X-Ray Crystallographic and Electron Microscopy Studies on Members of the ClyA/ Nhe Family of the Pore-Forming Toxins

Avian Pathogenic *E. coli* Cytolysin A and *B. cereus* Non-Hemolytic Enterotoxin

Magdah GANASH, M.Sc., B.Sc. Hons.

A thesis submitted in part fulfilment of the
requirements for the degree of Doctor of Philosophy

Department of Molecular Biology and Biotechnology
The University of Sheffield
Sheffield

November 2012

X-Ray Crystallographic and Electron Microscopy Studies on Members of the ClyA/ Nhe Family of the Pore-Forming Toxins: Avian Pathogenic *E. coli* Cytolysin A *and B. cereus* Non-Hemolytic Enterotoxin

Abstract

Escherichia coli cytolysin A (ClyA, also known as hemolysin E, HlyE) is a 34 kDa cytolytic α-helical pore-forming toxin. The crystal structure of soluble monomeric *E. coli* K-12 ClyA was previously solved at high resolution and this showed that ClyA had a novel structure that had not previously been seen in the data bank of proteins. Avian pathogenic *E. coli* (APEC), strain JM4660 ClyA is 75% sequence identical to *E. coli* K-12 ClyA and has many significant similarities. However, two significant differences between them are that JM4660 ClyA pores are more homogeneous when observed by electron microscopy (EM), and JM4660 ClyA is more thermostable than *E. coli* K-12 ClyA. Consequently, JM4660 ClyA could be a good model system to investigate ClyA membrane interaction. The expression and purification of JM4660 ClyA was successful. Crystals were grown of the pore form, but so far diffract to about 7 Å resolution.

Bacillus cereus Nhe is a complex, pore-forming toxin consisting of three related proteins: NheA (43 kDa), NheB (39 kDa), and NheC (40 kDa). It is able to lyse mammalian cells from several organisms including humans in both Agar and liquid media and all three components of Nhe are necessary for toxicity. NheA has only 22% identity with NheB and NheC and its binding is the final stage of pore formation. Nhe proteins have sequence identity to *B. cereus* Hbl proteins. A previous structure solution of HblB revealed a structural resemblance to *E. coli* ClyA. To date, no crystal structure of any Nhe protein has been available in either the soluble form or in the pore form. Therefore, the project aimed to resolve the crystal structures of Nhe proteins in their water-soluble forms. The expression and purification of NheA was very successful. NheA was crystallized using PEG3350 as a precipitant by the sitting-drop vapour-diffusion method. The crystals belonged to space group *C2*, with unit-cell parameters a = 308 Å, b = 58 Å, c = 172.4 Å, $\alpha = \gamma = 90°$ and $\beta = 110°$ and were estimated to contain 8 protein molecules per asymmetric unit. The three-dimensional crystal structure was solved at 2.05 Å resolution using Multi-wavelength anomalous dispersion (MAD) data sets. NheA is a rod-shopped structure. The main body is formed by a bundle of four helices, each at least 70 Å long and near the end of C terminal there is a extra fifth helix known as αG. NheA has two subdomains: the tail domain and the head domain that contains amphipathic α/β hairpin. The structure of NheA reveals strong structure similarity to ClyA and HblB, which further suggests that Nhe, Hbl and ClyA belong to the some novel family of toxins.

In addition, *B. cereus* NheB expressed in *B. subtilis* strain JH642, was purified and crystallized. This thesis also presents the initial electron microscopy studies involving NheA and NheB pore formation.

Acknowledgements

I thank Allah Almighty for carrying me everyday so far and making my dream come true.

Several people have been instrumental in allowing this project to be completed. I would like to thank my supervisor: Prof. Pete Artymiuk for his wide knowledge, his guidance, encouragement and patience throughout the duration of my work. Thank you also to Prof. Jeff Green for his guidance and his important ideas. I am deeply grateful to my advisors Prof. David Rice, as he was always there to discuss about my research problems and shared his knowledge for supporting my project, and to Prof. Peter Sudbery for his advice.

I would like to thank Dr. Svetlana Sedelnikova for her guidance, for teaching me many protocols in protein purification and for working with me very hard to purify NheA and NheB; many thanks for Prof. Per Bullough and Dr. Svetomir Tzokov for their kindness and for helping me a lot to deal with electron microscopy; Dr. Arthur Moir for carrying out N-terminal sequences of my proteins; Dr. Avinash Kale for helping me in mounting the crystals and Dr. Pat Baker for his valuable insight and guidance. I want to say very a special thank you to my colleague Danh Phung; it is a pleasure to work with him and thank you to the group of Prof. Per Einar Granum, of the Norwegian School of Veterinary Science for giving me an opportunity to work in collaboration with them. Thank you for all the people who have been very helpful; thanks to Fiona, Claudia, Stephen, Glyn, Mona, Fiyzah**,** Rodrick in E36, Matt in F10 and everyone in D11 and E36.

There are no words that say thank you enough for my wonderful parents for all the unconditional love and their encouragement. Many thanks to my brothers and my sisters, especially Entesar, thank you Entesar for the laughing and the fighting, for supporting and believing in me. Thank you to my husband Abdulghani for his love and support, he always being there for me, I could not have made it through without him by my side. I express special acknowledgement to my government, Kingdom of Saudi Arabia, for scholarship.

To my baby daughter Raneem, whose love is worth it all,
you were my inspiration and my motivation.

List of terms, symbols and abbreviations

Amino acids

ALA (A)	Alanine	**LEU (L)**	Leucine
ARG (R)	Arginine	**LYS (K)**	Lysine
ASN (N)	Asparagine	**MET (M)**	Methionine
ASP (D)	Aspartic acid	**PHE (F)**	Phenylalanine
CYS (C)	Cysteine	**PRO (P)**	Proline
GLN (Q)	Glutamine	**SER (S)**	Serine
GLU (E)	Glutamic acid	**THR (T)**	Threonine
GLY (G)	Glycine	**TRP (W)**	Tryptophan
HIS (H)	Histidine	**TYR (Y)**	Tyrosine
ILE (I)	Isoleucine	**VAL (V)**	Valine

Nucleotides

A	Adenine	**C**	Cytosine
G	Guanine	**T**	Thymine

Crystallographic terms

α (hkl)- Structure factor phases.

θ- Theta angle.

φ, ψ, ω- phis, psi and omega angles, Polar angles.

Å- Angstrom.

a, b, c- Real space unit cell dimensions.

α, β, γ- Alpha, beta, and gamma the real space unit cell angles.

B factor- Crystallographic temperature factor (Debye-Waller temperature factor).

FT- Fourier transform.

F or F (hkl)- Structure factor for reflection.

|F| or |F (hkl)|- Structure factor amplitude.

$\mathbf{F_{obs}}$, $\mathbf{F_{calc}}$ - Observed and Calculated structure factor.

f - Atomic scattering factor.

f′ - The dispersive component.

f′′ - The anomalous component.

Δ*f* - Heavy atom dispersion.

hkl -Reciprocal lattice indices.

I- Intensities.

$\mathbf{I}_{hkl}$- Intensities of the reflection *hkl.*

ρ(x, y, z)- Fourier transform of a structure factor to calculate the electron density in which for every x, y and z position in unit cell.

ρ(uvw)- Patterson function.

MAD- Multi wavelength anomalous dispersion.

MR- Molecular replacement.

MIR- Multiple isomorphous replacement.

NCS- Non-Crystallographic Symmetry.

R-factor - Crystallographic refinement R-factor (also known as **R-work)**.

$\mathbf{R}_{free}$- Free R-factor.

$\mathbf{R}_{merge}$ **or** $\mathbf{R}_{sym}$ - Mergeing R- factor. R-factor relating agreement between symmetry related reflections.

RMS- Root mean square root.

RMSD- Root mean square root deviation.

SAD- Single anomalous dispersion.

SIR- Single isomorphous replacement.

σ (I)- Sigma I.

λ- X-ray wavelength.

u*, *v* and *w - Patterson cell coordinates.

V- Volume of unit cell.

V_M- Matthew's number (Volume per protein mass).

Protein abbreviations, and biological terms

αHL- *Staphylococcus* aureus α- hemolysin.

ATP -Adenosine-5'-triphosphate.

APEC- Avian pathogenic *Escherichia coli.*

B. cereus- *Bacillus cereus.*

B. subtilis- *Bacillus subtilis.*

cAMP - Cyclic adenosine monophosphate.

CDCs- Cholesterol-dependent cytolysins.

ClyA- *Escherichia coli* Cytolysin A (also known as HlyE and sheA).

CRP- cAMP receptor protein (Regulatory protein which binds cAMP).

CytK- *Bacillus cereus* Cytotoxin K.

DNA- Deoxyribonucleic acid.

E. coli- *Escherichia coli.*

ETEC- Entero-pathogenic *E. coli.*

FNR- Fumarate and nitrate reduction.

Hbl - *Bacillus cereus* Hemolysin BL.

HblB- *Bacillus cereus* Hemolysin BL component B.

HblL1- *Bacillus cereus* Hemolysin BL component L1.

HblL2- *Bacillus cereus* Hemolysin BL component L2.

HlyA- *Escherichia coli* α-hemolysin.

HlyE- *Escherichia coli* hemolysin (also known as ClyA and sheA).

HlyX- Homologues from the FNR family.

HRP- Horseradish peroxidase.

IHF and H-NS- Two histone-like proteins.

LDH -Lactate dehydrogenase.

LLO- listeriolysin O.

LukF- *Staphylococcus aureus* Leukocidin F.

MAC- The complement membrane attack complex.

MscL- Mechanosensitive ion channel from *Mycobacterium tuberculosis*.

mRNA- Messenger ribonucleic acid.

Nhe - *Bacillus cereus* non hemolytic enterotoxin.

NheA- *Bacillus cereus* non hemolytic enterotoxin component A.

NheB- *Bacillus cereus* non hemolytic enterotoxin component B.

NheC- *Bacillus cereus* non hemolytic enterotoxin component C.

PH0828- A hypothetical protein from *Pyrococcus horikoshii OT3*

PF- Perforin.

PFO- Perfringolysin O.

PFT- Pore forming toxin.

RBC- Red blood cell.

RTX- Repeats in toxin family.

S. aureus - *Staphylococcus* aureus

SheA- *Escherichia coli* silent hemolysin A (also known as ClyA and HlyE).

SNARE -SNAP receptor, these proteins act as receptors for NSF (N-ethylmaleimide Sensitive Fusion protein) and SNAPs (Soluble NSF Attachment Protein).

Miscellaneous terms

° - Degrees.

°C - Degrees centigrade (temperature scale).

2D- Two-dimensional.

3D- Three-dimensional.

Å- Ångstrom.

ß-OG -ß-octylglucoside.

BSA -Bovine Serum Albumin

Bis Tri- (Bis(2-hydroxyethyl)-amino-tris(hydroxymethyl)-methane).

BL- Brain total lipid extract.

BLAST- Basic local alignment search tool, available on http://blast.ncbi.nlm.nih.gov/Blast.cgi.

bp- base pairs.

CCD- Charged Couple Device.

CCP4- Collaborative Computational Project, Number 4.

DDM-n-dodecyl beta-D-maltoside.

DDT-Dichlorodiphenyltrichloroethane.

DM- n-Decyl-maltoside.

EL- *Escherichia coli* total lipid extract.

EM- Electron microscopy.

EMP-Ethylmercury phosphate.

ExPASy- Expert Protein Analysis System; available on http://www.expasy.ch/.

h-hours.

HA- Hydroxylapatite Chromatography.

HEPES- (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid).

His tag- Hexa histidine-tag**.**

IPTG- Isopropyl-1-thio-B-D-galactopyranoside.

IP- Isoelectric points.

K- Kelvin (temperature scale).

kDa- KiloDalton.

keV- kiloelectron volt

LB- Luria-Bertani broth.

LC-MS- Liquid chromatography–mass spectrometry.

MAb- Monoclonal antibodies.

MES - 2-(N-morpholino)ethanesulfonic acid.

min- minutes.

MPD- 2-Methyl-2, 4-pentanediol.

MW- Molecular weight.

Ni-NTA- Nickel Nitrotriacetic acid.

NMR- Nuclear Magnetic Resonance.

OD- Optical density.

PEG- Polyethylene glycol.

rpm- Revolutions per minute.

SDS- PAGE- Sodium dodecyl sulfate polyacrylamide gel electrophoresis.

Sec or s- seconds.

SeMet- Selenomethionine.

SEM- Scanning electron microscopy.

TEM- Transmission electron microscopy.

TEMED- N,N,N',N'-tetramethylethylenediamine.

Tris- tris (hydroxymethyl) methylamine.

UV- Ultraviolet.

V- Volt.

v/v- volume by volume.

w/v- weight by volume.

XDS -X-ray Detector Software.

Chapter 1: Introduction

Bacteria are small micro-organisms, typically 1-10 μm in diameter, that are found in all environments (Turner et al, 2005). Bacterial pathogens produce toxins which destroy and lyse host cells including human and animal cells (Andreeva et al, 2006). Schmitt et al, (1999) have defined a bacterial toxin as "a soluble substance that alters the normal metabolism of host cells with a deleterious effect on the host". Depending on their mechanism of action, toxins can be classified into various groups such as membrane damaging, affecting protein synthesis, activation of second messenger pathways, inhabitation of the production of neurotransmitters, or activation of an autoimmune response (Schmitt et al, 1999).

1.1 Membrane damaging toxins

Pathogenic bacteria are able to lyse cells and damage the membrane of the target cells by two main mechanisms (Linder & Bernheimer, 1991). In the first mechanism, the cell surface may be enzymatically damaged by disrupting a component of the lipid bilayer. One well-characterised toxic enzyme is phospholipase, which is able to cleave phospholipids at the junction between the charged polar group (head group) and the long fatty acyl chains (Titball, 1999). The head group is responsible for stabilizing the lipid bilayer of the membrane. Thus, the cell membrane becomes leaky and finally lyses.

In the second mechanism, the toxins insert into the membrane and form functional trans-membrane pores (Bhakdi & Tranumjensen, 1991). Therefore, this group of toxins is known as pore forming toxins. This thesis mainly focus in two related pore forming toxins from different species of bacteria: the first one is Cytolysin A (ClyA) from avian pathogenic *Escherichia coli,* strain JM4660 and the second one is the non-hemolytic

enterotoxin (Nhe) from *Bacillus cereus*, strain NVH 0075/95.

1.2 Pore forming toxins (PFTs)

Pore Forming Toxins (PFTs) are one of the major groups of bacterial toxins (Ludwig et al, 1999) that are able to cause membrane damage. PFTs constitute approximately a quarter of all known bacterial protein toxins (Aroian & van der Goot, 2007) and are produced by many different organisms including both gram-positive and gram-negative bacterial species (Rosado et al, 2008) . Gonzalez et al, (2008) have shown that PFTs can be essential virulence factors of a pathogen. PFTs are released by pathogens as soluble, monomeric proteins (protomers) and then transformed to integral membrane proteins (oligomers) to form trans-membrane channels. Thus, they must undergo large conformational changes to create new hydrophobic surfaces that able to enter the core of lipid bilayer of the host membrane (Parker & Feil, 2005). PFTs destroy the plasma membrane by either delivering a toxic compound through the pore or disrupting the selective efflux and influx of ions through host membranes (Ludwig et al, 2004; Tilley et al, 2005) - in other words water diffuses into the cytoplasm, causing the cell to swell and burst (Figure 1.1). The ability of PFTs to convert, in a controlled manner, from a water soluble to a transmembrane form is common in this type of toxins and constitutes one of their most significant features (Bischofberger et al, 2009; Tilley et al, 2005). The mechanism of pore formation is still not understood totally but all pore forming toxins require the following steps: export into the medium, binding to the host cell, penetration into the membrane phospholipid bilayer and then finally formation of a trans-membrane channel (Menestrina et al, 1995). PFTs can be categorized according to their size or to the category of the structure of PFTs utilized to insert into plasma membrane (Feil et al, 2010).

According to the secondary structure that inserts into the phospholipid bilayer of the plasma membrane, bacterial pore forming toxins have been

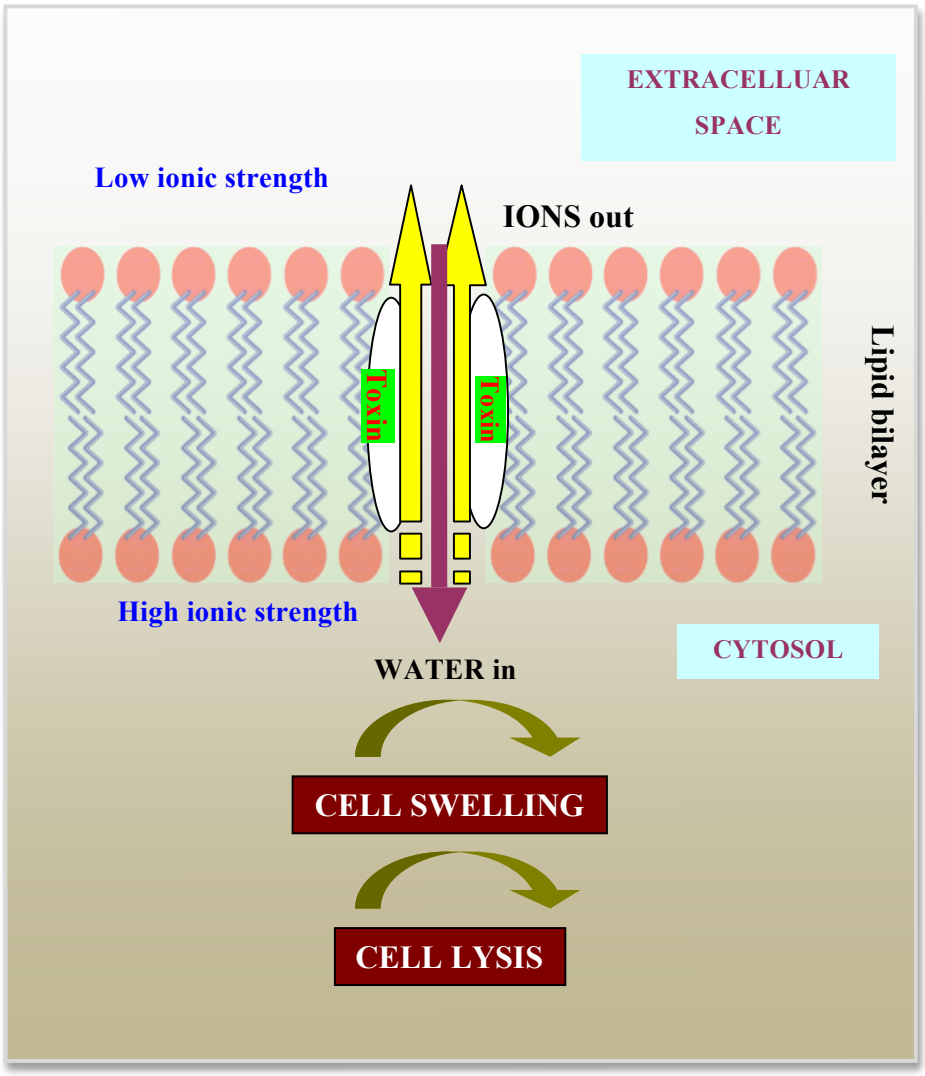


Figure 1.1: Diagram illustrating the mechanism of a PFT lysing host cells.
PFT causes leakage of ions and entry of water into the cell.

classified into two divisions (Gilbert, 2002; Tilley et al, 2005) (Figure 1.2): α–PFTs, which form helical channels made up of α–helices; and β–PFTs, which form β sheet structures such as β barrels. The formation of a pore involves three steps: binding to a specific receptor, unfolding of the pore forming part of the protein and insertion into the membrane (Tilley et al, 2005). β-PFTs require refolding during insertion in the membrane so they can form transmembrane β barrels. (Tilley et al, 2005). In the two classes of PFTs, the hydrophobic amino acids of the protein face the acyl chains of phospholipid of the membrane bilayers while the hydrophilic amino acids face the aqueous environment of transmembrane channel (Montoya & Gouaux, 2003).

