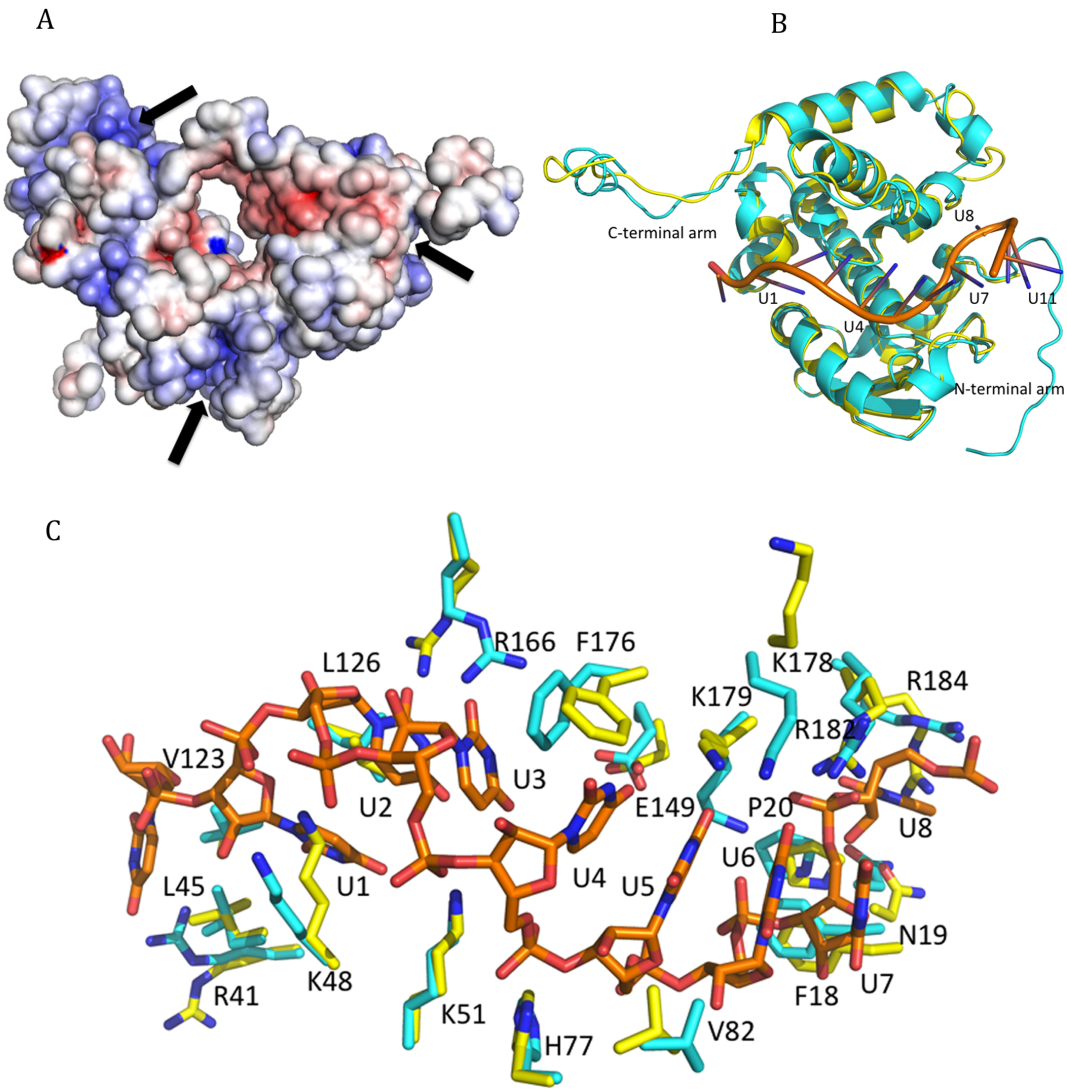


Figure 5.10 Refolded protein undergoes conformational changes upon RNA binding. A, Electrostatic surface potential map of trimetric SBV NP refolded. The arrows point to the exposed RNA binding clefts within each subunit. The positive charge is in blue and negative charge is in red. B, Overall conformational changes between the protomer of refolded SBV NP (yellow) and refolded SBV NP in complex (cyan) with RNA (orange). C, The residues located on RNA-binding cleft undergo conformational changes for RNA binding. The residues shown on yellow before RNA binding and after RNA binding show in cyan (Figure adapted from H. Dong et al. 2013).

5.4 Discussion and conclusion

The genomic RNA of negative- stranded viruses is encapsidated by nucleoprotein to form a ribonucleoprotein complex, which act as templates for genomic RNA replication and transcription (Elliott et al. 2013).

In chapter 4, a putative RNA-binding cleft was identified between the N-terminal and C-terminal domains. The N-terminal arm is crucial for RNA binding, which is confirmed by our mutagenesis and RNA binding studies. In this chapter, we have extended this work to report the crystal structure of the SBV NP in complex with 42-nt RNA to high resolution, which for the first time provide the new insights into details of interactions within the nucleoprotein and RNA complex of Orthobunyavirus of the *Bunyaviridae* family.

Only two nucleoprotein structures from the entire bunyavirus family have been reported to date, one nucleoprotein is from RVFV (Raymond et al. 2010; Ferron et al. 2011), and the other is nucleoprotein of CCHFV (Guo et al. 2012; Carter et al. 2012; Wang et al. 2012). Molecular details of how nucleoproteins encapsidate their cognate RNA genomes are best understood for RVFV.