Most PFTs are homo-oligomeric, for example *Staphylococcus* aureus α-hemolysin (Montoya & Gouaux, 2003). However, hetero-oligomeric PFTs also exist, for example staphylococcal γ-hemolysin and the related leukocidins, which form binary toxin complexes (Prevost et al, 1995) or tripartite PFTs such as non-hemolytic enterotoxin (Nhe) (Granum et al, 1999; Lund & Granum, 1996), which is one of the subjects of this thesis.

1.3 An overview of some pore forming toxins

1.3.1 Repeats in toxin family (RTX): *Escherichia coli* α-hemolysin (HlyA)

RTX toxins are featured by a specific structure known as the glycine and asparate-rich nonapeptide motif which is near the carboxy-terminal portion of the protein (Cortajarena et al, 2002; Goñi & Ostolaza, 1998; Linhartova et al, 2010), and which is repeated between 9 to 42 times (Goñi & Ostolaza, 1998). An additional significant characteristic of the RTX family is that the operon comprises four genes *hlyC, A, B,* and *D* in transcriptional order (Boehm et al, 1990). This means that all genes coding the synthesis, activation, and secretion are located in the same operon.

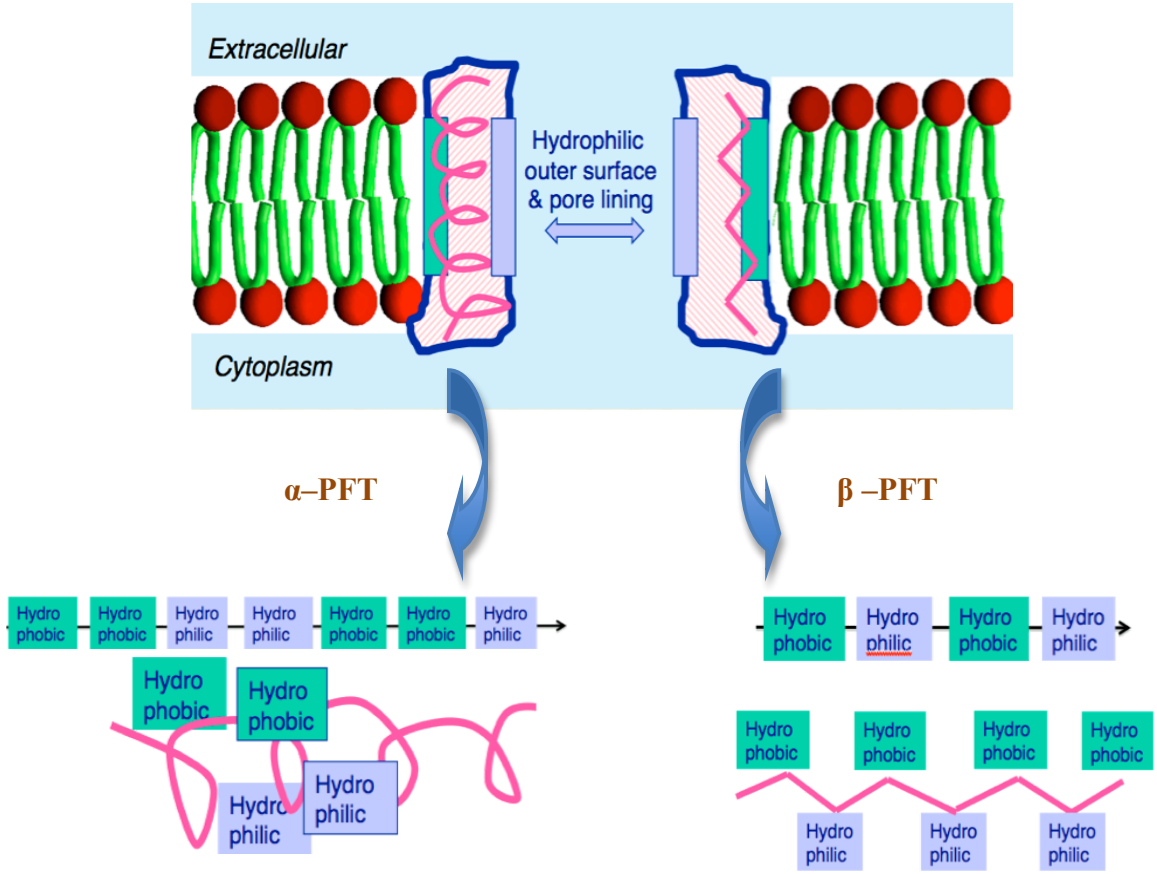


Figure 1.2: The classification of a PFT according to the secondary structure that inserts into the membrane: α–PFTs that form amphipathic helical channels and β–PFTs that form amphipathic β sheet structures such as β barrels.
The diagram was drawn by Prof. Pete Artymiuk (Department of MBB/University of Sheffield/United Kingdom).

The RTX toxins have similar structures that include a receptor domain at the C–terminal and a pore forming domain at the N-terminal. The receptor domain is dependent on the addition of fatty acyl groups and the binding of calcium ions to glycine and asparate in the C-terminal domain. The N-terminal domain is predicted to form pores by the insertion of four-transmembrane α-helical segments into the lipid bilayer. The sequences preceding and between these α-helices have strong amphipathic α-helical properties and possess membrane-binding features. For that reason, it is believed that RTX may be classified as a α-PFT (Stanley et al, 1998). The regulation of pore stability is flanked by two regions: a central domain containing glycine-rich repeated nonapeptide which binds calcium ions and the C-terminal export region (Menestrina et al, 1995).

The pore-forming HlyA of *Escherichia coli* is the best known example of an RTX toxin (Cortajarena et al, 2002). This protein is an extracellular toxin with a molecular mass of 107 kDa (Cortajarena et al, 2002; Menestrina et al, 1995) and homologues are secreted by gram-negative pathogenic strains, e.g. *Escherichia*, *Proteus*, *Morganella*, *Pasteurella*, *Actinobacillus* and *Bordetella* (Cortajarena et al, 2001). In order to lyse cells, HlyA requires post-translational modification and this requires two steps:

1) Intracellular activation by specific fatty acylation of the structure *hlyA* gene product, which requires the association of the *hlyC* gene product as well as an acyl carrier protein.
2) Then extracellular activation takes place through binding of calcium ions (Goñi & Ostolaza, 1998; Stanley et al, 1994). Binding of calcium ions lead to the formation of short β–strands known as a parallel β-barrel or β–super-helix (Stanley et al, 1998).

1.3.2 *Staphylococcus* aureus α-hemolysin

Staphylococcus aureus α-hemolysin (αHL) is a hemolytic protein that classified as a β-PFT and has a molecular mass of 33 kDa in its soluble

monomeric form and 232.4 kDa on pure lipid bilayers or on biological membranes (Gouaux, 1998; Panchal & Bayley, 1995).

The structure of αHL was determined to 1.9 Å resolution by Song et al, (1996), using X-ray crystallography. This showed that the transmembrane channel has a mushroom-like shape consisting of a cap domain, seven rim domains and a stem domain (Figure 1.3). The pore is a homo-oligomeric heptamer with general dimensions of 100 Å x 100 Å and the channel varies in diameter from 14 to 46 Å. To span the membrane the stem region in this heptameric complex, transmembrane region, uses two β strands from each protomer to form a 14-stranded antiparallel β sheet. The β-barrel is amphiphatic and contains hydrophilic and hydrophobic residues so that when the pore is formed the hydrophilic residues make the lumen of the pore whereas the hydrophobic residues interact with the membrane lipid (Feil et al, 2010).

1.3.3 Cholesterol-dependent cytolysins (CDCs): perfringolysin O (PFO)

CDCs are a superfamily of pore-forming proteins which are secreted by different species e.g. genera *Clostridium*, *Streptococcus*, *Listeria*, *Bacillus*, and *Arcanobacterium* (Tweten, 2005). CDCs also include the human proteins perforin (PF) and the complement membrane attack complex (MAC) (Gilbert 2010; Rosado et al, 2008). The formation of oligomeric transmembrane pores are dependent on the presence of cholesterol-rich membranes in all CDCs (Polekhina et al, 2005). The significant sequence similarity between CDCs (40-80%) suggests that they all have similar mechanism of pore formation and 3D structures (Polekhina et al, 2005).

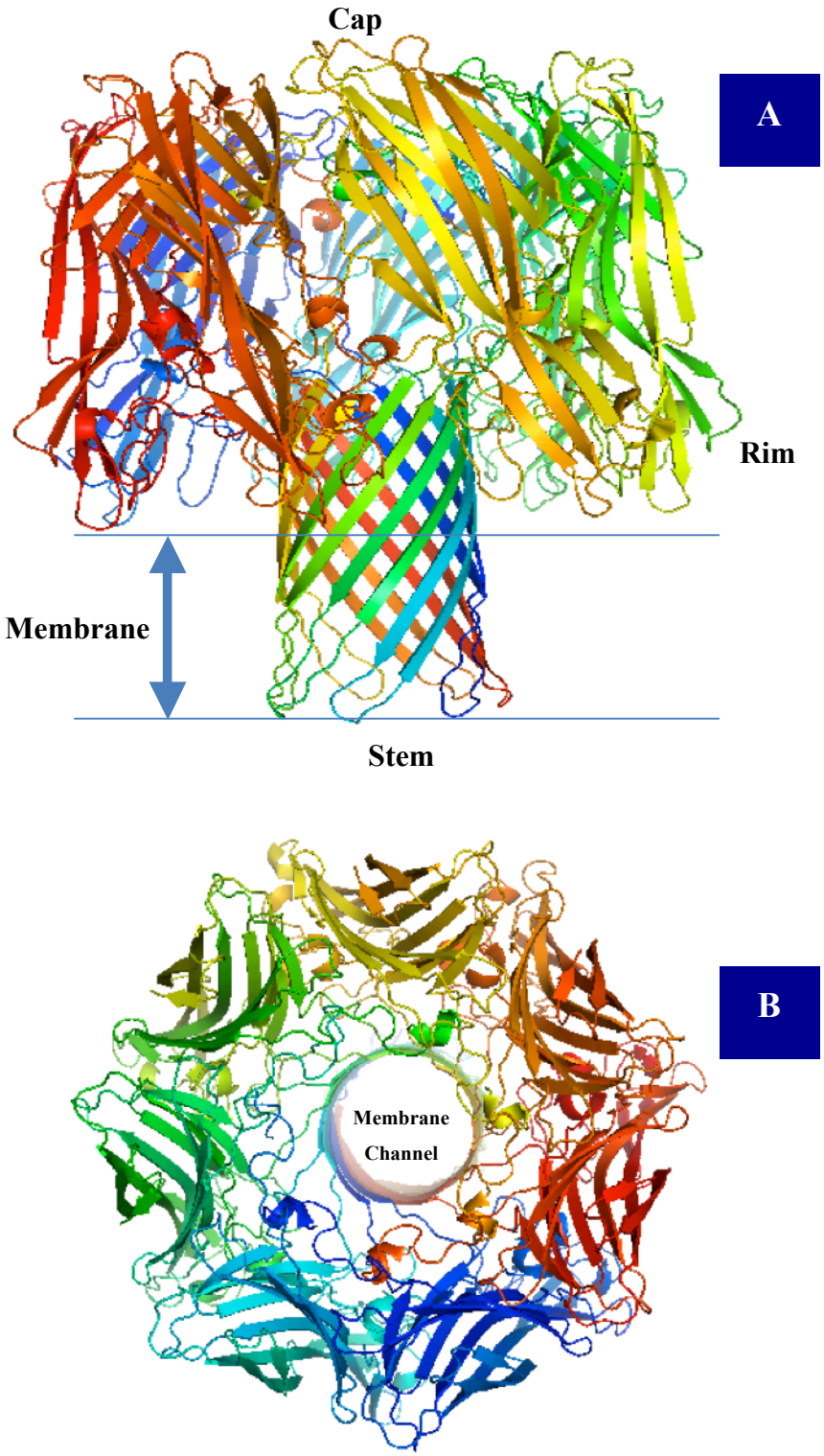


Figure 1.3: Ribbon image of the structure of the protein α-hemolysin heptamer. (A) From the side. (B) From above. Each protomer is a different color. The cap domain, seven rim domains and stem domain are labeled (Song et al, 1996).
Produced using PyMol (DeLano and Lam, 2005).

The first 3D structure of a CDC toxin was that of perfringolysin O (PFO) which was determined by multiple isomorphous replacement (MIR) (Rossjohn et al, 1997). PFO is secreted in a water-soluble form by the pathogen *Clostridium perfringens* (Dang et al, 2005). It has a molecular mass of 53kDa and is composed of 500 amino acid residues.

PFO has an elongated rod-shaped appearance (approximately 115 Å × 30 Å × 55 Å) and like the other β-PFTs, the monomeric structure of PFO is rich in β sheet and consists of four domains (Figure 1.4): the first domain is comprised of an α/β structure which, containing a 7-stranded antiparallel β sheet, the second domain has 4 β strands, the third domain consists of an α/β/α three layer structure, and the last domain consist of a β-sandwich topology (Rossjohn et al, 1997).

A consensus model of the PFO pore suggests that the membrane-bound oligomer has between 300 Å and 450 Å outer diameter, depending on the number of subunits of the pore which varies approximately between 40-50 sub-units. Side views of the pore show that it forms a mushroom shape, between 90 Å and 100 Å in length (Rossjohn et al, 1997).

1.4 *Escherichia coli* cytolysin A (ClyA)

Cytolysin A (ClyA) is an α-helical pore-forming toxin. It is a hemolytic protein with approximate molecular weight 34 kDa, which was originally recognized in *E. coli* K-12 and is able to produce pores in the cell membranes of eukaryotic cells (Wallace et al, 2000). The crystal structure of the soluble form showed that ClyA had a novel structure that did not occur elsewhere in the data bank of proteins (Wallace et al, 2000). Accordingly, ClyA revealed the molecular architecture of a new family of toxins.

The ClyA protein has a number of names: ClyA (cytolysin A), SheA (silent hemolysin A), HlyE (hemolysin E) because the earliest literature referring to the gene and gene product was produced by different independent research

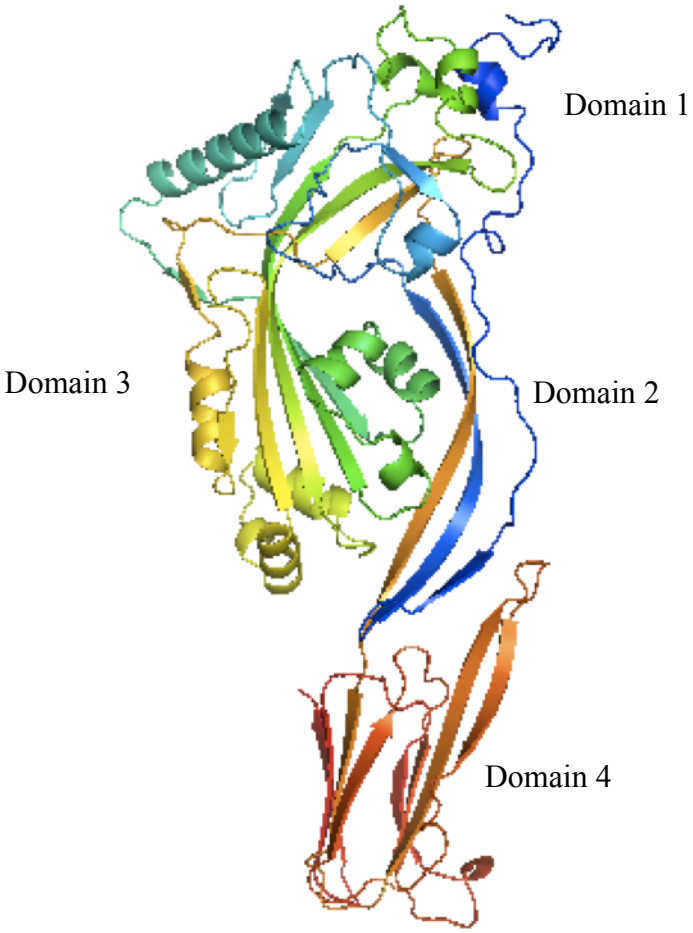


Figure 1.4: 3D Crystal structure of the PFO monomer.
The structure is revealed in ribbon representation with the four domains (Rossjohn et al, 1997). Produced using PyMol (DeLano and Lam, 2005).

Groups (del Castillo et al, 1997; Green & Baldwin, 1997; Oscarsson et al, 1996). Thus the same gene had three names *hlyE, clyA*, and *sheA*. This gene is located at position 26.4 min on the *E. coli* K-12 chromosome (GenBank accession no. U57430; (Green & Baldwin, 1997; Westermark et al, 2000). In this thesis the name ClyA will be used.

1.4.1 Biological activity of *Escherichia coli* ClyA

The biological activity of ClyA can be summarized as follows:

1) Lysis of mammalian erythrocytes from various organisms, including human, horse, sheep, goat, and hen (del Castillo et al, 1997; Oscarsson et al, 1996; Yadav et al, 2008).

2) Provoking apoptosis in leukocytes especially in macrophages (Oscarsson et al, 1996). Lai et al, (2000) showed that ClyA induced apoptosis in both human and murine macrophages showing host cell DNA fragmentation. Consequently, it destroys the immune cells of the host organism causing tissue damage. Oscarsson et al, (1999) found 0.9 μg/ml of ClyA was sufficient to induce cytotoxicity measured by release of Lactate dehydrogenase (LDH).

3) Induced disruption of the amount of intracellular Ca^{2+} in epithelial cells by induction of slow, intracellular Ca^{2+} oscillations (Soderblom et al, 2005).

1.4.2 The homologues of *Escherichia coli* ClyA

Certain *E. coli* laboratory strains are considered as non-pathogenic and safe organisms (Rowe & Welch, 1994). Despite this long-standing belief there is a gene that encodes a hemolytic protein in *E. coli* K12 (del Castillo et al, 1997). In fact, it seems that most non-pathogenic *E. coli* strains carry this gene and have the ability to express cytotoxic protein (Westermark et al, 2000).

ClyA is not related to the well-known pore forming *E. coli* hemolysins of the RTX family, e.g. *E. coli* HlyA (Coote, 1992; Oscarsson et al, 1999; Oscarsson et al, 2002). ClyA requires no post-translational processing to become active, whereas RTX toxins are created firstly as a protoxin then activated to produce mature toxins (Stanley et al, 1994). Furthermore, ClyA does not need Ca^{2+} ions to be translocated to the periplasm to become activated, and it does not have sites for conserved fatty acylation or signals for C-terminal secretion as RTX toxins do (del Castillo et al, 1997; Oscarsson et al, 2002).