RVFV is a prototypic membrane of the *phlebovirus* genera within *Bunyaviridae* family (Pepin et al. 2010). The crystal structure has been solved for both monomer (Raymond et al. 2010) (Figure 4.2 A) and hexamer forms (Ferron et al. 2011) (Figure 4.2 B). More recently, the crystal structure of RVFV-NP in complex with different lengths of single-stranded RNA or DNA revealed tetrameric, pentameric, hexameric forms of the protein structure (Raymond et al. 2012). These structures showed that linkage between adjacent subunits is mediated by highly flexible α-helical N-terminal

arm (Figure 5.11 A- C). Each protomer binds 7 nucleotides of which 4 nucleotides are sequestered in the hydrophobic binding slot, while 3 additional nucleotides bind between adjacent subunits. In the crystal structures, the flexible α-helical N-terminal arm allows formation of RVFV NP tetramers, pentamers and hexamers (Figure 5.11 A- C). These oligomers were also observed by EM analysis of the protein in solution (Figure 5.11 D) (Raymond et al. 2012).

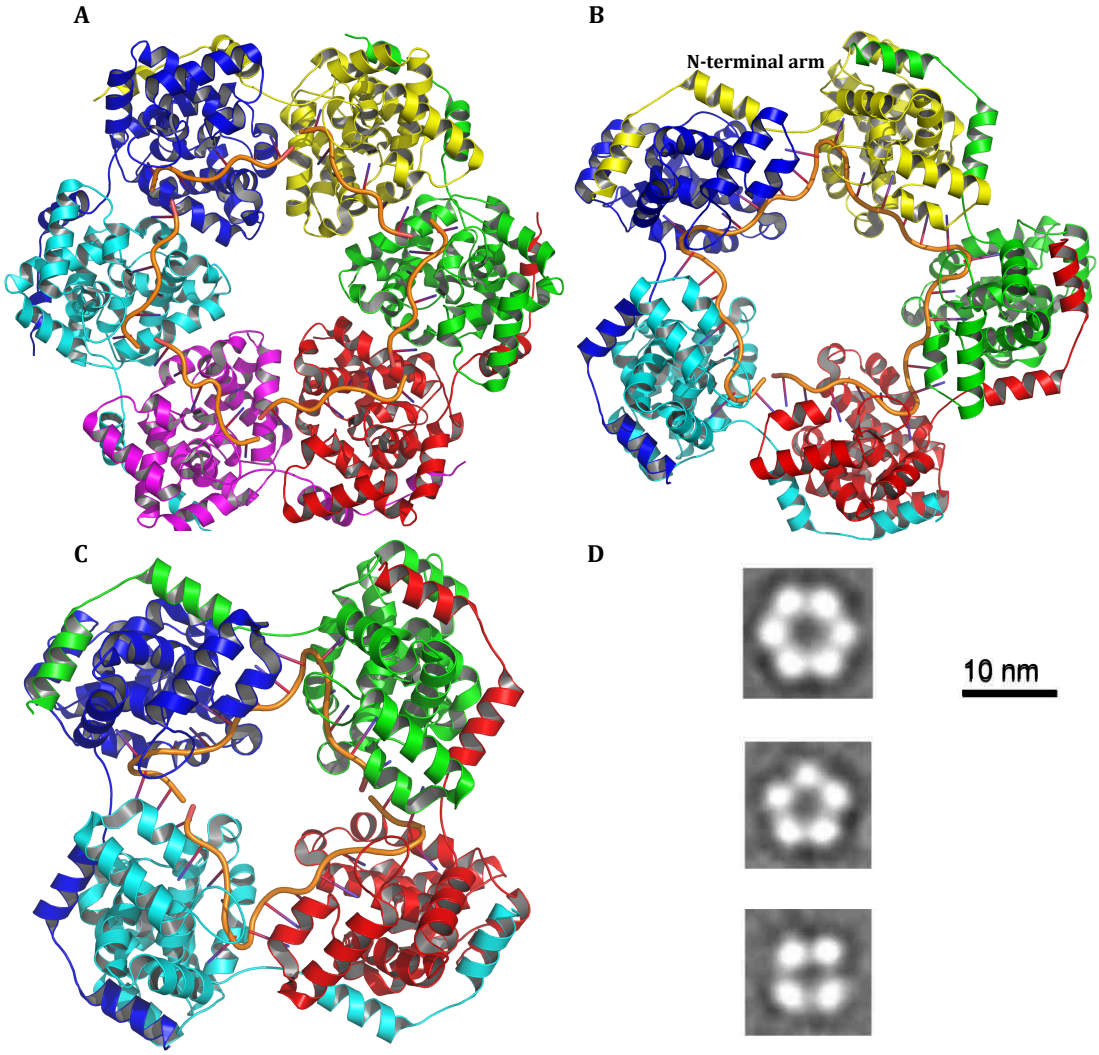


Figure 5.11 Structure of RVFV NP in complex with RNA. A, Hexameric-ring of structure in complex with 35-nt RNA (N_6-RNA_{35}) (orange stick inside the hexameric ring) in cartoon representation. The RNA were sequestered in the hydrophobic binding slot, and between adjacent subunits. The α-helical N-terminal arm in contact with adjacent subunit on the outside of the

multimers. Six subunits colored blue, yellow, green, red, magenta, cyan respectively. B, Pentameric ring of RVFV-NP in complex with 35-nt RNA (N_5-RNA_{35}). Five protomers are shown in blue, yellow, green, red and cyan respectively. C, The tetrameric ring of structure in complex with 28-nt RNA (N_4-RNA_{28}). Four protomers are shown in blue, green, red, cyan respectively. D, EM class averages of N_6-RNA_{35} (top), N_5-RNA_{35} (middle), and N_4-RNA_{28} (bottom) respectively. EM analysis shows the protein in complexes with RNAs exists in hexamer, pentamer, or tetramer in solution, which are consistent with crystal structures. The flexibility of the α-helical N-terminal arm allows the NP to form different oligomeric forms. (Figure D adapted from Raymond et al. 2012).

In contrast, CCHFV NP belongs to the *Nairovirus* genera within the Bunyaviridea family (Elliott and Schmaljohn 2013). The crystal structures of CCHFV NP has been reported (Figure 4.3) (Guo et al. 2012; Carter et al. 2012; Wang et al. 2012) and revealed a double superhelix with "head-to-tail" interaction, which suggests a positively charged RNA binding crevice on the outside of the double helix (Wang et al. 2012). However, the CCHFV NP structures were not structurally homologous with RVFV-NP, and the mechanism of nairovirus RNP assembly is still unknown.

In this study, the crystal structure of SBV NP in complex with 42-nt RNA was elucidated. SBV is an important representative of Orthobunyavirus genus (Lambert and Lanciotti 2008). Although RVFV and SBV are classified in different genera within the *Bunyaviridea* family, the two nucleoproteins RVFV (245 amino acids) and SBV (233 amino acids) have similarly size, the N-terminal arm of the nucleoprotein of RVFV is involved in interaction with adjacent protomers during oligomerization (Ferron et al. 2011; Raymond et al. 2012). In contrast, when SBV NP was complexed with single strand RNA (21, 28, 42 base in length), only protein in complex with 42-nt RNA diffracted to high resolution. This is because protein in complex with RNA (21, 28-nt) is heterogenous in nature, which limits diffraction quality. Similarly, poor diffraction was attained from large and fragile crystal of SBV NP co-purified with (*E.*

coli) RNA under native conditions. The structure of SBV NP was tetrameric and trimeric (Dong et al. 2013a), and revealed that the N and C-terminal arms were essential for multimerisation and the N-terminal arm was crucial for RNA binding. The crystal structure of the SBV NP-42-nt RNA complex is also significantly different form RVFV-NP-RNA complex. In SBV NP-42-nt RNA complex, each protomer can bind 11 nucleotides of which 8 nucleotides are located in the positively charged RNA-binding cleft. Bases U1, U2, U3, and U8 are located to facing inward to protein and bases U4- U7 were exposed to solvent. The remaining 3 nucleotides are bound to the N- terminal arms (Figure 5.8 D). Both the N- and C- terminal arms of SBV NP are highly flexible and necessary to orient the protein-RNA complex in a tetrameric ring structure by interacting with different adjacent protomers. When the crystal structure of denatured and refolded (RNA free) SBV NP was compared to the structure of the SBV NP-RNA complex, there were conformational changes within the positively charged RNA-binding cleft and also the N-terminal arm, which became ordered in protein-RNA complex structure.

In summary, the structure of SBV NP in complex with 42-nt RNA was determined to 2.16 Å where the RNA was wrapped in an RNA-binding cleft that suggests a new RNA sequestration mechanism in orthobunyavirus RNP formation. This data will provide a good foundation for understanding the mechanism of RNA encapsulation, replication and transcription of the largest Orthobunyavirus family, and this might provide an opportunity in the development of novel vaccine against diseases caused not only by SBV, but other pathogenic human and animal Orthobunyviruses.

Chapter 6

References

Bakke, A. M., Tashjian, D. H., Wang, C. F., Lee, S. H., Bai, S. C., & Hung, S. S. O. (2010). Competition between selenomethionine and methionine absorption in the intestinal tract of green sturgeon (Acipenser medirostris). *Aquatic Toxicology*, *96*, 62–69. doi:10.1016/j.aquatox.2009.09.014

Barr, J. N., & Wertz, G. W. (2004). Bunyamwera Bunyavirus RNA Synthesis Requires Cooperation of 3 J - and 5 J -Terminal Sequences, *78*(3), 1129–1138. doi:10.1128/JVI.78.3.1129

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., & Leslie, A. G. W. (2011). iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography*, *67*, 271–281. doi:10.1107/S0907444910048675

Beer, M., Conraths, F. J., & van der Poel, W. H. M. (2013). "Schmallenberg virus"--a novel orthobunyavirus emerging in Europe. *Epidemiology and Infection*, *141*(1), 1–8. doi:10.1017/S0950268812002245

Beveridge, T. J., & Davies, J. a. (1983). Cellular responses of Bacillus subtilis and Escherichia coli to the Gram stain. *Journal of Bacteriology*, *156*(2), 846–858.

Blitvich, B. J., Saiyasombat, R., Da Rosa, A. T., Tesh, R. B., Calisher, C. H., Garcia-Rejon, J. E., … Loroño-Pino, M. A. (2012). Orthobunyaviruses, a common cause of infection of livestock in the Yucatan Peninsula of Mexico. *American Journal of Tropical Medicine and Hygiene*, *87*, 1132–1139. doi:10.4269/ajtmh.2012.12-0188

Boles, J. O., Cisneros, R. J., Weir, M. S., Odom, J. D., Villafranca, J. E., & Dunlap, R. B. (1991). Purification and characterization of selenomethionyl thymidylate synthase from Escherichia coli: comparison with the wild-type enzyme. *Biochemistry*, *30*, 11073–11080.

Bos, M. P., & Tommassen, J. (2004). Biogenesis of the Gram-negative bacterial outer membrane. *Current Opinion in Microbiology*, *7*(6), 610–6. doi:10.1016/j.mib.2004.10.011

Bridgen, A., Weber, F., Fazakerley, J. K., & Elliott, R. M. (2001). Bunyamwera bunyavirus nonstructural protein NSs is a nonessential gene product that contributes to viral pathogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 664–669. doi:10.1073/pnas.98.2.664

Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, *355*, 472–475. doi:10.1038/355472a0

Brünger, A. T., Kuriyan, J., & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science (New York, N.Y.)*, *235*, 458–460. doi:10.1126/science.235.4787.458

Carter, S. D., Surtees, R., Walter, C. T., Ariza, A., Bergeron, É., Nichol, S. T., Hiscox, J.A., Edwards, T.A., & Barr, J. N. (2012). Structure, function, and evolution of the Crimean-Congo hemorrhagic fever virus nucleocapsid protein. *Journal of Virology*, *86*(20), 10914–23. doi:10.1128/JVI.01555-12