The *clyA* gene is present in all *E. coli* strains including pathogenic strains of *E. coli*, O157:H7 (del Castillo et al, 1997), and porcine ETEC (Entero-pathogenic *E. coli*) strain 3030-2 (Fernandez et al, 1998) as well as in the dysentery causing organism *Shigella flexneri* (del Castillo et al, 1997), and the typhoid fever causing bacteria *Salmonella enterica* serovars Typhi and Paratyphi (del Castillo et al, 2001). However the gene locus is apparently absent in *Salmonella Klebsiella pneumoniae or Serratia marcescens* (del Castillo et al, 1997) and other *Salmonella* serovars, including Typhimurium (Oscarsson et al, 2002).

These results show that there is a family of ClyA-like hemolysins and they are likely to be an important component of these organisms armory of toxins. Figure 1.5 summarizes the homologus ClyA proteins from members of the family of enterobacteriaceae.

1.4.3 Regulation of ClyA expression

1.4.3.1 positive regulation of *clyA*

The *clyA* gene is silenced under normal laboratory growth conditions in *E. coli* K-12. The expression of *clyA* is regulated by both negative and positive controls in *E. coli*. ClyA is expressed during anaerobic growth of *E. coli* when the amounts of oxygen and glucose levels are low. This expression is affected positively by the cAMP receptor protein (CRP).

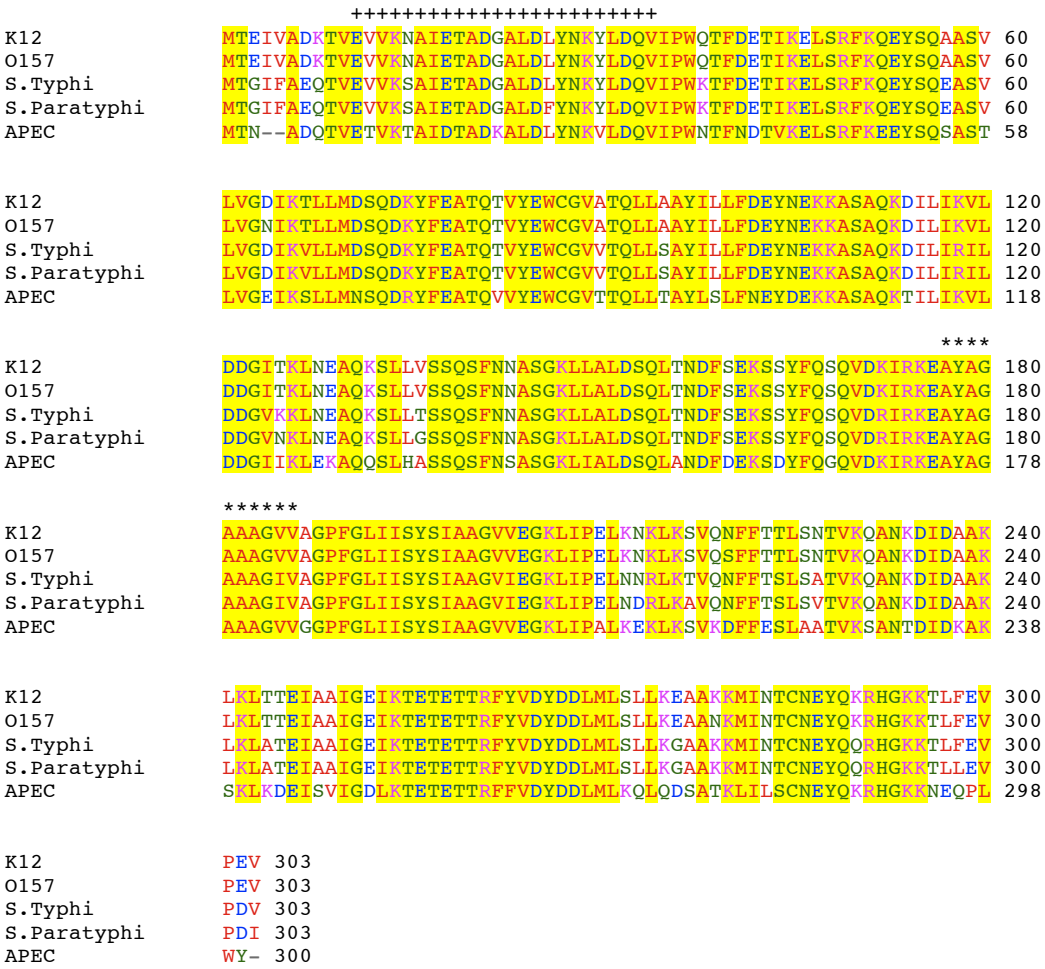


Figure 1.5: Sequence alignment of ClyA proteins.
Amino acid sequence alignments are shown for ClyA from *E. coli* K-12, *E. coli* O157, *E. coli* JM4660 (APEC), *Salmonella typhi*, and *Salmonella paratyphi*. Aliphatic/hydrophobic residues are colored red; Acidic residues are colored blue; basic residues are colored magenta; Hydroxyl, sulfhydryl, amine, and G residues are colored green. Conserved residues are shown in yellow background. The hydrophobic transmembrane sequence is shown by the (*) above the sequence. The amphipathic (αA1), transmembrane region of ClyA-K12, is indicated by the (+) above the sequence. The amino acid sequence alignment was generated using ClustalW2 (Goujon, et al. 2010& Larkin, et al. 2003) (available on: http://www.ebi.ac.uk/Tools/msa/).

The homologous transcription factor of fumarate and nitrate reduction (FNR) also positively controlled the expression of ClyA in the absence of oxygen (Green & Baldwin, 1997). Homologues from the FNR family such as HlyX, which is encoded by the *hlyx* gene of the pig pathogen *Actinobacillus pleuropneumoniae,* are able to regulate positively a hemolytic phenotype on *E. coli* (Green & Baldwin, 1997; MacInnes et al, 1990; Ralph et al, 1998). The anaerobic expression of ClyA can also be controlled by *Pasteurella multocida ahpA* gene or *mesA* in *E. coli* (Cox et al, 2000; Hunt et al, 2000). Positive regulation of *clyA* can also be induced by *Salmonella typhimurium* $SlyA_{ST}$ by binding to the *clyA* promoter region (Oscarsson et al, 1996). Other $SlyA_{ST}$ homologous proteins such as $SlyA_{EC}$ and MprA (also known as EmrR), which have bacterial regulatory function in *E. coli* also activated *clyA* expression (del Castillo et al, 1997; Ludwig et al, 1995; Oscarsson et al, 1996).

1.4.3.2 Negative regulation of *clyA*

Two histone-like proteins, IHF and H-NS, appear to control the negative regulation of *clyA* (del Castillo et al, 1997). H-NS is able to repress many genes in *E. coli* and so it seems that it may indeed repress the hemolysin of *E. coli* K-12 (GomezGomez et al, 1996). It was shown that a mutant lacking H-NS exhibits a hemolytic phenotype (GomezGomez et al, 1996). According to DNA bending analysis, the *clyA* promoter has high intrinsic curvature and it was concluded that this represents a control region that may be susceptible to H-NS silencing (Westermark et al, 2000).

1.5 The crystal structure of *E. coli* K 12 ClyA

1.5.1 Crystal Structure of *E. coli* K 12-ClyA as monomeric form

The three–dimensional structure of the ClyA monomer in its water-soluble form was solved by using X-ray crystallography at 2.0 Å resolution

(Wallace et al, 2000) and the major points are discussed below:

ClyA belongs to a novel class of bacterial toxins and it is an elongated, rod - shaped molecule with approximate dimensions 100 Å x 30 Å x 20 Å.

The toxin mainly consists of a long helical bundle: the main part consists of four α helices: αA1/ αA2 with distinct kink at residue PRO 36, α B, α C, and α F. Each helix is approximately 70-80 Å long and they are coiled around each other with a simple up-down-up-down topology (Figure 1.6).

However, ClyA has significant novel structure elaborations at both ends:

At the lower end of the ClyA monomer (also known as the tail domain) there is a fifth G helix at the C-terminal. The G helix consists of 24 residues and is approximately 35 Å long and adds a fifth helix to the bundle and it is located in a space between helices A and B. The tail domain also hosts two cysteine residues, which are located on the B and G helices and which are sufficiently close to form a disulfide bond. However, the γ-sulphydryl atoms of two cysteine residues (CYS 87, 285) are separated by 4.1 Å, which is longer than the expected 2.2 Å for a disulphide bond.

Replacement of CYS by SER in ClyA shows that neither is necessary for the structure to cause hemolytic activity but the two cysteines seem to contribute to the thermostability of ClyA. The incorporation of a disulphide bond within ClyA caused a decrease in hemolytic activity, hence suggesting that a disulphide bond may inhibit pore formation by preventing premature oligomerisation of ClyA (Atkins et al, 2000).

Wai et al, (2003) found that most of ClyA in the periplasm was monomeric and in an oxidised form, whereas ClyA exported in outer-membrane vesicles (OMVs) was oligomeric and present in the reduced form. This led to further speculation that oligomerisation to form active pores is dependent on redox status (Wai et al, 2003). In spite of this, another research group Eifler et al, (2006) found that the redox state of ClyA had no effect on oligomerisation or activity of ClyA as they found ClyA remained monomeric in any case of redox state and that both forms of ClyA, reduced and oxidised, have similar

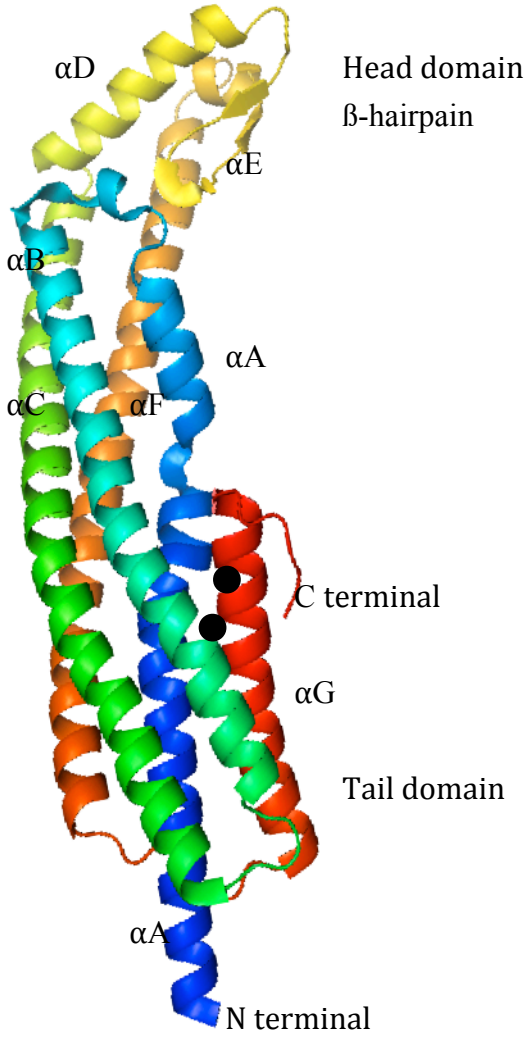


Figure. 1.6: The 3D crystal structure of monomeric soluble *E.coli* K-12 ClyA consist of four α–helices (αA, αB, αC and αF), two Cysteine residues (black circles), and α ß-hairpain in the head domain. The C and N terminal and αG helix in the tail domain are labelled (Wallace et al, 2000). Produced using PyMol (DeLano and Lam, 2005).

activity against horse erythrocytes (Eifler et al, 2006).

The second novel elaboration, the upper end (also known as the head domain) of ClyA consists of a short ß hairpin (the "ß tongue") flanked by two shorter helices D and E. The ß tongue packs between helices C and, F. The ß hairpin contains the main hydrophobic region on the ClyA surface. Hence, Wallace et al, (2000) suggested that the ß tongue interacts with lipid membrane.

Wallace et al, (2000) stated that the most significant region of the amino acid sequence of ClyA is between residues 177 and 203, which are hydrophobic. In addition, there is a shorter hydrophobic region from residues 89 to 101. It has been suggested that the longer hydrophobic region in ClyA plays an important role in the insertion of toxin into cell membranes and it was predicted as transmembrane helix.

1.5.2 Models for pore formation

Wallace et al, (2000) proposed a conservative preliminary, octameric model for pore formation after describing the 3D crystal structure of ClyA. It was proposed that the flat face of the head domains “β tongue” would interact with the hydrophobic lipid bilayer with no major conformational change to allow ClyA to form a pore.

By using cryo-electron microscopy and single particle analysis Eifler et al, (2006) visualized the oligmeric pore structure at 12 Å resolution. This showed the ClyA pore form appeared to be an assembly of 13 monomers of ClyA. They suggested that the head domain “β tongue” of ClyA could form a 26-stranded β barrel cap structure during membrane insertion. However, in direct contradiction of this, Tzokov et al, (2006) showed the pores appeared to the pores consist of 8 ClyA subunits.

Both Eifler et al, (2006) and Tzokov et al, (2006) found that the pore assemblies are much longer than the soluble ClyA structure (Wallace et al,

2000) which suggested that conformational changes of ClyA monomers must take place.

Hunt et al, (2008) also showed that there are rate limiting conformational changes when ClyA forms a pore and they proposed that conformational changes of ClyA are expected when the protein binds to the lipid bilayer of the plasma membrane.

1.5.3 Crystal Structure of ClyA *E. coli* K 12 pore form

After the conflicting hypotheses of octameric (Tzokov et al, 2006) and 13-meric (Eifler et al, 2006) pores, it is now widely accepted that a 12-meric pore is produced in *E. coli* ClyA K12. This is a result of the publication of the crystal structure of the detergent-solubilized soluble pore form of ClyA (Mueller et al, 2009), the first crystal structure of a membrane-bound α-PFT and second of any membrane-bound PFT.

The 12-meric pore structure has been determined by X-ray crystallography to 3.3 Å resolution and it shows that the ClyA monomers undergo large conformational changes before oligomerization can occur (Mueller et al, 2009). The assembled ClyA pore forms a hollow cylinder in the membrane and measures up to 130 Å in height, an inner diameter of 70 Å and outer diameter of 105 Å (Figure 1.7). Mueller et al, (2009) reported that huge conformational changes occur in the conversion of the ClyA monomer to the protomer, with more than half the residues being affected and the head domain being completely rearranged**.** In fact the content of α helices in the ClyA pore form increased from 79% to 83% compared with the soluble form.

The protomers are arranged in a parallel head to head orientation and in contrast to the four main helices (αA, αB, αC, and αF) that were observed in the water-soluble form of ClyA, the pore form of ClyA consists of a three α-helix bundle (αB, αC, and αF) in each protomer there is a large translocation of helix αA, whose N-terminal end forms a constriction at top of the pore.

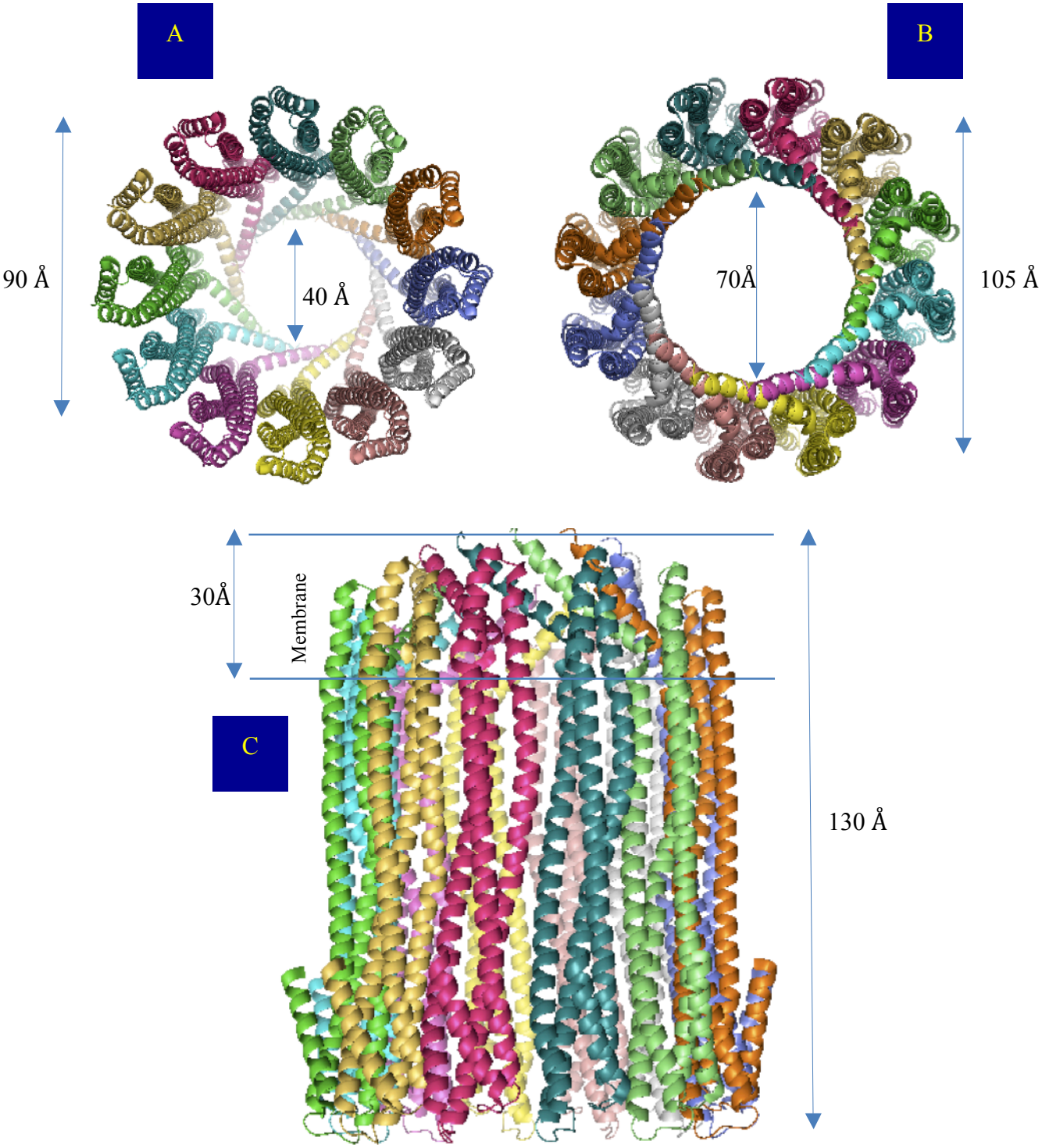


Figure 1.7: Ribbon image of the structure of the dodecameric transmembrane ClyA-K12 pore form (Mueller et al, 2009).
(A) From the top, (B) From bottom, and (C) from the side. Each protomer is a different color. The protomers are oriented head to head and parallel to the symmetry axis of the pore. Produced using PyMol (DeLano and Lam, 2005).

At the bottom of the pore, αG at the C-terminus flanks the three helices in the outside of the pore.

The protomers are tied together by the formation of a network of a 25 hydrogen bounds and 13 salt bridges between each pair of protomers in the pore.