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson J.S., Richardson, D. C. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, *66*, 12–21. doi:10.1107/S0907444909042073

Chimalakonda, G., Ruiz, N., Chng, S.-S., Garner, R. a, Kahne, D., & Silhavy, T. J. (2011). Lipoprotein LptE is required for the assembly of LptD by the beta-barrel assembly machine in the outer membrane of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(6), 2492–7. doi:10.1073/pnas.1019089108

Chng, S.-S., Ruiz, N., Chimalakonda, G., Silhavy, T. J., & Kahne, D. (2010). Characterization of the two-protein complex in Escherichia coli responsible for lipopolysaccharide assembly at the outer membrane. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(12), 5363–8. doi:10.1073/pnas.0912872107

Chng, S.-S., Xue, M., Garner, R. A., Kadokura, H., Boyd, D., Beckwith, J., & Kahne, D. (2012). Disulphide rearrangement triggered by translocon assembly controls lipopolysaccharide export. *Science (New York, N.Y.)*, *337*(6102), 1665–8. doi:10.1126/science.1227215

Cowan, S. W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R. A., Jansonius, J. A., & Rosenbusch, J. P. (1992). Crystal structures explain functional properties of two E. coli porins. *Nature*, *358*, 727–733. doi:10.1038/358727a0

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D: Biological Crystallography*, *62*, 1002–1011. doi:10.1107/S0907444906022116

Cuesta-Seijo, J. A., Neale, C., Khan, M. A., Moktar, J., Tran, C. D., Bishop, R. E., Pomes, R., & Privé, G. G. (2010). PagP crystallized from SDS/Cosolvent reveals the route for phospholipid access to the hydrocarbon ruler. *Structure*, *18*, 1210–1219. doi:10.1016/j.str.2010.06.014

Davies, J. A., Anderson, G. K., Beveridge, T. J., & Clark, H. C. (1983). Chemical mechanism of the gram stain and synthesis of a new electron-opaque marker for electron microscopy which replaces the iodine mordant of the stain. *Journal of Bacteriology*, *156*, 837–845.

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L.W., Arendall, W.B., Richardson, J. S., Richardson, D. C. (2007).

MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research*, *35*. doi:10.1093/nar/gkm216

De Regge, N., Deblauwe, I., De Deken, R., Vantieghem, P., Madder, M., Geysen, D., Smeets, F., Losson, B., Van Den, B. T., & Cay, A. B. (2012). Detection of Schmallenberg virus in different Culicoides spp. by real-time RT-PCR. *Transboundary and Emerging Diseases*, *59*, 471–475. doi:10.1111/tbed.12000

DeLano, W. L. (2002). The PyMOL Molecular Graphics System. *Schrödinger LLC Wwwpymolorg*, *Version 1.*, http://www.pymol.org. doi:citeulike-article-id:240061

Delcour, A. H. (2009). Outer membrane permeability and antibiotic resistance. *Biochimica et Biophysica Acta*, *1794*, 808–816. doi:10.1016/j.bbapap.2008.11.005

Dong, A., Xu, X., Edwards, A. M., Chang, C., Chruszcz, M., Cuff, M., … Zhu, H. (2007). In situ proteolysis for protein crystallization and structure determination. *Nature Methods*, *4*, 1019–1021. doi:10.1038/nmeth1118

Dong, C., Beis, K., Nesper, J., Brunkan-Lamontagne, A. L., Clarke, B. R., Whitfield, C., & Naismith, J. H. (2006). Wza the translocon for E. coli capsular polysaccharides defines a new class of membrane protein. *Nature*, *444*, 226–229. doi:10.1038/nature05267

Dong, H., Xiang, Q., Gu, Y., Wang, Z., Paterson, N. G., Stansfeld, P. J., He, C., Zhang, Y., Wang, W., & Dong, C. (2014). Structural basis for outer membrane lipopolysaccharide insertion. *Nature*, *511*(7507), 52–6. doi:10.1038/nature13464

Dong, H., Li, P., Böttcher, B., Elliott, R. M., & Dong, C. (2013). Crystal structure of Schmallenberg orthobunyavirus nucleoprotein – RNA complex reveals a novel RNA sequestration mechanism. *RNA* 19 (8): 1129-36. doi:10.1261/rna.039057.113.5

Dong, H., Li, P., Elliott, R. M., & Dong, C. (2013). Structure of schmallenberg orthobunyavirus nucleoprotein suggests a novel mechanism of genome encapsidation. *Journal of Virology*, *87*(10), 5593–601. doi:10.1128/JVI.00223-13

Elliott, R. M., & Schmaljohn, C. (2013). Bunyaviridae. In *Fields Virology, 6th edition*.

Elliott, R. M. (1990). Molecular biology of the Bunyaviridae. *Journal of General Virology*. doi:10.1099/0022-1317-71-3-501

Elliott, R. M., Blakqori, G., van Knippenberg, I. C., Koudriakova, E., Li, P., McLees, A., Shi, X., & Szemiel, A. M. (2013). Establishment of a reverse genetics system for Schmallenberg virus, a newly emerged orthobunyavirus in Europe. *The Journal of General Virology*, *94*(Pt 4), 851–9. doi:10.1099/vir.0.049981-0

Elliott, R. M., & Weber, F. (2009). Bunyaviruses and the type I interferon system. *Viruses*. doi:10.3390/v1031003

Emsley, P., & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, *60*, 2126–2132. doi:10.1107/S0907444904019158

Evans, P. (2006). Scaling and assessment of data quality. In *Acta Crystallographica Section D: Biological Crystallography* (Vol. 62, pp. 72–82). doi:10.1107/S0907444905036693

Fairman, J. W., Noinaj, N., & Buchanan, S. K. (2011). The structural biology of β-barrel membrane proteins: A summary of recent reports. *Current Opinion in Structural Biology*. doi:10.1016/j.sbi.2011.05.005

Ferron, F., Li, Z., Danek, E. I., Luo, D., Wong, Y., Coutard, B., Walz, T., & Lescar, J. (2011). The hexamer structure of the Rift Valley fever virus nucleoprotein suggests a mechanism for its assembly into ribonucleoprotein complexes. *PLoS Pathogens*, *7*. doi:10.1371/journal.ppat.1002030

Freinkman, E., Chng, S.-S., & Kahne, D. (2011). The complex that inserts lipopolysaccharide into the bacterial outer membrane forms a two-protein plug-and-barrel. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(6), 2486–91. doi:10.1073/pnas.1015617108

Freinkman, E., Okuda, S., Ruiz, N., & Kahne, D. (2012). Regulated assembly of the transenvelope protein complex required for lipopolysaccharide export. *Biochemistry*, *51*(24), 4800–6. doi:10.1021/bi300592c

Garigliany, M. M., Bayrou, C., Kleijnen, D., Cassart, D., Jolly, S., Linden, A., & Desmecht, D. (2012). Schmallenberg virus: A new Shamonda/Sathuperi-like virus on the rise in Europe. *Antiviral Research*. doi:10.1016/j.antiviral.2012.05.014

Gibbens, N. (2012). Schmallenberg virus: a novel viral disease in northern Europe. *Veterinary Record*. doi:10.1136/vr.e292

Goller, K. V, Höper, D., Schirrmeier, H., Mettenleiter, T. C., & Beer, M. (2012). Schmallenberg Virus as Possible Ancestor of Shamonda Virus, *18*(10), 1644–1646.

Goodsell, D. S. (2005). Representing structural information with RasMol. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis. Chapter 5*, Unit 5.4. doi:10.1002/0471250953.bi0504s11

Grabowicz, M., Yeh, J., & Silhavy, T. J. (2013). Dominant negative LptE mutation that supports a role for LptE as a plug in the LptD barrel. *Journal of Bacteriology*, *195*, 1327–1334. doi:10.1128/JB.02142-12

Gram, H. C. (1884). "{U}ber die isolierte F"{a}rbung der Schizomyceten in Schnitt- und Trockenpr"{a}paraten. *Fortschritte Der Medizin*, *2*, 185–189. Retrieved from citeulike-article-id:1574067

Gu, Y., Stansfeld, P. J., Zeng, Y., Dong, H., Wang, W., & Dong, C. (2015). Lipopolysaccharide is Inserted into the Outer Membrane through An Intramembrane Hole, A Lumen Gate, and the Lateral Opening of LptD. *Structure*, *23*(3), 496–504. doi:10.1016/j.str.2015.01.001

Guo, Y., Wang, W., Ji, W., Deng, M., Sun, Y., Zhou, H., Yang, C., Deng, F., …Rao, Z. (2012). Crimean-Congo hemorrhagic fever virus nucleoprotein reveals endonuclease activity in bunyaviruses. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(13), 5046–51. doi:10.1073/pnas.1200808109

Hagan, C. L., Silhavy, T. J., & Kahne, D. (2011). β-Barrel membrane protein assembly by the Bam complex. *Annual Review of Biochemistry*, *80*, 189–210. doi:10.1146/annurev-biochem-061408-144611

Hendrickson, W. A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science (New York, N.Y.)*, *254*, 51–58. doi:10.1126/science.1925561

Hendrickson, W. A., Horton, J. R., & LeMaster, D. M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *The EMBO Journal*, *9*, 1665–1672.

Ho, B. K., & Brasseur, R. (2005). The Ramachandran plots of glycine and pre-proline. *BMC Structural Biology*, *5*, 14. doi:10.1186/1472-6807-5-14

Hoffmann, a, Bukau, B., & Kramer, G. (2010). Structure and function of the molecular chaperone Trigger Factor. *Biochimica et Biophysica Acta*, *1803*, 650–661. doi:10.1016/j.bbamcr.2010.01.017

Hoffmann, B., Scheuch, M., Höper, D., Jungblut, R., Holsteg, M., Schirrmeier, H., … Beer, M. (2012). Orthobunyavirus in Cattle , *18*(3), 469–472.

Holm, L., & Rosenström, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Research*, *38*(Web Server issue), W545–9. doi:10.1093/nar/gkq366

Hong, H., Patel, D. R., Tamm, L. K., & Van Den Berg, B. (2006). The outer membrane protein OmpW forms an eight-stranded β-barrel with a hydrophobic channel. *Journal of Biological Chemistry*, *281*, 7568–7577. doi:10.1074/jbc.M512365200

Kabsch, W. (2010). XDS. *Acta Crystallogr D Biol Crystallogr*, *66*, 125–132. doi:10.1107/S0907444909047337

Karow, M., & Georgopoulos, C. (1993). The essential Escherichia coli msbA gene, a multicopy suppressor of null mutations in the htrB gene, is related to the universally conserved family of ATP-dependent translocators. *Molecular Microbiology*, *7*, 69–79.

Khan, M. A., & Bishop, R. E. (2009). Molecular mechanism for lateral lipid diffusion between the outer membrane external leaflet and a beta-barrel hydrocarbon ruler. *Biochemistry*, *48*(41), 9745–56. doi:10.1021/bi9013566

Kim, S., Malinverni, J. C., Sliz, P., Silhavy, T. J., Harrison, S. C., & Kahne, D. (2007). Structure and function of an essential component of the outer membrane protein assembly machine. *Science*, *317*, 961–964. doi:10.1126/science.1143993

Lambert, A. J., & Lanciotti, R. S. (2008). Molecular characterization of medically important viruses of the genus Orthobunyavirus. *Journal of General Virology*, *89*, 2580–2585. doi:10.1099/vir.0.2008/002253-0

Leslie, A. G. W., & Powell, H. R. (2007). Processing diffraction data with MOSFLM. In *Evolving methods for macromolecular Crystallography* (pp. 41–51). doi:10.1007/978-1-4020-6316-9_4