To summarize the mechanism of the rearrangements of the body of the ClyA molecule to allow the conformational change to occur caused by membrane insertion (Mueller et al, 2009) (Figure 1.8):

1. The ß-tongue associates with the lipid membrane.

2. The ß-tongue spontaneously inserts into the membrane due to its strong hydrophobic character. This disrupts the hydrophobic core of the head region and causes the 2 flanking helices (αD and αE) to straighten out and refold to become transmembrane α helix region as previously suggested by Hunt et al, (2008). As a result, αD with ß1 becomes a helical extension to α C, and α E with ß2 becomes an extension to αF.

3. Helix αA2 becomes an extension to αB, converting the body of the protomers from a 4 to a 3 helices bundle. However the amphipathic αA1 relocates by approximately 180°, resulting in a repositioning of its N terminus by more than 140 Å and then attaching to the membrane. Thus, distinct kink in the soluble form between αA1 and αA2 at residue PRO 36 becomes now part of the turn between αA1 and the extended αB of the protomer.

4. αF moves into the position previously occupied by αA1. Then the oligomerization of protomers can begin.

5. The αA1 helix is now lying at the membrane surface, and closing of the pore causes αA to buckle and form an iris-like interlocking region, which keeps the pore open. The helices then wedge up to form a hole in the membrane.

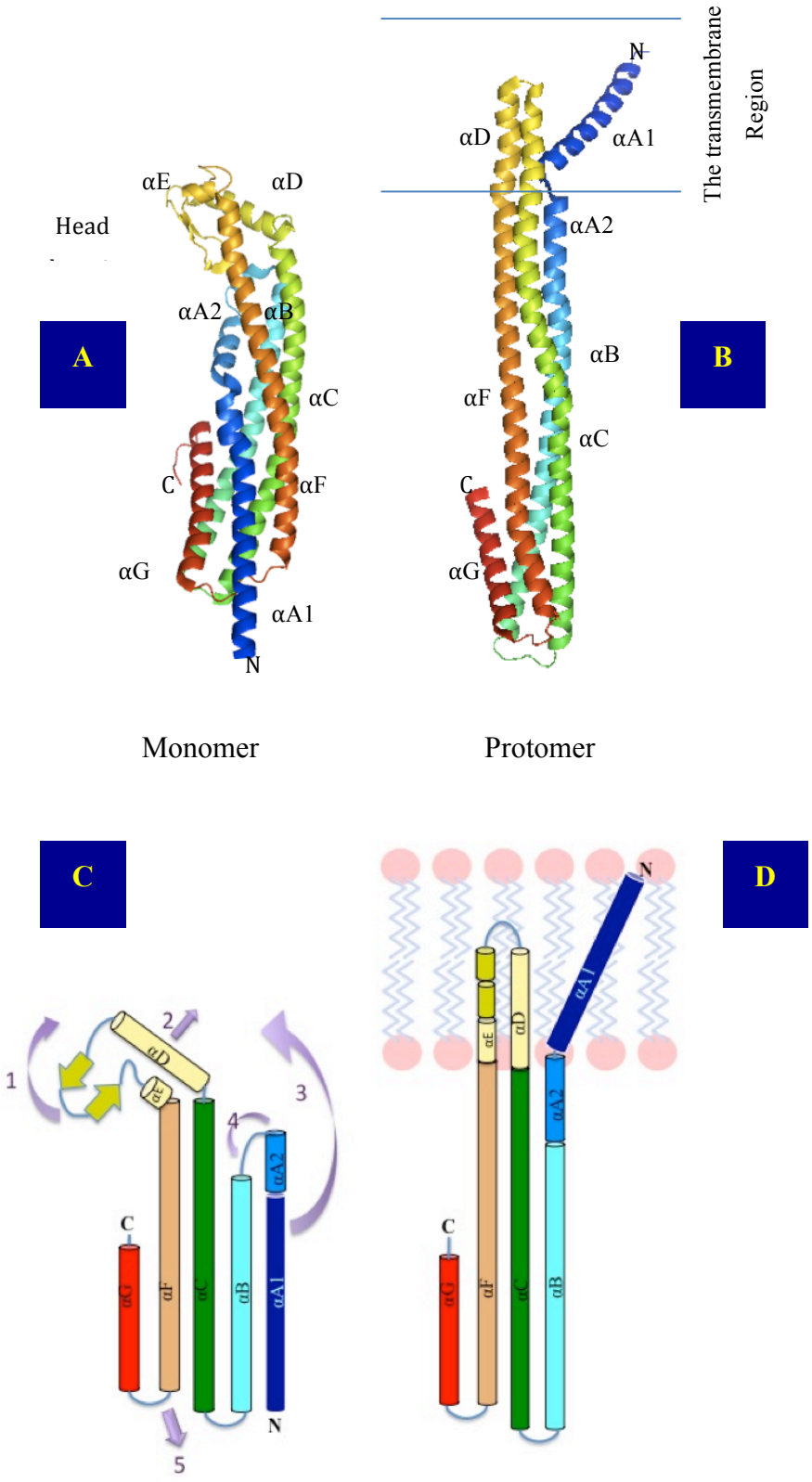


Figure 1.8: Conformational changes between the monomer and protomer forms of ClyA-K12. (A, C) the soluble form, and (B, D) the protomer in the pore form. In (C) the numbered arrows summariz the order of the rearrangements of the body of ClyA molecule proposed to allow the conformational change that occurs on membrane insertion (Mueller et al, 2009). Ribbon images in (A, B) Produced using PyMol (DeLano and Lam, 2005).

The pore assembly structure fits fairly well in the electron microscopy envelope (Figure 1.9). Extra density at the lower edge of the microscopy envelope represents lipids and detergent.

1.6 Comparison of Avian pathogenic *E. coli* strain (APEC) JM4660 ClyA with *E. coli* K-12 ClyA

ClyA was first discovered in *E. coli* K-12, but alignment of the encoded APEC ClyA with *E. coli* K-12 ClyA revealed 75% identity in their amino acid sequences and this protein will be used in this project.

APEC ClyA isolate (JM4660) encodes a gene, which has 71% identity at the nucleotide level to *clyA* (Reingold et al, 1999). The isoelectric points (IP) proved to be the same for both K-12 ClyA and APEC ClyA proteins (5.14), together with similar molecular weights of 34.6 kDa for APEC ClyA and 32.2 kDa for the *E. coli* K-12 ClyA.

There was an error in the published APEC ClyA sequence (Reingold et al, 1999). Wyborn et al, (2004) reported that at amino acid 188, JM4660 ClyA sequence encodes aliphatic neutral amino acid Glycine (GLY or G; GGG codon) instead of the polar hydrophilic charged amino acid Arginine (ARG or R; AGG codon) that was shown by Reingold et al, (1999). Residue 188 is in the hydrophobic β-tongue of ClyA and would have had serious implications for the interaction of ClyA with cell membranes (Figure 1.10) if it had been an ARG. APEC ClyA contains 76 other correct amino acid substitutions as reported by Reingold et al, (1999) when compared to the K-12 ClyA (Wyborn et al, 2004).

Both K-12 ClyA and APEC ClyA have similar properties, being active against red blood cells. However, APEC JM4660 ClyA may be advantageous to study due to its higher thermostability (consistent with the fact that birds have a higher core temperature than mammals). In addition APEC ClyA is more homogenous as the estimation of the native molecular mass; K-12 ClyA presents as mixture of monomers, dimers, and superior

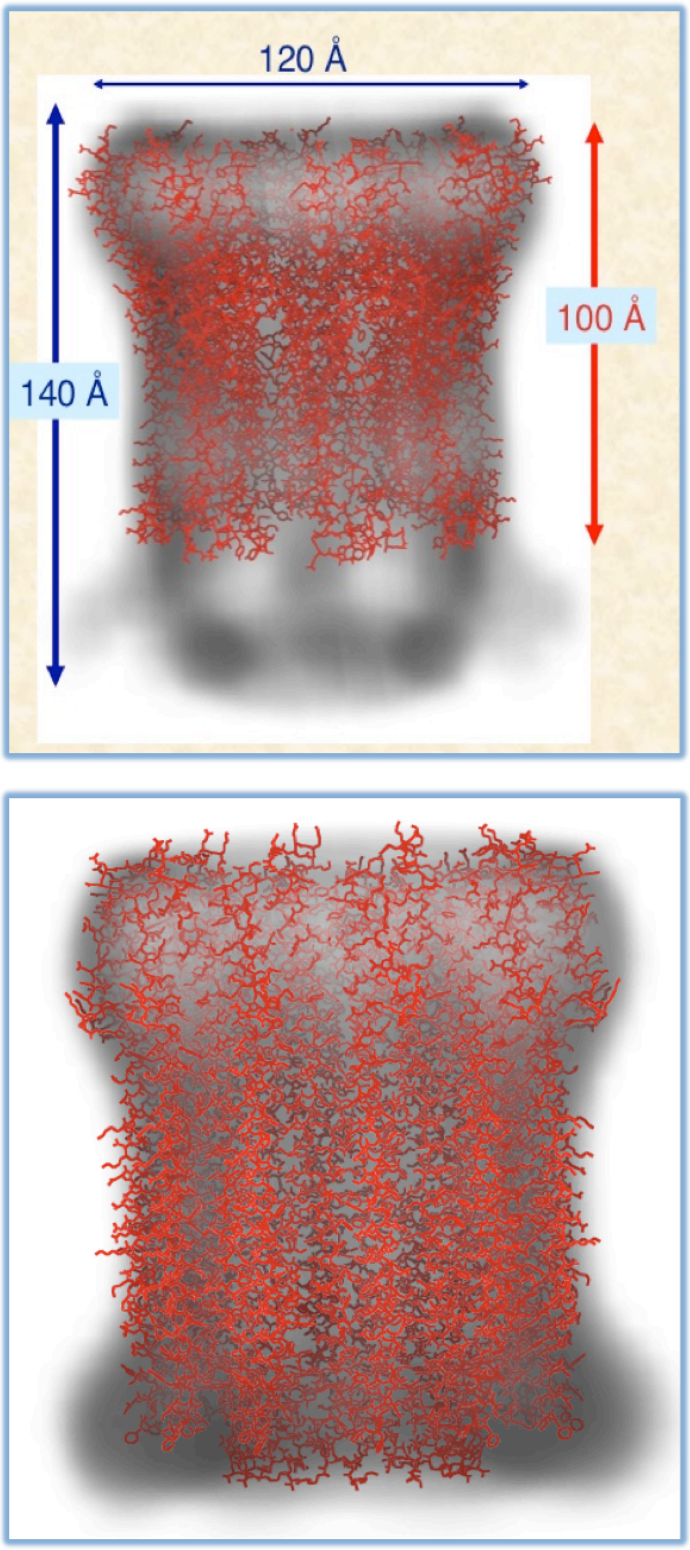


Figure 1.9: Single particle electron microscopy derived envelope of the ClyA pore assemblies.
(A) One of a big problem with the ClyA monomer structure is that the pore assemblies are much longer than the soluble ClyA structure, which means that the conformational changes must take place.
(B) The dodecameric transmembrane ClyA-K12 (Mueller et al, 2009) assembly structure fits fairly well in the electron microscopy envelope after conformational changes take place. The diagram was drawn by Prof. Pete Artymiuk (department of MBB/University of Sheffield/United Kingdom).

```
APEC1           MTN--ADQTVETVKTAIDTADKALDLYNKVLDQVIPWNTFNDTVKELSRFKEEYSQSAST 58
APEC2           MTN--ADQTVETVKTAIDTADKALDLYNKVLDQVIPWNTFNDTVKELSRFKEEYSQSAST 58
K12             MTEIVADKTVEVVKNAIETADGALDLYNKYLDQVIPWQTFDETIKELSRFKQEYSQAASV 60


APEC1           LVGEIKSLLMNSQDRYFEATQVVYEWCGVTTQLLTAYLSLFNEYDEKKASAQKTILIKVL 118
APEC2           LVGEIKSLLMNSQDRYFEATQVVYEWCGVTTQLLTAYLSLFNEYDEKKASAQKTILIKVL 118
K12             LVGDIKTLLMDSQDKYFEATQTVYEWCGVATQLLAAYILLFDEYNEKKASAQKDILIKVL 120


APEC1           DDGIIKLEKAQQSLHASSQSFNSASGKLIALDSQLANDFDEKSDYFQGQVDKIRKEAYAG 178
APEC2           DDGIIKLEKAQQSLHASSQSFNSASGKLIALDSQLANDFDEKSDYFQGQVDKIRKEAYAG 178
K12             DDGITKLNEAQKSLLVSSQSFNNASGKLLALDSQLTNDFSEKSSYFQSQVDKIRKEAYAG 180


APEC1           AAAGVVGRPFGLIISYSIAAGVVEGKLIPALKEKLKSVKDFFESLAATVKSANTDIDKAK 238
APEC2           AAAGVVGGPFGLIISYSIAAGVVEGKLIPALKEKLKSVKDFFESLAATVKSANTDIDKAK 238
K12             AAAGVVAGPFGLIISYSIAAGVVEGKLIPELKNKLKSVQNFFTTLSNTVKQANKDIDAAK 240


APEC1           SKLKDEISVIGDLKTETETTRFFVDYDDLMLKQLQDSATKLILSCNEYQKRHGKKNEQPL 298
APEC2           SKLKDEISVIGDLKTETETTRFFVDYDDLMLKQLQDSATKLILSCNEYQKRHGKKNEQPL 298
K12             LKLTTEIAAIGEIKTETETTRFYVDYDDLMLSLLKEAAKKMINTCNEYQKRHGKKTLFEV 300


APEC1           WY- 300
APEC2           WY- 300
K12             PEV 303
```

Figure 1.10: Amino acid sequence alignment of ClyA-K12 and APEC ClyA.
APEC**1** ClyA was determined by Reingold et al (1999); APEC**2** ClyA was corrected by Wyborn et al (2004). The ClyA K-12 sequence is used as the reference sequence**.** The correct APEC ClyA (APEC**2)** sequence encodes the aliphatic neutral amino acid Glycine (GLY or G) (Wyborn et al 2004) at residues 188 instead of the polar hydrophilic charged amino acid Arginine (ARG or R) as incorrectly reported by Reingold et al (1999). The original APEC1 ClyA sequence is coloured in red. The amino acid sequence alignment was generated using ClustalW2 (Goujon, et al. 2010& Larkin, et al. 2003) (available on: http://www.ebi.ac.uk/Tools/msa/).

order aggregates (Atkins et al, 2000) while the most of JM4660 ClyA was dimeric in solution (Wyborn et al, 2004).

1.7 Pore formation By *E. coli* K-12 ClyA and APEC ClyA by elecron microscopy

Electron microscopy of negatively-stained ClyA *E. coli* K-12 in lipid vesicles showed many stain-filled ring-shaped pore assemblies (Wallace et al, 2000). They showed that most of the pores are circular shaped structures when the pore is viewed from above (Figure 1.11). They estimated that 62% of the individual pores had an internal diameter of 42-52 Å and an external diameter of 70-90 Å while 36% of them were bigger than this (having an inner pore diameter of 55-60 Å and an external diameter of 90-105 Å). Also, the pores can be seen as spikes when they are viewed from the side, approximately 105 -110 Å long. These observations by electron microscopy show that the ClyA *E. coli* K-12 monomers oligomerize to build channels that project from the cell membrane. Pores formed by APEC ClyA were more homogeneous than ClyA K-12 ones with an inner diameter 36-41 Å (Wyborn et al, 2004). As observed with ClyA *E. coli* K-12, the side views of the APEC ClyA pore assembly can be seen as spikes 90-100 Å long (Wyborn et al, 2004). This greater consistency in size means they may be more suitable for structural studies.

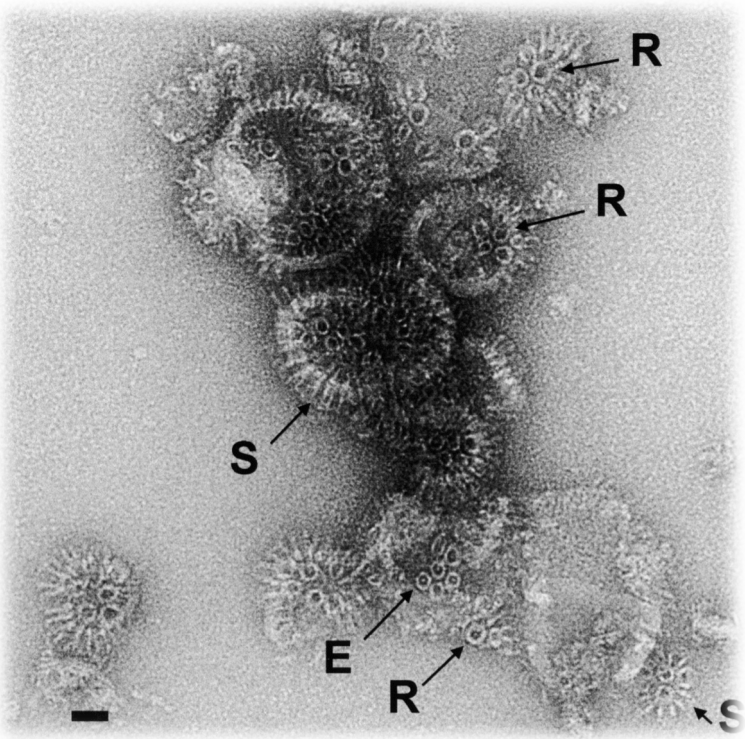


Figure 1.11: Electron micrograph of negative stained ClyA-K12 in lipid vesicles.
Stain-filled rings (R) are apparent in views normal to the membrane. Some pores show a central stain-excluding density (E). Side views of the pore appear as spikes (S). Scale bar (lower left) represents 200 Å.
Republished from Wallace et al (2000), with permission from Elsevier.

1.8 The first aim of this project

When my project started in November 2008 using mainly X-ray crystallographic analysis, my aim was to grow crystals of the pore form of APEC ClyA. I was eventually able to grow crystals of APEC ClyA in the presence of detergent that produced low-resolution diffraction to 7 Å as described in chapter 4. While these crystallization attempts were on-going, the structure of the pore form of *E. coli* K-12 ClyA, was determined to 3.3 Å resolution (Mueller et al, 2009). This structure revealed the 12-meric pore structure (described in 1.5.3). Because of the similarity between K-12 ClyA and APEC ClyA, it is very likely that a 12-meric pore is will also be produced by APEC ClyA, with the monomers undergoing the same large conformational changes before oligomerization can occur.

However, other oligomeric states and conformations may also possible as a result of the variability in pore size which has been observed by EM (Eifler et al, 2006; Tzokov et al, 2006) and there is interest in solving alternative structure to clarify the following questions:

- The pores appear to exist in two forms: an open pore in mammalian that is hemolytic, and a closed pore in *E. coli* that is not haemolytic (Wai et al, 2003). So, what is the appearance of the "closed" form in relationship to the open pore (Mueller et al, 2009)? i.e. what are the conformational differences between closed and open pores?
- Does the G-helix move and does its movement contribute to closure of the pore?
- Is there any mobility in β tongue region in the soluble form?

Therefore, we attempted to solve the crystal structure of APEC ClyA pores with detergents or lipids by X-ray crystallography and EM at higher resolution. This would hopefully have helped to answer the above questions and could have provided further important details to the understanding of pore formation and structure.