Lovell, S. C., Davis, I. W., Arendall, W. B., De Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., Richardson, D. C. (2003). Structure validation by Cα geometry: φ,ψ and Cβ deviation. *Proteins: Structure, Function and Genetics*, *50*, 437–450. doi:10.1002/prot.10286

Malojčić, G., Andres, D., Grabowicz, M., George, A. H., Ruiz, N., Silhavy, T. J., & Kahne, D. (2014). LptE binds to and alters the physical state of LPS to catalyze its assembly at the cell surface. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(26), 9467–72. doi:10.1073/pnas.1402746111

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, *40*(Pt 4), 658–674. doi:10.1107/S0021889807021206

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., … Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, *67*, 355–367. doi:10.1107/S0907444911001314

Narita, S., & Tokuda, H. (2009). Biochemical characterization of an ABC transporter LptBFGC complex required for the outer membrane sorting of lipopolysaccharides. *FEBS Letters*, *583*(13), 2160–4. doi:10.1016/j.febslet.2009.05.051

Nikaido, H. (2003). Molecular Basis of Bacterial Outer Membrane Permeability Revisited Molecular Basis of Bacterial Outer Membrane Permeability Revisited, *67*(4). doi:10.1128/MMBR.67.4.593

Noinaj, N., Kuszak, A. J., Gumbart, J. C., Lukacik, P., Chang, H., Easley, N. C., Lithgow, T., & Buchanan, S. K. (2013). Structural insight into the biogenesis of β-barrel membrane proteins. *Nature*, *501*, 385–90. doi:10.1038/nature12521

Okuda, S., Freinkman, E., & Kahne, D. (2012). Cytoplasmic ATP hydrolysis powers transport of lipopolysaccharide across the periplasm in E. coli. *Science (New York, N.Y.)*, *338*, 1214–7. doi:10.1126/science.1228984

Okuda, S., & Tokuda, H. (2011). Lipoprotein Sorting in Bacteria. *Annual Review of Microbiology*. doi:10.1146/annurev-micro-090110-102859

Otwinowski, Z., & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*. doi:10.1016/S0076-6879(97)76066-X

Pepin, M., Bouloy, M., Bird, B. H., Kemp, A., & Paweska, J. (2010). Rift Valley fever virus (Bunyaviridae: Phlebovirus): An update on pathogenesis, molecular epidemiology, vectors, diagnostics and prevention. *Veterinary Research*. doi:10.1051/vetres/2010033

Phan, G., Remaut, H., Wang, T., Allen, W. J., Pirker, K. F., Lebedev, A., … Waksman, G. (2011). Crystal structure of the FimD usher bound to its cognate FimC-FimH substrate. *Nature*, *474*(7349), 49–53. doi:10.1038/nature10109

Raetz, C. R., & Dowhan, W. (1990). Biosynthesis and function of phospholipids in Escherichia coli. *The Journal of Biological Chemistry*, *265*(3), 1235–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2404013

Raetz, C. R. H., & Whitfield, C. (2002). Lipopolysaccharide endotoxins. *Annual Review of Biochemistry*, *71*, 635–700. doi:10.1146/annurev.biochem.71.110601.135414

Rasmussen, L. D., Kristensen, B., Kirkeby, C., Rasmussen, T. B., Belsham, G. J., Bødker, R., & Bøtner, A. (2012). Culicoids as vectors of Schmallenberg virus. *Emerging Infectious Diseases*. doi:10.3201/eid1807.120385

Raymond, D. D., Piper, M. E., Gerrard, S. R., Skiniotis, G., & Smith, J. L. (2012). Phleboviruses encapsidate their genomes by sequestering RNA bases. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(47), 19208–13. doi:10.1073/pnas.1213553109

Raymond, D. D., Piper, M. E., Gerrard, S. R., & Smith, J. L. (2010). Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 11769–11774. doi:10.1073/pnas.1001760107

Remaut, H., Tang, C., Henderson, N. S., Pinkner, J. S., Wang, T., Hultgren, S. J., Thanassi, D. G., Waksman, G., & Li, H. (2008). Fiber Formation across the Bacterial Outer Membrane by the Chaperone/Usher Pathway. *Cell*, *133*, 640–652. doi:10.1016/j.cell.2008.03.033

Rhodes, G. (2006). *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*. *Complementary science series* (Vol. 35, p. 352). doi:10.1002/bmb.89

Rius, J. (2011). Patterson-function direct methods for structure determination of organic compounds from powder diffraction data. XVI. *Acta Crystallographica Section A: Foundations of Crystallography*, *67*, 63–67. doi:10.1107/S0108767310047823

Rogulin, E. A., Perevyazova, T. A., Zheleznaya, L. A., & Matvienko, N. I. (2004). Plasmid pRARE as a vector for cloning to construct a superproducer of the site-specific nickase N.BspD6I. *Biochemistry (Moscow)*, *69*, 1123–1127. doi:10.1023/B:BIRY.0000046886.19428.d5

Ruiz, N., Chng, S.-S., Hiniker, A., Kahne, D., & Silhavy, T. J. (2010). Nonconsecutive disulfide bond formation in an essential integral outer membrane protein. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(27), 12245–50. doi:10.1073/pnas.1007319107

Ruiz, N., Falcone, B., Kahne, D., & Silhavy, T. J. (2005). Chemical conditionality: A genetic strategy to probe organelle assembly. *Cell*, *121*, 307–317. doi:10.1016/j.cell.2005.02.014

Ruiz, N., Gronenberg, L. S., Kahne, D., & Silhavy, T. J. (2008). Identification of two inner-membrane proteins required for the transport of lipopolysaccharide to the outer membrane of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(14), 5537–42. doi:10.1073/pnas.0801196105

Ruiz, N., Kahne, D., & Silhavy, T. J. (2009). Transport of lipopolysaccharide across the cell envelope: the long road of discovery. *Nature Reviews. Microbiology*, *7*(9), 677–83. doi:10.1038/nrmicro2184

Sali, A., Potterton, L., Yuan, F., Van Vlijmen, H., & Karplus, M. (1995). Evaluation of comparative protein modeling by MODELLER. *Proteins: Structure, Function and Genetics*, *23*, 318–326. doi:10.1002/prot.340230306

Sankaran, K., & Wu, H. C. (1994). Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol. *Journal of Biological Chemistry*, *269*, 19701–19706.