1.9 ClyA like toxins from *Bacillus cereus*

Sequence searches show that ClyA has sequence homologues in *E. coli, Shigella flexneri* and *Salmonella typhi* and *Salmonella paratyphi.* The three dimensional crystal structure of ClyA had a novel structure that did not occur elsewhere in the data bank of proteins, based on searches of the structural databases (Wallace et al, 2000). Recently, however the three dimensional crystal structure of hemolysin BL component B (HblB) from *Bacillus cereus* was solved (Madegowda et al, 2008). This showed that HblB has very strong structure similarity with ClyA even through there is a little sequence homology between ClyA and *B. cereus* toxins (Madegowda et al, 2008). In addition, the strong sequence relationship of Hbl with the non hemolytic enterotoxin (Nhe) from *B. cereus* implied a possible similarity in structure and function between Nhe and ClyA (Fagerlund et al, 2008). This showed that *E. coli* ClyA, Nhe, and Hbl of *B. cereus* could represent a novel super-family of pore forming toxins.

The next sections provide a context for consideration of the protein that is the main focus of this thesis: the Non-hemolytic enterotoxin Nhe protein from *B. cereus*.

1.10 *Bacillus cereus*

B. cereus belongs to the family Bacillaceae. It is a gram positive, rod shaped, and spore-forming bacterium, ubiquitous, aerobic or facultative anaerobic, motile, widespread in nature including soils (mainly rice paddy soil), sediments, meat, egg, dust, natural water, hair of animal, dairy products and vegetation (Granum & Lund, 1997; Kramer & Gilbert., 1989). It can also be found in drugs, including both topical and oral pharmacetutical products (Garcia Arribas et al, 1988). The full genomes have been sequenced for *B. cereus* strains ATCC 14579 (Ivanova et al, 2003), ATCC 10987 (Rasko et al, 2004), G9241(Hoffmaster et al, 2004).

Due to the ability of *B. cereus* to form spores it has strong resistance to

extreme environmental conditions such as heat and chemical treatments, as well as freezing, drying, dehydration, enzymatic attack and radiation (Kotiranta et al, 2000; Setlow & Setlow, 1995).

1.11 Illness caused by *Bacillus cereus*

B. cereus is now recognized as a common agent of food- poisoning illness in humans. It can cause two different forms of food-poisoning: the emetic and diarrheal type (Kotiranta et al, 2000).

1.11.1 The emetic gastrointestinal diseases

The emetic syndrome was identified early in the 1970s in the United Kingdom, being caused by the ingestion of the preformed toxins in food. The emetic toxin that is known as cereluide causes nausea and vomiting a few hours after ingestion of contaminated food, rarely followed by abdominal cramping or diarrhea (Kramer & Gilbert, 1989). Other studies have indicated the high toxicity potential of cereluide including immunomodulatory and neuro- and hepatotoxic modes of action (Andersson et al, 2007; Mikkola et al, 1999; Paananen et al, 2002).

Cereluide is an ionophoric, cyclic dodecadepsipeptide with molecular weight 1.2 kDa and consisting of α-amino and α-hydroxy acids (D-*O*-Leu-D-Ala-L-*O*-Val-L-Val)$_3$ (Agata et al, 1994), synthesized by a non-ribosomal peptide synthetase, encoded by the *ces* genes (Ehling-Schulz et al, 2005a; Ehling-Schulz et al, 2005b). It is a stable toxin due to its strong resistance to low pH (as low as 2), heat and to proteolysis. Therefore, this toxin cannot be inactivated in the acidic environment of the stomach or by enzyme proteolytic activity in the intestinal tract (Ehling-Schulz et al, 2004; Frenzel et al, 2011; Pexara & Govaris, 2010). Frenzel et al, (2011) shows that use of commercial polyphosphate blends could control and inhibit cereulide synthesis in food, mainly where *B. cereus* spores cannot be inactivated

during processing or preparation steps.

1.11.2 The diarrheal gastrointestinal diseases

B. cereus caused diarrhea was identified in Norway in 1947-1949 (Hauge, 1955). It is caused by enterotoxins which form during vegetative growth of *B. cereus* in the small intestine of the host organism (Granum et al, 1993). The diarrheal syndrome appears 8–16 hours after ingestion and is characterized by abdominal pain and watery diarrhea without vomiting or fever.

At the present time, at least three different pore forming enterotoxins appear to be responsible for diarrheal type and involved in food-poisoning: two protein complexes, hemolysin BL (Hbl) (Beecher & Macmillan, 1991; Beecher et al, 1995) and non-hemolytic enterotoxin (Nhe) (Granum et al, 1999; Lund & Granum, 1996), and the cytotoxin K (CytK) (Lund et al, 2000).

1.12 The *Bacillus cereus* protein enterotoxins

1.12.1 *Bacillus cereus* Cytotoxin K (CytK)

CytK is a single pore forming cytotoxic protein with a molecular mass 34 kDa and encoded by *cytK*. CytK has vascular permeability activity against epithelial cells in the small intestine, and causes fluid accumulation in cells (Lund et al, 2000). This toxin has been isolated from *B. cereus* strain 391/98 that does not contain genes for the other known *B. cereus* enterotoxins Hbl and Nhe (Fagerlund et al, 2004). This strain was responsible for an acute food poisoning in outbreak France (Lund et al, 2000). CytK forms weakly anion selective pores in planar lipid bilayers with 7 Å in diameter.

The CytK amino acid sequence shows similarity to *Staphylococcus aureus* α-hemolysin toxin and *Clostridium perfringens* β-toxin (Gouaux, 1998; Lund et al, 2000; Steinthorsdottir et al, 2000), which belong to the family of β-barrel pore-forming toxins (discussed in section 1.3.2). This suggests that CytK may also classified as a β-barrel pore-forming toxin (Lund et al, 2000).

1.12.2 *Bacillus cereus* hemolysin BL (Hbl)

Hemolysin BL (Hbl) is a complex pore forming toxin, consisting of three-components: Hbl component B which is reported as a binding protein (35 kDa) to lipid bilayer, and two lytic proteins: Hbl component L1 (36 kDa), and Hbl component L2 (45 kDa) (Beecher & Macmillan, 1991). None of these proteins are active individually and all of them are required for maximal cytotoxicity (Beecher & Macmillan, 1991; Beecher et al, 1995). However, Beecher & Wong, (1997) reported that all the three components are able to bind separately to the red blood cells (RBCs) surface in experiments where one Hbl proteins was added to RBCs , then washed, then the another two Hbl protein caused hemolysis regardless of which Hbl proteins was first added.

HblB, HblL1, and HblL2 are encoded by the genes *hbA*, *hbD*, and *hbC* respectively and have been cloned and sequenced from the *B. cereus* strain F837/76. These genes are located in the same operon and give a 5.5 kb mRNA transcript (Heinrichs et al, 1993; Lindbäck et al, 1999; Ryan et al, 1997). Lindbäck et al, (1999) indicated that the cytotoxic activity of Hbl was decreased by disruption of the *hbl* operon in *B. cereus* strain ATCC14579.

Hbl is able to lyse Vero cells, and it has hemolytic activity against some mammalian cells such as sheep, bovine and mouse erythrocytes but not against human, horse and rabbit erythrocytes (Beecher & Macmillan, 1990; Beecher & Wong, 2000). It also caused fluid accumulation in rabbit ileal loops (Beecher et al, 1995).

Madegowda et al, (2008) solved the 3D crystal structure of Hbl component B at 2 Å resolution using multiple anomalous dispersion (MAD) phasing. HblB has long α-helical bundle and elongated structure with dimensions of 90 Å x 40 Å x 30 Å. There are four major helices in the polypeptide chain, α A (GLU 19- LYS 45), αB (ASP 61- ASN 114) with distinct kinks in LYS 83-VAL 84, αC (GLY 118- ASN 174), and αF (TYR 244- ALA 297) (Figure 1.12).

Like ClyA, HblB has two sub-domains at both ends, which resemble the two novel domains in ClyA:

-The head domain, which consists of an α/ β hydrophobic hairpin, α D (ASP 177- LEU 189) with distinct kinks at LEU 189- GLY 190, αE (ARG 225- GLN 240) with distinct kinks at LEU 228-GLY 229, and two short strands β1 and β2.

-The tail domain that is made up of αG (LYS 307-LYS 335) at the C terminal with distinct kinks at LYS 310- PRO 311.

Using DALI (Holm & Rosenström, 2010), it was shown that there is a high structural similarity between HblB and K-12 ClyA (Figure 1.13) with a Z-score of 13.7, although sequence alignment showed a limited match (sequence identity = 16%) (Figure 1.14) (Madegowda et al, 2008). As well as the relationship between HblB and K-12 ClyA in structure, both of them possess the same function, which includes pore formation with cytotoxic and hemolytic activity (Madegowda et al, 2008) and so may represent a new superfamily of pore forming toxins.

The main difference between HblB and K-12 ClyA is the orientation of the head domain. The head domain in HblB is turned downward and makes interactions with the tail domain. However the head domain in K-12 ClyA turned is turned upward with a only small interaction with the tail domain (Figure 1.13) (Madegowda et al, 2008). This could represent two different conformations in the two structures (Fagerlund et al, 2008).

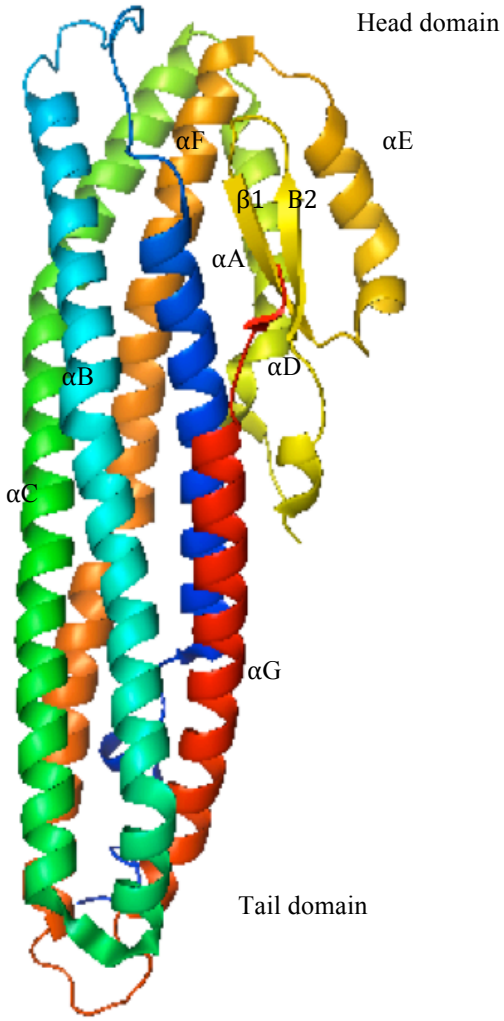


Figure. 1.12: The 3D crystal structure of monomeric HblB (Madegowda et al, 2008).
The main body consists of αA, αB, αC, and αF. The head domain has two shorter α-helices, labelled αD and αE, with two strands β 1 and β 2. αG is located in tail domain.
Figure was produced using PyMol (DeLano and Lam, 2005).

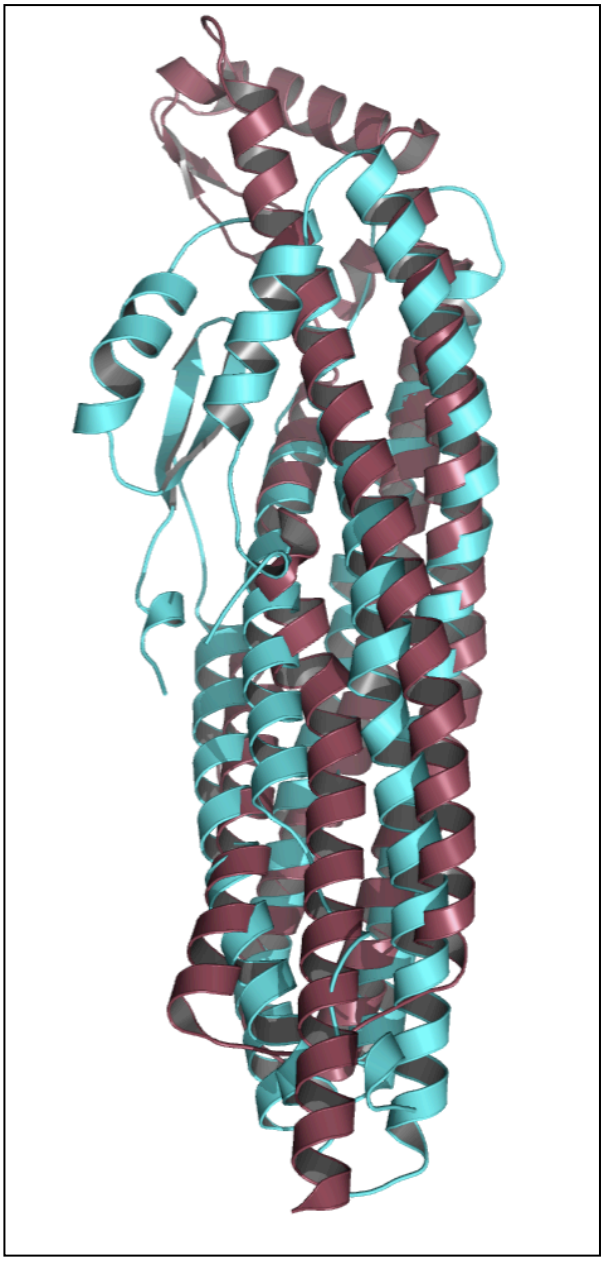


Figure 1.13: Superimposed structures of HblB (cyan) and K-12 ClyA (pink).
It is clear that the main difference is in the β-tongue region. The β-tongue is turned upward and connects with the tail domain in HblB, but it is turned downward in ClyA. The image was produced using the Dali server (Holm and Rosenström 2010).

Figure 1.14: Amino acid alignment of *B. cereus* HblB and *E. Coli* ClyA-K12.

Hydrophobic residues are colored red; acidic residues are colored blue; basic residues are colored magenta; hydroxyl, sulfhydryl, amine, and G residues are colored green. (* Asterisk) shows a fully conserved residue.
(: Colon) shows conservation between groups of strongly similar properties.
(. Period) shows conservation between groups of weakly similar properties.
The hydrophobic transmembrane sequence is highlighted by yellow background in ClyA-K12 and grey background in HblB.
The amino acids sequence alignment was generated using ClustalW2 (Goujon, et al. 2010& Larkin, et al. 2003) (available on: http://www.ebi.ac.uk/Tools/msa/).

1.12.3. *Bacillus cereus* non-hemolytic enterotoxins (Nhe)

Besides Hbl, another enterotoxin complex known as the non-hemolytic enterotoxin (Nhe), represents a major pathogenicity factor involved in the etiology of *B. cereus* diarrheal syndrome.

Nhe is a complex pore-forming toxin consisting of three proteins (Table 1.1): NheA (41 kDa), NheB (39.8 kDa), NheC (36.5 kDa) (Granum et al, 1999) that was isolated firstly from *B. cereus* strain 0075/95, which lacks the *hbl* genes and *cytK* gene and was involved in a large food–poisoning outbreak in Norway in 1995 (Lund & Granum, 1996). In initial studies Lund & Granum, (1999) suggested that a collagenase (105 kDa) was part of the Nhe complex but this was later found to be wrong (Granum et al, 1999). Sequencing of the *nhe* operon (Figure 1.15) indicated that it is composed of three open reading frames: *nheA*, *nheB*, and *nheC* and it was found that all three proteins are required for maximal toxicity (Granum et al, 1999).

NheC is produced by *B. cereus* in much lower amounts than NheA and NheB, and thus NheC has never been isolated and purified from *B. cereus* supernatant culture. Use of a specific antibody confirmed the presence of NheC in lower amounts than NheA and NheB (Dietrich et al, 2005). In contrast to the Nhe proteins, Hbl proteins have been isolated from *B. cereus* supernatant culture and all of them appear to present in approximately the same amount (Beecher & Macmillan, 1991).

Because an original study (Lindbäck et al, 2004) on Nhe wrongly indicated this toxin did not have hemolytic activity it was misnamed “non-hemolytic enterotoxin”. However, further study confirmed the ability of Nhe to lyse mammalian erythrocytes from various organisms, including human, horse, cat, bovine, dog and pig (Fagerlund et al, 2008) although with less hemolytic activity than Hbl. In addition, Nhe has cytotoxic activity against epithelial cells, indicated by loss of cellular Adenosine-5'-triphosphate (ATP) and LDH release (Fagerlund et al, 2008). To obtain the maximum cytotoxic activity the ratio of NheA: NheB: NheC should be 10:10:1 as determined in a Vero cells assay and reported by Lindbäck et al, (2004).

Table 1.1: Shows the characteristics of Nhe proteins. According to http://web.expasy.org/protparam/ (Gasteiger et al, 2005).

Protein	Number of amino acids	Molecular weight (MW)	Theoretical pI (isoelectric point)
NheA	386	43742.5	5.22
NheB	402	43119.0	6.85
NheC	359	39877.3	5.91

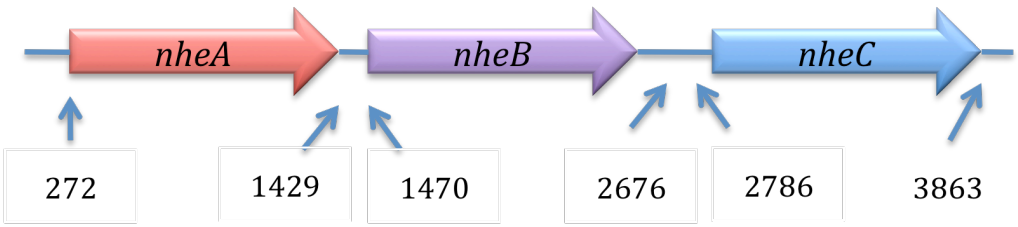


Figure 1.15: The *nhe* operon shows the site of *nheA*, *nheB* and *nheC*. The numbers illustrate the start and stop point of the three genes. (Granum et al, 1999).

Also, it was noticed that increasing the concentration of NheC inhibited the cytotoxic activity. The function of NheC is not clear yet but it may act as catalyst by bringing NheA to NheB after binding of NheB to target membranes or it may enhance conformational changes (Lindbäck et al, 2004). However, it has also been reported that NheA and NheB in the absence of NheC are necessary and sufficient for formation of large-conductance channels in rodent pituitary cells (GH_4), which are the largest channels observed in cell membranes (Haug et al, 2010). However, in primate epithelial cells (Vero) addition of NheC is required for forming channels (Haug et al, 2010).

The detailed mechanism behind pore-formation of the toxin is unknown, but is presumed to follow a pattern that involves membrane binding, oligomerization and finally insertion of the transmembrane regions to form the pore. It has previously been shown that NheB is the only component which can bind alone to the cell surface of the Vero cells (Lindbäck et al, 2004) but recently it has been shown that NheC also binds to the membrane of Vero cells but NheA lacks this ability (Lindbäck et al, 2010).

NheB and NheC share 44% sequence identity but NheA has only 22% identity with NheB and NheC. To produce cytotoxic activity, it has been demonstrated that Nhe requires a specific binding order of the individual proteins whereby the presence of NheB and NheC is necessary in the priming step, and NheA is required in the final incubation stage of pore formation (Lindbäck et al, 2010)

Nhe proteins have significant sequence homology to Hbl components (Figure 1.16) and (Table 1.2) but do not have significant sequence similarity to other known proteins. This means that all Nhe proteins and Hbl proteins belong to same protein family. Beside sequence similarities NheA and HblL2 have no predicted transmembrane region, while NheB and HblL1 have two transmembrane helices. Finally, NheC and HblB each have one predicted transmembrane helix, in the same position of the two proteins (Granum et al, 1999) (Figure 1.17).

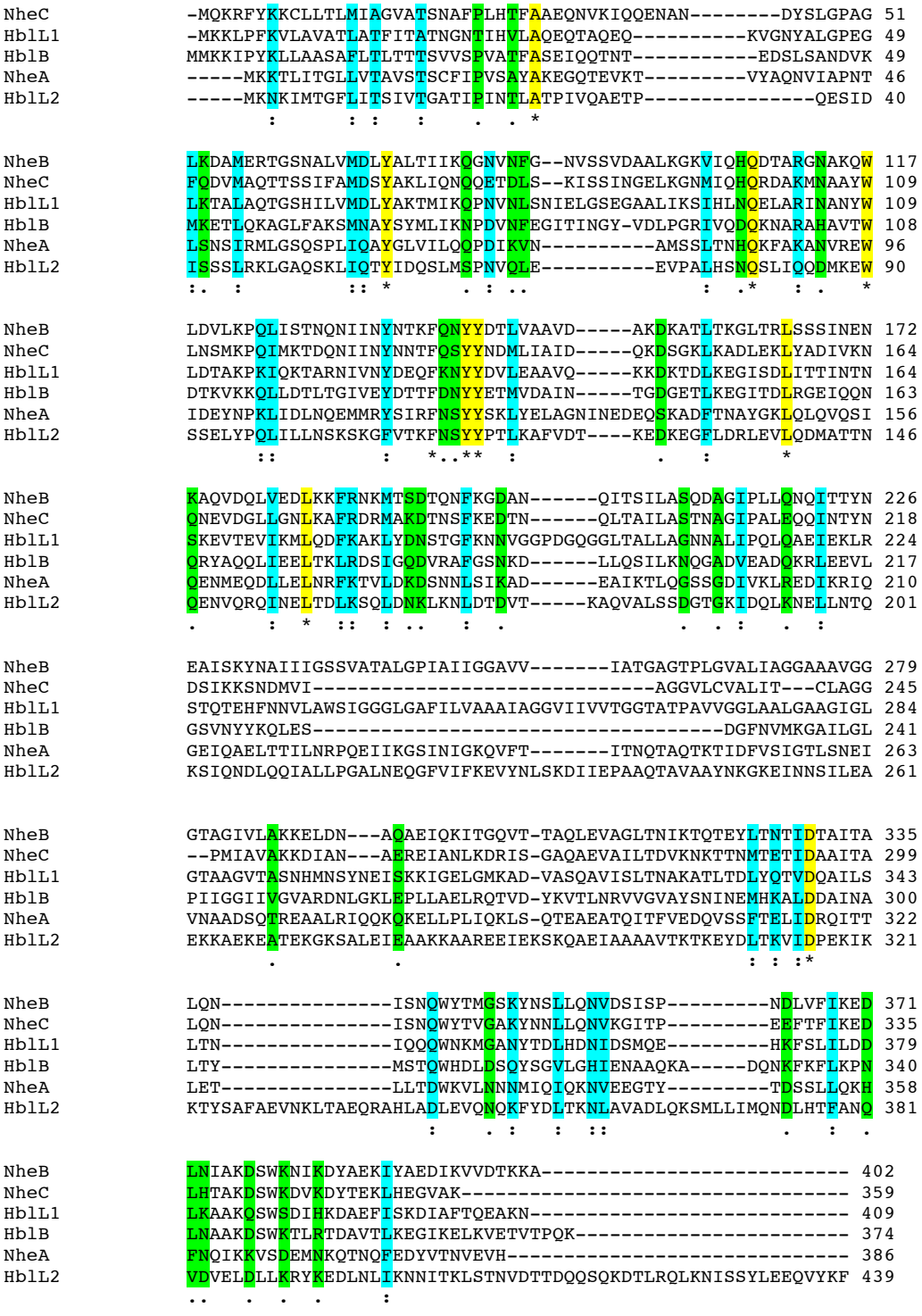


Figure 1.16: Amino acid alignment of *B. cereus* Nhe and Hbl proteins. Yellow background indicates a fully conserved residue. Blue background indicates conservation between groups of strongly similar properties. Green background indicates conservation between groups of weakly similar properties. The amino acid sequence alignments were generated using ClustalW2 (Goujon, et al. 2010& Larkin, et al. 2003) (available on: http://www.ebi.ac.uk/Tools/msa/).

Table 1.2 The percentage sequence identity between the Nhe proteins and the Hbl portions (Fagerlund et al, 2008).

	HblL2	**HblL1**	**Hbl B**	**NheA**	**NheB**	**NheC**
HblL2	-	18	23	23	21	19
HblL1	18	-	25	18	**40**	32
HblB	25	25	-	20	27	25
NheA	23	23	20	-	22	22
NheB	21	**40**	27	22	-	**44**
NheC	19	32	25	22	**44**	-

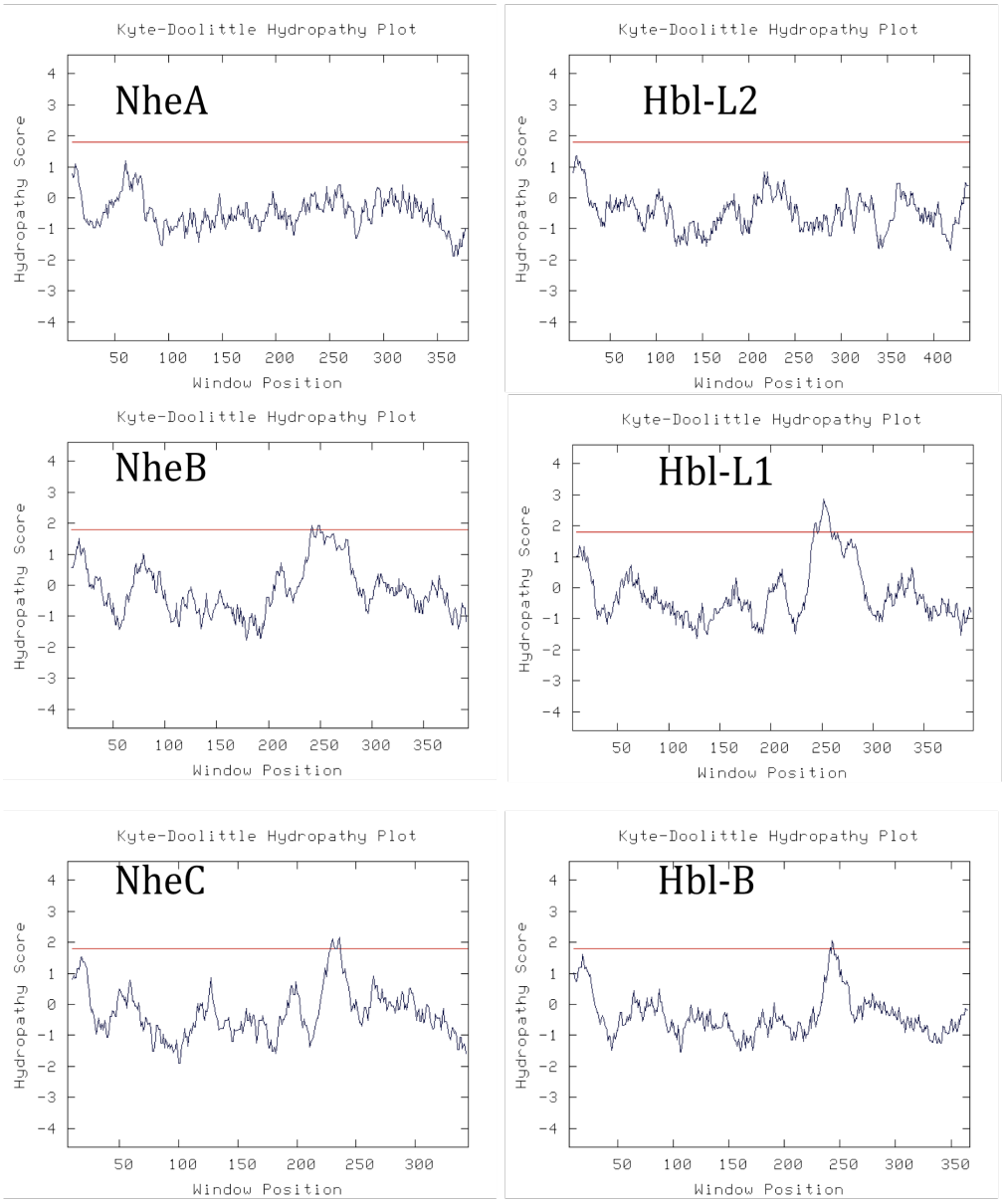


Figure 1.17: Hydrophobicity plots for Nhe and Hbl portions.
The Kyte-Doolittle hydropathy calculation (Kyte, and Doolittle 1982) is used to score the hydrophobic character of the protein. Peaks (with scores >1.8; red line) show possible trans-membrane regions.
NheA and HblL2 have no predicted transmembrane region, while NheB and HblL1 have two transmembrane helices. Finally, NheC and HblB each have one predicted transmembrane helix.
Calculated at (http://gcat.davidson.edu/rakarnik/KD.html).

Sequence similarity between HblB and NheB, NheC (Table 1.2) has allowed generation of 3D homology computer models using HblB as template (Fagerlund et al, 2008). The predicated 3D structures are mainly α-helical, including four main chains with β- hairpin that flanked by two α- helices.

As mention before, HblB has strong structural similarities with ClyA (see Figure 1.13) although sequence alignment showed only limited match between ClyA and the Nhe and Hbl components (Table 1.3). Consequently, the strong sequence relationship of Hbl with Nhe implied a possible similarity in structure and function between Nhe and ClyA. This showed that ClyA, Nhe, and Hbl represent a novel super-family of pore forming toxin.

1.13 Regulation of Nhe expression

Nhe expression is positively regulated by PlcR/PapR quorum-sensing system. PlcR is a protein with 34 kDa transcriptional regulator. PlcR encoded by the *B. cereus plcR* gene and control the expression of *B. cereus* cytotoxin, including Nhe, CytK, Hbl, degradative enzymes (phospholipases, proteases), and cell-surface proteins (Agaisse et al, 1999; Slamti & Lereclus, 2002). Activation of PlcR is dependent on the presence of PapR, a small signaling peptide around a 48 amino acids and encoded by *papR* gene, which is located 70 bp downstream from *plcR*. PlcR acts as a quorum-sensing effector and exported by the bacterial cell, processed, possibly as a pentamer, and then reimported into the cell. Once inside the cell, PapR interacts with PlcR to facilitate binding to its DNA targets and activate their transcription (Slamti & Lereclus, 2002). It was shown deletion of the *plcR* gene in *B. cereus* decreased hemolytic activity and virulence of the bacterium compared to the wild-type *B. cereus* strain (Gohar et al, 2002).

Table 1.3 The percentage of sequences identity between *B. cereus* Nhe, Hbl proteins and *E. coli* ClyA-K12 (Fagerlund et al, 2008)

	ClyA	HblL2	HblL1	Hbl B	NheA	NheB	NheC
ClyA	-	**18**	**20**	**20**	**18**	**19**	**20**
HblL2	18	-	18	23	23	21	19
HblL1	20	18	-	25	18	40	32
Hbl B	20	25	25	-	20	27	25
NheA	18	23	23	20	-	22	22
NheB	19	21	40	27	22	-	44
NheC	20	19	32	25	22	44	-

1.14 The main Thesis aims and objectives

Experiments from different laboratories around the world, described above, shows that the pore forming toxins are an interesting and important proteins but that there are many outstanding issues concerning their structures in soluble and membrane forms, their mode of action and the roles of lesions in biological membranes.

To date, few studies of Nhe proteins have been available and there is no 3D crystal structure for any of the Nhe proteins in either the soluble form or in the pore form. Consequently, more studies of pore formation and the hetero-oligomeric Nhe complex are critical to understand the mechanism of this super-family of pore forming toxin. NheA and NheB were are necessary and sufficient for pore formation and cell lysis (Haug et al, 2010). According to that, we are interested to elucidate the 3D crystal structure of Nhe component A and B (NheA, NheB) using X-ray crystallography and electron microscopy. Hopefully this would provide new insights into the structure of the pore forming toxins.

References

Agaisse H, Gominet M, Okstad OA, Kolsto AB, Lereclus D (1999) PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis*. *Mol Microbiol* **32:** 1043-1053

Agata N, Mori M, Ohta M, Suwan S, Ohtani I, Isobe M (1994) A novel dodecadepsipeptide, cereulide, isolated from *Bacillus cereus* causes vacuole formation in hep-2 cells. *FEMS Microbiol Lett* **121:** 31-34

Andersson MA, Hakulinen P, Honkalampi-Hamalainen U, Hoornstra D, Lhuguenot JC, Maki-Paakkanen J, Savolainen M, Severin I, Stammati AL, Turco L, Weber A, von Wright A, Zucco F, Salkinoja-Salonen M (2007) Toxicological profile of cereulide, the *Bacillus cereus* emetic toxin, in functional assays with human, animal and bacterial cells. *Toxicon* **49:** 351-367

Andreeva ZI, Nesterenko VF, Yurkov IS, Budarina ZI, Sineva EV, Solonin AS (2006) Purification and cytotoxic properties of *Bacillus cereus* hemolysin II. *Protein Expr Purif* **47:** 186-193

Aroian R, van der Goot FG (2007) Pore-forming toxins and cellular non-immune defenses (CNIDs). *Curr Opin Microbiol* **10:** 57-61

Atkins A, Wyborn NR, Wallace AJ, Stillman TJ, Black LK, Fielding AB, Hisakado M, Artymiuk PJ, Green J (2000) Structure-function relationships of a novel bacterial toxin, hemolysin E - The role of alpha(G). *J Biol Chem* **275:** 41150-41155

Beecher DJ, Macmillan JD (1990) A novel bicomponent Hemolysin from *Bacillus cereus*. *Infect Immun* **58:** 2220-2227

Beecher DJ, Macmillan JD (1991) Characterization of the components of Hemolysin bl from *Bacillus cereus*. *Infect Immun* **59:** 1778-1784

Beecher DJ, Schoeni JL, Wong ACL (1995) Enterotoxic activity of Hemolysin bl from *Bacillus cereus*. *Infect Immun* **63:** 4423-4428

Beecher DJ, Wong ACL (1997) Tripartite Hemolysin bl from *Bacillus cereus* - Hemolytic analysis of component interactions and a model for its characteristic paradoxical zone phenomenon. *J Biol Chem* **272:** 233-239

Beecher DJ, Wong ACL (2000) Cooperative, synergistic and antagonistic haemolytic interactions between Haemolysin bl, phosphatidylcholine phospholipase C and sphingomyelinase from *Bacillus cereus*. *Microbiology-Sgm* **146:** 3033-3039

Bhakdi S, Tranumjensen J (1991) Complement lysis - a hole is a hole. *Immunol Today* **12:** 318-320

Bischofberger M, Gonzalez MR, van der Goot FG (2009) Membrane injury by pore-forming proteins. *Curr Opin Cell Biol* **21:** 589-595

Boehm DF, Welch RA, Snyder IS (1990) Domains of *Escherichia coli* hemolysin (HlyA) involved in binding of calcium and erythrocyte membranes. *Infect Immun* **58:** 1959-1964

Coote JG (1992) Structural and functional relationships among the RTX toxin determinants of Gram negative bacteria. *Fems Microbiol Rev* **88:** 137-162

Cortajarena AL, Goni FM, Ostolaza H (2002) His-859 is an essential residue for the activity and pH dependence of *Escherichia coli* RTX toxin alpha-hemolysin. *Journal of Biological Chemistry* **277:** 23223-23229

Cortajarena AL, Goñi FM, Ostolaza H (2001) Glycophorin as a receptor for *Escherichia coli* α-hemolysin in erythrocytes. *J Biol Chem* **276:** 12513-12519

Cox AJ, Hunt ML, Ruffolo CG, Adler B (2000) Cloning and characterisation of the *Pasteurella multocida ahpA* gene responsible for a haemolytic phenotype in *Escherichia coli*. *Vet Microbiol* **72:** 135-152

Dang TX, Hotze EM, Rouiller I, Tweten RK, Wilson-Kubalek EM (2005) Prepore to pore transition of a cholesterol-dependent cytolysin visualized by electron microscopy. *J Struct Biol* **150:** 100-108

del Castillo FJ, Leal SC, Moreno F, del Castillo I (1997) The *Escherichia coli* K-12 sheA gene encodes a 34-kDa secreted haemolysin. *Mol Microbiol* **25:** 107-115

del Castillo FJ, Moreno F, del Castillo I (2001) Secretion of the *Escherichia coli* K-12 SheA hemolysin is independent of its cytolytic activity. *FEMS Microbiol Lett* **204:** 281-285

DeLano WL, Lam JW (2005) PyMOL: A communications tool for computational models. *Abstr Pap Am Chem Soc* **230:** 254-COMP

Dietrich R, Moravek M, Burk C, Granum PE, Martlbauer E (2005) Production and characterization of antibodies against each of the three subunits of the *Bacillus cereus* nonhemolytic enterotoxin complex. *Appl Environ Microbiol* **71:** 8214-8220

Ehling-Schulz M, Fricker M, Scherer S (2004) *Bacillus cereus*, the causative agent of an emetic type of food-borne illness. *Mol Nutr Food Res* **48:** 479-487

Ehling-Schulz M, Svensson B, Guinebretiere MH, Lindback T, Andersson M, Schulz A, Fricker M, Christiansson A, Granum PE, Martlbauer E,

Nguyen-The C, Salkinoja-Salonen M, Scherer S (2005a) Emetic toxin formation of *Bacillus cereus* is restricted to a single evolutionary lineage of closely related strains. *Microbiology* **151:** 183-197

Ehling-Schulz M, Vukov N, Schulz A, Shaheen R, Andersson M, Martlbauer E, Scherer S (2005b) Identification and partial characterization of the nonribosomal peptide synthetase gene responsible for cereulide production in emetic *Bacillus cereus*. *Appl Environ Microbiol* **71:** 105-113

Eifler N, Vetsch M, Gregorini M, Ringler P, Chami M, Philippsen A, Fritz A, Muller SA, Glockshuber R, Engel A, Grauschopf U (2006) Cytotoxin ClyA from *Escherichia coli* assembles to a 13-meric pore independent of its redox-state. *EMBO J* **25:** 2652-2661

Fagerlund A, Lindback T, Storset AK, Granum PE, Hardy SP (2008) *Bacillus cereus* Nhe is a pore-forming toxin with structural and functional properties similar to the ClyA (HIyE, SheA) family of haemolysins, able to induce osmotic lysis in epithelia. *Microbiology* **154:** 693-704