Schirmer, T., Keller, T. A., Wang, Y. F., & Rosenbusch, J. P. (1995). Structural basis for sugar translocation through maltoporin channels at 3.1 A resolution. *Science (New York, N.Y.)*, *267*, 512–514. doi:10.1126/science.7824948

Schneider, T. R., & Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Crystallographica Section D: Biological Crystallography*, *58*, 1772–1779. doi:10.1107/S0907444902011678

Shao, J., Li, M., Jia, Q., Lu, Y., & Wang, P. G. (2003). Sequence of Escherichia coli O128 antigen biosynthesis cluster and functional identification of an α-1,2-fucosyltransferase. *FEBS Letters*, *553*, 99–103. doi:10.1016/S0014-5793(03)00980-3

Sheldrick, G. M. (2007). A short history of SHELX. *Acta Crystallographica Section A: Foundations of Crystallography*. doi:10.1107/S0108767307043930

Sheldrick, G. M. (2010). Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. *Acta Crystallographica Section D: Biological Crystallography*, *66*, 479–485. doi:10.1107/S0907444909038360

Shi, X., Kohl, A., Li, P., & Elliott, R. M. (2007). Role of the cytoplasmic tail domains of Bunyamwera orthobunyavirus glycoproteins Gn and Gc in virus assembly and morphogenesis. *Journal of Virology*, *81*, 10151–10160. doi:10.1128/JVI.00573-07

Shi, X., van Mierlo, J. T., French, A., & Elliott, R. M. (2010). Visualizing the replication cycle of bunyamwera orthobunyavirus expressing fluorescent protein-tagged Gc glycoprotein. *Journal of Virology*, *84*, 8460–8469. doi:10.1128/JVI.00902-10

Silhavy, T. J., Kahne, D., & Walker, S. (2010). The bacterial cell envelope. *Cold Spring Harbor Perspectives in Biology*, *2*(5), a000414. doi:10.1101/cshperspect.a000414

Soldan, S. S., & González-Scarano, F. (2005). Emerging infectious diseases: the Bunyaviridae. *Journal of Neurovirology*, *11*, 412–423. doi:10.1080/13550280591002496

Sperandeo, P., Cescutti, R., Villa, R., Benedetto, C. Di, Candia, D., Dehò, G., & Polissi, A. (2007). Characterization of lptA and lptB , Two Essential Genes Implicated in Lipopolysaccharide Transport to the Outer Membrane of Escherichia coli Characterization of lptA and lptB , Two Essential Genes Implicated in Lipopolysaccharide Transport to the Outer Me. doi:10.1128/JB.01126-06

Sperandeo, P., Lau, F. K., Carpentieri, A., De Castro, C., Molinaro, A., Dehò, G., Silhavy, T. J., & Polissi, A. (2008). Functional analysis of the protein machinery required for transport of lipopolysaccharide to the outer membrane of Escherichia coli. *Journal of Bacteriology*, *190*(13), 4460–9. doi:10.1128/JB.00270-08

Sperandeo, P., Villa, R., Martorana, A. M., Samalikova, M., Grandori, R., Dehò, G., & Polissi, A. (2011). New insights into the Lpt machinery for lipopolysaccharide transport to the cell surface: LptA-LptC interaction and LptA stability as sensors of a properly assembled transenvelope complex. *Journal of Bacteriology*, *193*(5), 1042–53. doi:10.1128/JB.01037-10

Srinivas, N., Jetter, P., Ueberbacher, B. J., Werneburg, M., Zerbe, K., Steinmann, J., … Robinson, J. A. (2010). Peptidomimetic antibiotics target outer-membrane biogenesis in Pseudomonas aeruginosa. *Science (New York, N.Y.)*, *327*, 1010–1013. doi:10.1126/science.1182749

Suits, M. D. L., Sperandeo, P., Dehò, G., Polissi, A., & Jia, Z. (2008). Novel structure of the conserved gram-negative lipopolysaccharide transport protein A and mutagenesis analysis. *Journal of Molecular Biology*, *380*(3), 476–88. doi:10.1016/j.jmb.2008.04.045

Tarlinton, R., Daly, J., Dunham, S., & Kydd, J. (2012). The challenge of Schmallenberg virus emergence in Europe. *Veterinary Journal (London, England : 1997)*, *194*(1), 10–8. doi:10.1016/j.tvjl.2012.08.017

Taylor, G. (2003). The phase problem. In *Acta Crystallographica - Section D Biological Crystallography* (Vol. 59, pp. 1881–1890). doi:10.1107/S0907444903017815

Terwilliger, T. C. (2003). Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallographica - Section D Biological Crystallography*, *59*, 38–44. doi:10.1107/S0907444902018036

Tokuda, H. (2009). Biogenesis of outer membranes in Gram-negative bacteria. *Bioscience, Biotechnology, and Biochemistry*, *73*, 465–473. doi:10.1271/bbb.80778

Touw, D. S., Patel, D. R., & van den Berg, B. (2010). The crystal structure of OprG from Pseudomonas aeruginosa, a potential channel for transport of hydrophobic molecules across the outer membrane. *PLoS ONE*, *5*. doi:10.1371/journal.pone.0015016

Tran, A. X., Dong, C., & Whitfield, C. (2010). Structure and functional analysis of LptC, a conserved membrane protein involved in the lipopolysaccharide export pathway in Escherichia coli. *The Journal of Biological Chemistry*, *285*(43), 33529–39. doi:10.1074/jbc.M110.144709