Fagerlund A, Ween A, Lund T, Hardy SP, Granum PE (2004) Genetic and functional analysis of the *cytK* family of genes in *Bacillus cereus*. *Microbiology*- **150:** 2689-2697

Feil SC, Polekhina G, Gorman MA, Parker MW (2010) Introduction. In *Proteins: membrane binding and pore formation* Anderluh G, Lakey J. (eds), Vol. 677, pp 1-13. Springer Science and Business Media, LLC

Fernandez SV, Xing J, Kapur V, Libby SJ, Barletta RG, Moxley RA (1998) Regulation of the *Escherichia coli* SheA gene and characterization of its encoded hemolytic activity. *FEMS Microbiol Lett* **168:** 85-90

Frenzel E, Letzel T, Scherer S, Ehling-Schulz M (2011) Inhibition of Cereulide Toxin Synthesis by Emetic *Bacillus cereus* via Long-Chain Polyphosphates. *Appl Environ Microbiol* **77:** 1475-1482

Garcia Arribas ML, Plaza CJ, de la Rosa MC, Mosso MA (1988) Characterization of *Bacillus cereus* strains isolated from drugs and evaluation of their toxins. *J Appl Bacteriol* **64:** 257-264

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, Walker JM (ed), pp 571-607

Gilbert RC (2010) Cholesterol-Dependent Cytolysins. In *Proteins: membrane binding and pore formation* Anderluh G, Lakey J. (eds), Vol. 677, pp 56-66. Springer Science and Business Media, LLC

Gilbert RJC (2002) Pore-forming toxins. *Cell Mol Life Sci* **59:** 832-844

Gohar M, Okstad OA, Gilois N, Sanchis V, Kolsto AB, Lereclus D (2002) Two-dimensional electrophoresis analysis of the extracellular proteome of *Bacillus cereus* reveals the importance of the PlcR regulon. *Proteomics* **2:** 784-791

GomezGomez JM, Blazquez J, Baquero F, Martinez JL (1996) Hns mutant unveils the presence of a latent haemolytic activity in *Escherichia coli* K-12. *Mol Microbiol* **19:** 909-910

Goñi FM, Ostolaza H (1998) *E. coli* α-hemolysin: a membrane-active protein toxin. *Brazilian Journal of Medical and Biological Research* **31:** 1019-1034

Gonzalez MR, Bischofberger M, Pernot L, van der Goot FG, Freche B (2008) Bacterial pore-forming toxins: The whole story. *Cell Mol Life Sci* **65:** 493-507

Gouaux E (1998) Alpha-hemolysin from *Staphylococcus aureus*: An archetype of beta-barrel, channel-forming toxins. *Journal of Structural Biology* **121:** 110-122

Granum PE, Brynestad S, O'Sullivan K, Nissen H (1993) The enterotoxin from *Bacillus cereus*: production and biochemical characterization. *Netherlands Milk Dairy J* **47:** 63-70

Granum PE, Lund T (1997) *Bacillus cereus* and its food poisoning toxins. *FEMS Microbiol Lett* **157:** 223-228

Granum PE, O'Sullivan K, Lund T (1999) The sequence of the non-haemolytic enterotoxin operon from *Bacillus cereus*. *FEMS Microbiol Lett* **177:** 225-229

Green J, Baldwin ML (1997) The molecular basis for the differential regulation of the *hlyE*-encoded haemolysin of *Escherichia coli* by FNR and HlyX lies in the improved Activating Region 1 contact of HlyX. *Microbiology* **143:** 3785-3793

Haug TM, Sand SL, Sand O, Phung D, Granum PE, Hardy SP (2010) Formation of very large conductance channels by *Bacillus cereus* Nhe in Vero and GH(4) cells identifies NheA + B as the inherent pore-forming structure. *J Membr Biol* **237:** 1-11

Hauge S (1955) Food poisoning caused by aerobic spore-forming bacilli. *J Appl Microbiol* **18:** 591–595

Heinrichs JH, Beecher DJ, Macmillan JD, Zilinskas BA (1993) Molecular-cloning and characterization of the *hbla* gene encoding the b-component of hemolysin bl from *Bacillus cereus*. *Journal of Bacteriology* **175:** 6760-6766

Hoffmaster AR, Ravel J, Rasko DA, Chapman GD, Chute MD, Marston CK, De BK, Sacchi CT, Fitzgerald C, Mayer LW, Maiden MCJ, Priest FG, Barker M, Jiang L, Cer RZ, Rilstone J, Peterson SN, Weyant RS, Galloway DR, Read TD, Popovic T, Fraser CM (2004) Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. *PNAS* **101:** 8449-8454

Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. *Nucl Acids Res* **38:** 545-549

Hunt ML, Cox AJ, Ruffolo CG, Rajakumar K, Adler B (2000) Characterisation of a Pasteurella multocida esterase gene which confers a hemolytic phenotype in *Escherichia coli* under anaerobic conditions. *FEMS Microbiol Let* **192:** 249-256

Hunt S, Moir AJG, Tzokov S, Bullough PA, Artymiuk PJ, Green J (2008) The formation and structure of *Escherichia coli* K-12 haemolysin E pores. *Microbiology* **154:** 633-642

Ivanova N, Sorokin A, Anderson I, Galleron N, Candelon B, Kapatral V, Bhattacharyya A, Reznik G, Mikhailova N, Lapidus A, Chu L, Mazur M, Goltsman E, Larsen N, D'Souza M, Walunas T, Grechkin Y, Pusch G, Haselkorn R, Fonstein M, Ehrlich SD, Overbeek R, Kyrpides N (2003) Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature* **423:** 87-91

Kotiranta A, Lounatmaa K, Haapasalo M (2000) Epidemiology and pathogenesis of *Bacillus cereus i*nfections. *Microbes Infect* **2:** 189-198

Kramer JM, Gilbert. RJ (1989) Foodborne Bacterial Pathogens. In *Bacillus cereus and other Bacillus species*, Dyle MP (ed), 21-70. New York: Marcel Dekker

Kyte J, Doolittle R (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**: 105-132

Lai XH, Arencibia I, Johansson A, Wai SN, Oscarsson J, Kalfas S, Sundqvist KG, Mizunoe Y, Sjostedt A, Uhlin BE (2000) Cytocidal and apoptotic effects of the ClyA protein from *Escherichia coli* on primary and cultured monocytes and macrophages. *Infect Immun* **68:** 4363-4367

Lindbäck T, Fagerlund A, Rodland MS, Granum PE (2004) Characterization of the *Bacillus cereus* Nhe enterotoxin. *Microbiology* **150:** 3959-3967

Lindbäck T, Hardy SP, Dietrich R, Sødring M, Didier A, Moravek M, Fagerlund A, Bock S, Nielsen C, Casteel M, Granum PE, Maertlbauer E (2010) Cytotoxicity of the *Bacillus cereus* Nhe enterotoxin requires specific binding order of its three exoprotein components. *Infect Immun* **78:** 3813-3821

Lindbäck T, Okstad OA, Rishovd AL, Kolsto AB (1999) Insertional inactivation of *hblC* encoding the L-2 component of *Bacillus cereus* ATCC 14579 haemolysin bl strongly reduces enterotoxigenic activity, but not the haemolytic activity against human erythrocytes. *Microbiology-Uk* **145:** 3139-3146

Linder R, Bernheimer AW (1991) Sourcebook of Bacterial Protein Toxins, Alouf JE, Freer JH (eds), pp 101–115. London: Academic Press

Ludwig A, Bauer S, Benz R, Bergmann B, Goebel W (1999) Analysis of the SlyA-controlled expression, subcellular localization and pore-forming activity of a 34 kDa haemolysin (ClyA) from *Escherichia coli* K-12. *Mol Microbiol* **31:** 557-567

Ludwig A, Tengel C, Bauer S, Bubert A, Benz R, Mollenkopf HJ, Goebel W (1995) SlyA, a regulatory protein from *Salmonella typhimurium*, induces a haemolytic and pore-forming protein in *Escherichia coli*. *Mol Gen Genet* **249:** 474-486

Ludwig A, von Rhein C, Bauer S, Huttinger C, Goebel W (2004) Molecular analysis of cytolysin A (ClyA) in pathogenic *Escherichia coli* strains. *J Bacteriol* **186:** 5311-5320

Lund T, De Buyser ML, Granum PE (2000) A new cytotoxin from *Bacillus cereus* that may cause necrotic enteritis. *Mol Microbiol* **38:** 254-261

Lund T, Granum PE (1996) Characterisation of a non-haemolytic enterotoxin complex from *Bacillus cereus* isolated after a foodborne outbreak. *FEMS Microbiol Lett* **141:** 151-156

Lund T, Granum PE (1999) The 105-kDa protein component of *Bacillus cereus* non-haemolytic enterotoxin (Nhe) is a metalloprotease with gelatinolytic and collagenolytic activity. *FEMS Microbiol Lett* **178:** 355-361

MacInnes JI, Kim JE, Lian CJ, Soltes GA (1990) *Actinobacillus pleuropneumoniae hlyX* gene homology with the fnr gene of *Escherichia coli J Bacteriol* **172:** 4587-4592

Madegowda M, Eswaramoorthy S, Burley SK, Swaminathan S (2008) X-ray crystal structure of the B component of hemolysin bl from *Bacillus cereus*. *Proteins-Structure Function and Bioinformatics* **71:** 534-540

Menestrina G, Serra MD, Pederzolli C, Bregante M, Gambale F (1995) Bacterial hemolysins and leukotoxins affect target cells by forming large exogenous pores into their plasma membrane. *Escherichia coli* hemolysin A as a case example. *Biosci Rep* **15:** 543-551

Mikkola R, Saris NEL, Grigoriev PA, Andersson MA, Salkinoja-Salonen MS (1999) Ionophoretic properties and mitochondrial effects of cereulide -

The emetic toxin of *B.cereus*. *European Journal of Biochemistry* **263:** 112-117

Montoya M, Gouaux E (2003) Beta-barrel membrane protein folding and structure viewed through the lens of alpha-hemolysin. *Biochim Biophys Acta Biomembr* **1609:** 19-27

Mueller M, Grauschopf U, Maier T, Glockshuber R, Ban N (2009) The structure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism. *Nature* **459:** 726-730

Oscarsson J, Mizunoe Y, Li L, Lai XH, Wieslander A, Uhlin BE (1999) Molecular analysis of the cytolytic protein ClyA (SheA) from *Escherichia coli*. *Mol Microbiol* **32:** 1226-1238

Oscarsson J, Mizunoe Y, Uhlin BE, Haydon DJ (1996) Induction of haemolytic activity in *Escherichia coli* by the *slyA* gene product. *Molecular Microbiology* **20:** 191-199

Oscarsson J, Westermark M, Lofdahl S, Olsen B, Palmgren H, Mizunoe Y, Wai SN, Uhlin BE (2002) Characterization of a pore-forming cytotoxin expressed by *Salmonella enterica* serovars Typhi and Paratyphi A. *Infect Immun* **70:** 5759-5769

Paananen A, Mikkola R, Sareneva T, Matikainen S, Hess M, Andersson M, Julkunen I, Salkinoja-Salonen MS, Timonen T (2002) Inhibition of human natural killer cell activity by cereulide, an emetic toxin from *Bacillus cereus*. *Clin Exp Immunol* **129:** 420-428

Parker MW, Feil SC (2005) Pore-forming protein toxins: from structure to function. *Prog Biophys Mol Biol* **88:** 91-142

Pexara A, Govaris A (2010) *Bacillus cereus*: an important foodborne pathogen. *J Hell Vet Med Soc* **61:** 127-133

Polekhina G, Giddings K, Tweten R, Parker MW (2005) Insights into the action of the superfamily of cholesterol-dependent cytolysins from studies of intermedilysin. *PNAS* **102:** 600-605

Prevost G, Cribier B, Couppie P, Petiau P, Supersac G, Finckbarbancon V, Monteil H, Piemont Y (1995) Panton-Valentine leucocidin and gamma-hemolysin from *Staphylococcus-aureus* ATCC-49775 are encoded by distinct genetic-loci and have different biological- activities. *Infect Immun* **63:** 4121-4129

Ralph ET, Guest JR, Green J (1998) Altering the anaerobic transcription factor FNR confers a hemolytic phenotype on *Escherichia coli K12*. *Proc Natl Acad Sci U S A* **95:** 10449-10452

Rasko DA, Ravel J, Okstad OA, Helgason E, Cer RZ, Jiang L, Shores KA, Fouts DE, Tourasse NJ, Angiuoli SV, Kolonay J, Nelson WC, Kolsto A-B, Fraser CM, Read TD (2004) The genome sequence of *Bacillus cereus* ATCC 10987 reveals metabolic adaptations and a large plasmid related to *Bacillus anthracis* pXO1. *Nucl Acids Res* **32:** 977-988

Reingold J, Starr N, Maurer J, Lee MD (1999) Identification of a new *Escherichia coli* SheA haemolysin homolog in avian *E. coli*. *Vet Microbiol* **66:** 125-134

Rosado C, S K, Bull T, Kuiper M, RH L, Buckle A, Voskoboinik I, Bird P, Trapani J, Whisstock J, Dunstone M (2008) The MACPF/CDC family of pore-forming toxins. *Cell Microbiol* **10:** 1765-1774

Rossjohn J, Feil S, McKinstry W, Tweten R, Parker M (1997) Structure of a cholesterol-binding, thiol-activated cytolysin and a model of its membrane form. *Cell* **89:** 685–692

Rowe GE, Welch RA (1994) Assays of hemolytic toxins. *Bacterial Pathogenesis, Pt A* **235:** 657-667

Ryan PA, Macmillan JD, Zilinskas BA (1997) Molecular cloning and characterization of the genes encoding the L (1) and L (2) components of hemolysin bl from *Bacillus cereus*. *Journal of Bacteriology* **179:** 2551-2556

Schmitt CK, Meysick KC, O'Brien AD (1999) Bacterial toxins: Friends or foes? *Emerging Infectious Diseases* **5:** 224-234

Setlow B, Setlow P (1995) Binding to DNA protects a/ß-type, small, acid-soluble spore proteins of *Bacillus* and *Clostridium* species against digestion by their specific protease as well as by other proteases. *J Bacteriol* **177:** 4149-4151

Slamti L, Lereclus D (2002) A cell-cell signaling peptide activates the PlcR virulence regulon in bacteria of the *Bacillus cereus* group. *EMBO Journal* **21:** 4550-4559

Soderblom T, Oxhamre C, Wai SN, Uhlen P, Aperia A, Uhlin BE, Richter-Dahlfors A (2005) Effects of the *Escherichia coli* toxin cytolysin A on mucosal immunostimulation via epithelial Ca^{2+} signalling and Toll-like receptor 4. *Cell Microbiol* **7:** 779-788

Song LZ, Hobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux JE (1996) Structure of *Staphylococcal* alpha-hemolysin, a heptameric transmembrane pore. *Science* **274:** 1859-1866

Stanley P, Koronakis V, Hughes C (1998) Acylation of *Escherichia coli* hemolysin: A unique protein lipidation mechanism underlying toxin function. *Microbiol Mol Biol Rev* **62:** 309-333

Stanley P, Packman LC, Koronakis V, Hughes C (1994) Fatty acylation of 2 internal lysine residues required for the toxic activity of *Escherichia coli* hemolysin. *Science* **266:** 1992-1996

Steinthorsdottir V, Halldorsson H, Andresson OS (2000) *Clostridium perfringens* beta-toxin forms multimeric transmembrane pores in human endothelial cells. *Microb Pathog* **28:** 45-50

Tilley SJ, Orlova EV, Gilbert RJC, Andrew PW, Saibil HR (2005) Structural basis of pore formation by the bacterial toxin Pneumolysin. *Cell* **121:** 247-256

Titball RW (1999) Membrane-damaging and cytotoxic phospholipases In *The Comprehensive Sourcebook of Bacterial Protein Toxins.*, Alouf JE, Freer JH (eds), 2nd edn, pp 310-329. London: Academic Press

Turner P, McLennan A, Bates A, W. M (2005) *Molecular Biology Instant Notes*, 3rd edn. New York, USA: Taylor and Francis.

Tweten RK (2005) Cholesterol-dependent cytolysins, a family of versatile pore-forming toxins. *Infect Immun* **73:** 6199-6209

Tzokov SB, Wyborn NR, Stillman TJ, Jamieson S, Czudnochowski N, Artymiuk PJ, Green J, Bullough PA (2006) Structure of the hemolysin E (HlyE, ClyA, and SheA) channel in its membrane-bound form. *J Biol Chem* **281:** 23042-23049

Wai SN, Westermark M, Oscarsson J, Jass J, Maier E, Benz R, Uhlin BE (2003) Characterization of dominantly negative mutant ClyA cytotoxin proteins in *Escherichia coli*. *J Bacteriol* **185:** 5491-5499

Wallace AJ, Stillman TJ, Atkins A, Jamieson SJ, Bullough PA, Green J, Artymiuk PJ (2000) *E. coli* hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy. *Cell* **100:** 265-276

Westermark M, Oscarsson J, Mizunoe Y, Urbonaviciene J, Uhlin BE (2000) Silencing and activation of ClyA cytotoxin expression in *Escherichia coli*. *J Bacteriol* **182:** 6347-6357

Wyborn NR, Clark A, Roberts RE, Jamieson SJ, Tzokov S, Bullough PA, Stillman TJ, Artymiuk PJ, Galen JE, Zhao L, Levine MM, Green J (2004) Properties of haemolysin E (HlyE) from a pathogenic *Escherichia coli* avian isolate and studies of HlyE export. *Microbiology* **150:** 1495-1505

Yadav SP, Ahmad A, Pandey BK, Verma R, Ghosh JK (2008) Inhibition of lytic activity of *Escherichia coli* toxin hemolysin E against human red blood cells by a leucine zipper peptide and understanding the underlying mechanisms. *Biochemistry* **47:** 2134-2142.

Chapter 2: Theory of X-Ray Crystallography of Biological Macromolecules

X-ray crystallography is a technique that is used to determine the three dimensional structures of biological samples that are in a crystalline state. It has mainly been used to determine molecular or atomic models of small molecules or nucleic acids or protein macromolecules. X-ray crystallography is unequivocally the most used structural determination technique. It surpasses nuclear magnetic resonance (NMR) and Cryo-electron microscopy by thousands in the number biological structures deposited in the Protein Data Bank (Berman et al, 2000).

This chapter provides a brief overview of the practice and theory of X-ray crystallography as a biological macromolecular structure elucidation technique and main technique that was used in this thesis. The information was mostly taken from (Blundell & Johnson, 1976), (Rhodes, 2006), (Blow, 2002), (Drenth, 1999) (Rupp, 2010) or otherwise as cited.

2.1 Sample preparations and techniques in protein crystallisation

The major objectives in X-ray crystallography are expression of enough quality protein and determining suitable conditions for crystal growth. There are no general rules for protein crystallisation experiments; and so numerous conditions have to be tried. There are generally three agreed stages to proteins' crystal growth: nucleation, growth or post-nucleation, and cessation of growth. Once the nucleus has been formed, protein particles move from the bulk of the solution and start to attach and incorporate in ordered layers to form a crystal, which can grow to macroscopic dimensions (Figure 2.1).