Vagin, A., & Teplyakov, A. (2010). Molecular replacement with MOLREP. *Acta Crystallographica Section D: Biological Crystallography*, *66*, 22–25. doi:10.1107/S0907444909042589

Van Den Berg, B. (2010). Going forward laterally: Transmembrane passage of hydrophobic molecules through protein channel walls. *ChemBioChem*. doi:10.1002/cbic.201000105

Van den Berg, B., Black, P. N., Clemons, W. M., & Rapoport, T. A. (2004). Crystal structure of the long-chain fatty acid transporter FadL. *Science (New York, N.Y.)*, *304*, 1506–1509. doi:10.1126/science.1097524

Vertommen, D., Ruiz, N., Leverrier, P., Silhavy, T. J., & Collet, J. F. (2009). Characterization of the role of the escherichia coli periplasmic chaperone SurA

using differential proteomics. *Proteomics*, *9*, 2432–2443. doi:10.1002/pmic.200800794

Villa, R., Martorana, A. M., Okuda, S., Gourlay, L. J., Nardini, M., Sperandeo, P., Deho, G., Bolognesi, M., Kahne, D., & Polissi, A. (2013). The Escherichia coli Lpt transenvelope protein complex for lipopolysaccharide export is assembled via conserved structurally homologous domains. *Journal of Bacteriology*, *195*(5), 1100–8. doi:10.1128/JB.02057-12

Vollmer, W., Blanot, D., & De Pedro, M. A. (2008). Peptidoglycan structure and architecture. *FEMS Microbiology Reviews*. doi:10.1111/j.1574-6976.2007.00094.x

Von Heijne, G. (1990). The Signal peptide. *J Membr Biol*, *115*, 195–201. doi:10.1007/bf01868635

Vonrhein, C., Blanc, E., Roversi, P., & Bricogne, G. (2007). Automated structure solution with autoSHARP. *Methods in Molecular Biology (Clifton, N.J.)*, *364*, 215–230. doi:10.1385/1-59745-266-1:215

Vorou, R., Pierroutsakos, I. N., & Maltezou, H. C. (2007). Crimean-Congo hemorrhagic fever. *Current Opinion in Infectious Diseases*, *20*, 495–500. doi:10.1097/QCO.0b013e3282a56a0a

Walter, C. T., & Barr, J. N. (2011). Recent advances in the molecular and cellular biology of bunyaviruses. *Journal of General Virology*, 92, 2467-2484 doi:10.1099/vir.0.035105-0

Wang, Y., Dutta, S., Karlberg, H., Devignot, S., Weber, F., Hao, Q., … Kotaka, M. (2012). Structure of Crimean-Congo hemorrhagic fever virus nucleoprotein: superhelical homo-oligomers and the role of caspase-3 cleavage. *Journal of Virology*, *86*(22), 12294–303. doi:10.1128/JVI.01627-12

Wang, Z., Xiang, Q., Zhu, X., Dong, H., He, C., Wang, H., Zhang, Y., Wang, W., & Dong, C. (2014). Structural and functional studies of conserved nucleotide-binding protein LptB in lipopolysaccharide transport. *Biochemical and Biophysical Research Communications*, *452*(3), 443–449. doi:10.1016/j.bbrc.2014.08.094

Ward, A., Reyes, C. L., Yu, J., Roth, C. B., & Chang, G. (2007). Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(48), 19005–10. doi:10.1073/pnas.0709388104

Wernimont, A., & Edwards, A. (2009). In Situ proteolysis to generate crystals for structure determination: An update. *PLoS ONE*, 4 (4): e5094 doi:10.1371/journal.pone.0005094

Whitfield, C. (2010). Biosynthesis and Assembly of Capsular Polysaccharides. *Microbial Glycobiology*, 351–373. doi:10.1016/B978-0-12-374546-0.00020-1

Whitfield, C. (2010). Glycan chain-length control. *Nature Chemical Biology*, *6*(6), 403–404. doi:10.1038/nchembio.376

Whitney, J. C., Hay, I. D., Li, C., Eckford, P. D. W., Robinson, H., Amaya, M. F., Wood, L. F., Ohman, D. E., Bear, C. E., Rehm, B. H., & Lynne Howell, P. (2011). Structural basis for alginate secretion across the bacterial outer membrane. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 13083–13088. doi:10.1073/pnas.1104984108

Winn, M. D. (2003). An overview of the CCP4 project in protein crystallography: An example of a collaborative project. *Journal of Synchrotron Radiation*. doi:10.1107/S0909049502017235

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., … Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*. doi:10.1107/S0907444910045749

World Health Organization. (2014). The evolving threat of antimicrobial resistance: Options for action. *WHO Publications*, 1–119.

Wu, T., McCandlish, A. C., Gronenberg, L. S., Chng, S.-S., Silhavy, T. J., & Kahne, D. (2006). Identification of a protein complex that assembles lipopolysaccharide in the outer membrane of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(31), 11754–9. doi:10.1073/pnas.0604744103

Yu, X.-J., Liang, M.-F., Zhang, S.-Y., Liu, Y., Li, J.-D., Sun, Y.-L., … Li, D.-X. (2011). Fever with thrombocytopenia associated with a novel bunyavirus in China. *The New England Journal of Medicine*, *364*, 1523–1532. doi:10.1056/NEJMoa1010095

Zhang, G., Meredith, T. C., & Kahne, D. (2013). On the essentiality of lipopolysaccharide to Gram-negative bacteria. *Current Opinion in Microbiology*, *16*(6), 779–85. doi:10.1016/j.mib.2013.09.007

Zwart, P. H., Afonine, P. V, Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., … Adams, P. D. (2008). Automated structure solution with the PHENIX suite. *Methods in Molecular Biology (Clifton, N.J.)*, *426*, 419–435. doi:10.1007/978-1-60327-058-8_28