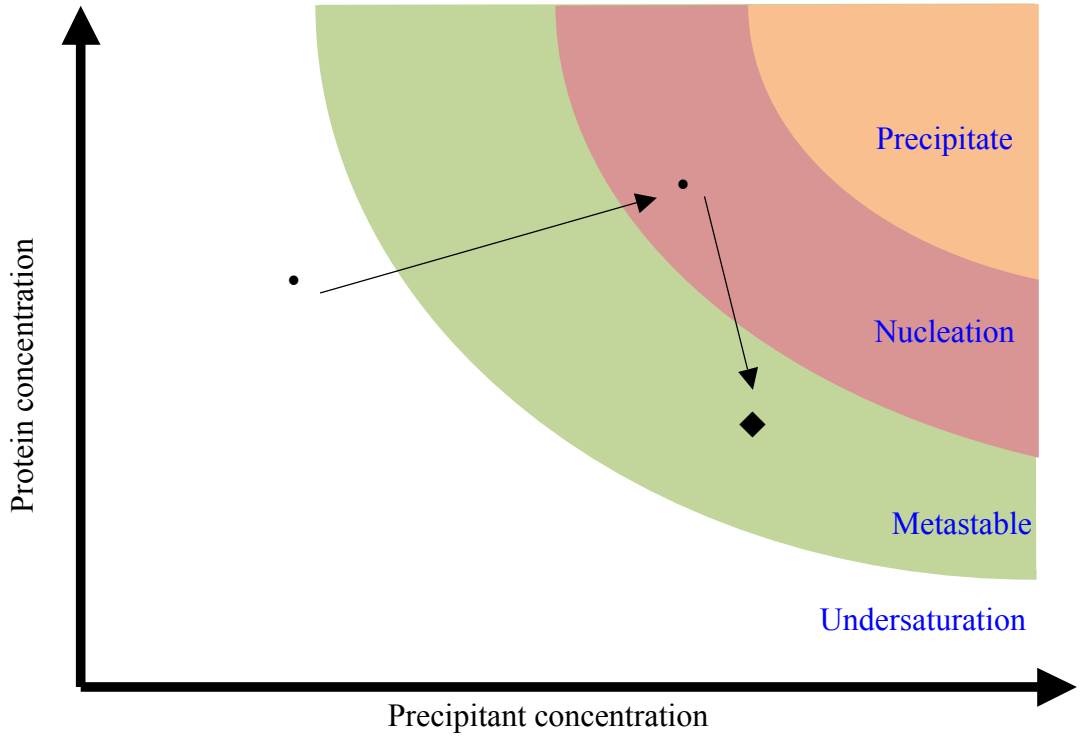


Figure 2.1: Schematic illustration of a protein crystallisation phase diagram.
The plot shows the protein concentration against precipitant concentration during a crystallization experiment. The black circles represent the protein and precipitant. As the drop equilibrates with the reservoir solution the protein and precipitant concentration will be increased and nuclei can be formed, then protein particles move from the bulk of the solution and the crystal can be growth. Adapted from Rhodes (2006).

Depending on whether the protein is soluble or a membrane protein some of the variables that can affect crystal growth include: pH, temperature, detergents, metal, ions, purity, homogeneity, protein concentration and the presence of numerous additives and their concentrations. Optimising each of these parameters can take a very long time, hence the common use of commercial kits to determine initial nucleation and growth conditions.

2.1.1 Sample preparation

Sample preparation is the most important step in order to get protein crystals. High concentrations of protein in the range of 10 mg/ml are often ideal. The purity and homogeneity of the sample are of paramount importance: up to 95% purity is required. It is usual to purify the protein with more than one purification step in order to improve its purity and homogeneity.

2.1.2 Crystallisation techniques

There are many techniques that are used to crystallise proteins, all of them aiming at inducing supersaturated solutions of protein macromolecules.

2.1.2.1 Batch technique

The batch technique is probably one of the first protein crystallisation techniques developed in 1930s. This technique involves adding a protein precipitant to large volume (about 1 ml) of highly concentrated protein until a slight opalescence was observed (Rayment, 2002). This technique fell out of favour because of its high requirement for large volumes at high protein concentration.

2.1.2.2 Vapour diffusion methods

The vapour diffusion method became popular in early 1960s and is still the most commonly used method today. The method involves mixing the protein and crystallisation solution in a one to one ratio and placing the mixture in a sealed environment close to a reservoir of a precipitant. Because the concentration of the crystallisation mixture is lower than that the reservoir, water/solvent normally diffuses out from the sample into the reservoir. This leads to increase in the concentration of protein and the precipitants, which causes the protein to come out of the solution. The vapour diffusion method is performed as either hanging drop or sitting drop (Figure 2.2).

2.1.2.2.1 Hanging drop method

Because of the large number of conditions one needs to screen in order to grow protein crystals, vapour diffusion methods are designed use small quantities to conserve samples. The hanging drop is performed on a pre-siliconised coverslip (Figure 2.2 A). Normally, 1 to 10 µl of purified protein solution at high concentration are mixed with crystallisation solution in 1:1 ratio. This mixture is then placed on the clean pre-siliconised coverslip, which is then inverted and placed above a small well containing 500 µl to 1 ml of the crystallisation buffer.

2.1.2.2.1 Sitting drop method

Obviously trying to load each crystallisation buffer and mixing it with the protein in 1:1 ratio for hundreds of crystallisation conditions and placing them one by one on coverslips is very time consuming. With the sitting drop method (Figure 2.2 B), sample loading is automated and can be handled using crystallisation robots. The principles of sitting drop are similar to those of hanging drop except that the 1:1 mixture is loaded in a depression

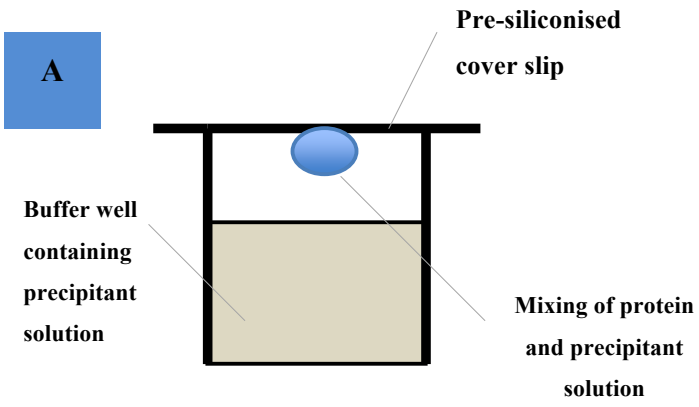
of a micro-bridge. Similar to the hanging drop method, the container is airtight sealed with grease, oil or transparent sticky tape.

2.1.2.3 Dialysis

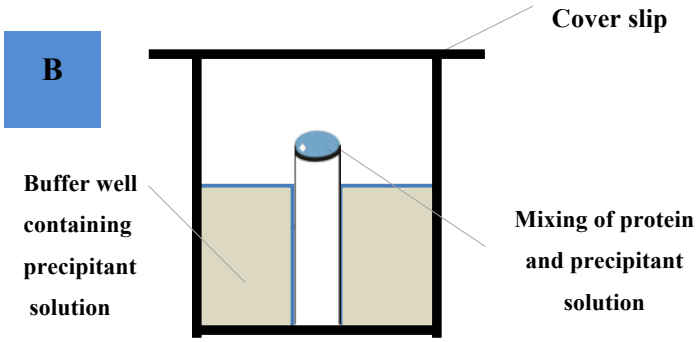
The dialysis method is commonly used with a "dialysis buttons". The dialysis button's well is loaded with a protein sample and then covered with a dialysis membrane. It is then placed in a small container. The precipitating solution is then added. The selectively permeable membrane covering the dialysis button only allows passage of precipitating solution into the protein sample. As the concentration of precipitating solution increases inside the membrane, nucleation can occur, leading to protein crystal growth. The precipitating solution can be changed over time to a fresh one. The container is usually sealed airtight with grease.

2.2 Protein crystal mounting

The crystal must then be mounted so that it can be exposed to an X-ray beam. The quality of crystals can be judged by sharp diffraction patterns at large angles emanating from X-ray beam. Protein crystal mounting is a delicate step that needs a great degree of care. Crystals are normally mounted in their mother liquor, the solution in which the crystal growth took place. This is very important to avoid the crystals cracking or disintegrating; rendering them useless for data collection and also to keep them hydrated. Crystal used to be mounted in sealed capillary tubes (see below) but now one of the commonly used techniques of mounting crystals is to freeze them in a loop.



Hanging drop crystallization method.



Sitting drop crystallization method.

Figure 2.2: **Schematic illustration of vapour diffusion protein crystallization methods**.
(A) The hanging drop method and (B) the sitting drop method that used in this thesis.

2.2.1 Room temperature mounting

During the early days of protein crystallography, when cryogenic techniques were not fully developed, crystals were mounted using capillary tubes. The protein crystal was slowly and gently drawn into the capillary tube and droplets of mother liquor were introduced on either side and the tube was then sealed. The capillary tube was then mounted on a goniometer head, which controls the crystal's orientation in the X-ray beam (Figure 2.3).

The biggest problem associated with collecting diffraction data from protein crystals mounted at room temperature is radiation damage. This means that a lot of crystals were needed to collect enough data for structure determination. To overcome this problem crystals can be soaked in cryoprotectants and then be cooled to liquid nitrogen temperature.

2.2.2 Cryogenic mounting (Cryocrystallography)

Cryo-crystallography was introduced to minimise radiation damage to the crystals. This technique, which is commonly applied these days, involves introducing a stream of nitrogen gas at 100 K or plunging the crystal into the liquid nitrogen. The liquid nitrogen leads to solvent vitrification of the disordered liquid phase. The major drawback for this technique is that the liquid nitrogen can lead to formation of ordered ice crystals, which can disrupt the protein crystal and lead to poor diffraction data. To overcome this problem, cryoprotectants are introduced into the mother liquor to prevent ice crystal formation. Commonly used cryoprotectants include glycerol, low molecular weight polyethylene glycol (PEGs), ethylene glycol and 2-methyl-2, 4-pentanediol (MPD). It is necessary to optimise the freezing stage and/ or the type/concentration of cryoprotectants as both can degrade the protein crystal.

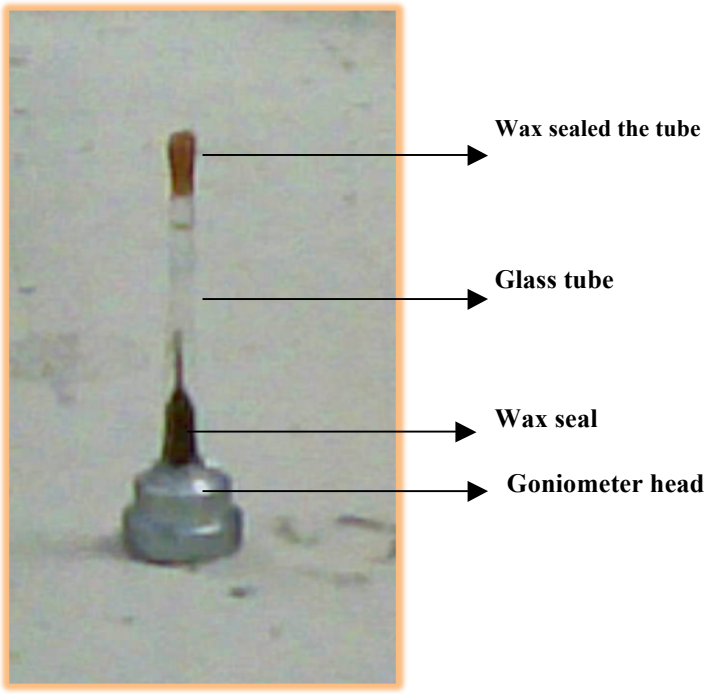


Figure 2.3: Room temperature mounting.
The protein crystal was drawn into the capillary tube and droplets of the mother liquor were introduced on either side and then sealed. The capillary tube was then mounted on a goniometer head, which controls the crystal's orientation in the X-ray beam.

2.3 X-Ray generation

X-ray crystallography can generally be described as a visualisation method for macromolecules that are in the crystalline state. In order for any object to be visualised, it needs to scatter light. The electromagnetic radiation of visible light, which has a wavelength of 400 to 700 nm, cannot produce images of individual atoms of proteins because they are much smaller than this. The typical bond length of atoms in proteins is 0.15 nm. X-rays are electromagnetic radiation and useful in elucidating atomic structures of proteins because they have a wavelength of 10 to 0.01 nm and are easy to produce. When X-rays are used at a wavelength of 0.1 nm (1 Å), they can resolve chemical covalent bonds and show the positions of atoms.

In most crystallographic structural biology laboratories, X-rays are produced by rotating anode tubes (Figure 2.4), but more intense X-rays are produced by electron synchrotrons (Figure 2.5). Whichever way the X-rays are produced and for whatever purpose that are intended to be used for, there are three main requirements to produce X-rays: an electron source, the acceleration of the electrons and the target that will eventually produce X-rays upon interaction with the electrons.

2.3.1 Rotating anode tubes

X-rays are generated by bombarding a metal target with electrons that have been produced from a heated filament (usually a tungsten filament) and propelled through an electric field. An X-ray tube is just a sealed container with an anode at one end and a cathode on the other end, maintained in a vacuum. The cathode is heated, which allows the electrons to be released and propelled through the electric field towards the anode. The anode is normally made of copper, molybdenum, or chromium. Copper is normally preferred due to its high heat conductivity. When electrons are released from the cathode, if they have sufficient energy, will knock off electrons in the target metal. The electronic shells of atoms are designated K, L, M, N, O, P,

and Q; starting from the innermost shells outwards. The energy level spacing of the metal target produces specific wavelengths that are typically characteristic of that particular metal. The lower the atomic number, the longer the emission lines of the metal. When electrons are knocked out from the innermost K electronic shell, and then are replaced with the electrons from L electronic shells, a K_α emission is observed. When the electrons in the K shell are replaced by electrons from M shell, a K_β emission is produced. In crystallography a single source of X-rays is desirable. This is normally achieved by filtering the weaker K_β emissions e.g. using graphite crystals or Ni for copper emissions.

2.3.2 Synchrotrons

Synchrotrons are the most powerful sources of X-rays. They produce much more intense X-rays ranging from 0.5 Å to 1.8 Å in wavelength. They are very expensive compared to rotating anode tubes. X-rays are produced by accelerating electrons in a vacuum, which are travelling near the speed of light, in storage rings. These rings consist of straight and curved parts much like a polygon, which can go up to 1.5 km in circumference. At each corner of this polygon, powerful electromagnets are placed. Intense radiation is produced either by bending magnets or with the help of insertion devices such as Wigglers or Undulators. These devices change the trajectory of the electrons, which leads to production of X-rays (Rhodes, 2006; and Rupp, 2010). This is achieved by alternating magnetics placed in north and south poles adjacent and opposite to each other. The X-rays are emitted tangentially to the direction of the electrons circular path around the synchrotron. These X-rays are filtered so that only monochromatic X-rays of specific wavelength are allowed to reach the protein sample on the beamline.

Synchrotron radiation has the major advantage of having a tuneable wavelength, which is a very useful property for phase determination using Multi-wavelength Anomalous Dispersion (MAD) or Single -wavelength

Anomalous Dispersion (SAD). Further, due to the intensity of X-rays, the exposure times are greatly reduced, hence huge amount of data can be collected before the protein crystals are completely damaged due to radiation. This also allows collecting data from almost any size of crystals and in some instances even from the crystals that showed poor diffractions pattern when using Rotation anode tubes.

2.4 X-Ray scattering

The interaction of X-rays with the electrons in protein crystals leads to scattering. The scattered waves are normally 180º out of phase to the incident wave, although they maintain the same wavelength. Two major types of scattering that can occur are:

1) Thomson scattering, also commonly known as elastic or coherent scattering. This kind of scattering is characterised by no change of energy between the incident beam and the scattered beam, although its direction may be changed. This kind of scattering is very useful to X-ray crystallography because when the incident X-ray photon collides with an electron in the crystal, the resultant oscillations produce radiation of frequency identical to that of the incident wave as the electron returns to its original position.

2) The other kind of scattering is inelastic scattering, also known as Compton or incoherent scattering. This kind of scattering leads to loss of energy; therefore the scattered wave has less energy and a changed wavelength.

Figure 2.4: A Rigaku Micromax 007 copper rotating anode generator used in our lab.
X-rays come from the left from a rotating anode source (not shown), and emerge through a metal collimator (1) to strike the crystal, which is placed in a cryoloop on the goniometer (2). An image plate detector (3) captures the reflections. The crystal is kept frozen by a stream of cold nitrogen gas coming from a liquid nitrogen supply (4).

Figure 2.5: Aerial view of Diamond Light Source, the UK's national synchrotron facility, in south Oxfordshire. Image courtesy Diamond Light Source.

2.4.1 Atomic scattering

The incident X-ray beam is scattered by the electrons around an atom. An atom with Z electrons (its atomic number) will be expected to scatter X-rays Z times as strongly as a single electron. The atomic scattering factor *f* is defined as the ratio between the amplitudes scattered by an atom and by a single electron (Hammond, 2009). The atomic scattering factor can be plotted against $sin\theta/\lambda$ for a particular atom. The higher the resolution (or the bigger the angle θ), the smaller is the atomic scattering factor. The atomic scattering factor for the entire atom is a description of the individual scattering contributions of each electron expressed as;

$$f = \int_r \rho(\mathbf{r}) \exp(2\pi \mathbf{r} \bullet \mathbf{S}) dr$$

where ρ is the electron density, **r** is the distance from the centre of the atom and **S** is the scattering vector. The amplitude of the scattering vector is expressed as

$$|S| = (2 \sin \theta/\lambda)$$

The atomic scattering factor is commonly affected by the atomic fluctuations caused by temperature changes. Atoms are not at a fixed position at all the time, but they are vibrating with respect to changes in temperature. Hence at higher diffraction angles, the scattering factor is greatly reduced due to these thermal vibrations. These thermal vibrations are corrected using the temperature factor (*Debye-Waller temperature factor*) or B-factor either at low resolution using isotropic temperature factor corrections or at near to atomic resolution with the anisotropic B-factor refinement. The B-factor can also explain variation due to static disorders (different conformations in different unit cell) as well as thermal fluctuations.

2.4.2 Unit cell scattering (Structure factors)

In order to determine the three-dimensional structure of a protein confined in a unit cell, the three key factors (direction of the beam, amplitude and phases) associated with each scattered or diffracted beam must be determined. The unit cell is the smallest group of molecules, which can be repeated in three dimensions by translation to produce a crystal lattice. Thus to determine the scattering amplitudes in the unit cell, it requires summing up all the amplitudes from all atoms within that unit cell, taking into account the phases and direction of the scattered waves. For a unit cell made up of *n* atoms located at position $\mathbf{r}_j$ in the unit cell, where *j* is an integer from 1 to n (1, 2, 3, …n) and **r** is the distance between the origin and atoms, the sum of diffraction of the unit cell is described as:

$$\mathbf{F} = f \exp(2\pi i \mathbf{r} \bullet \mathbf{S})$$

Where F(**S**) is the structure factor, **S** is the scattering vector and $\mathbf{2\pi r.S}$ is the phase difference between the scattered wave at position **r** and the origin.

2.4.3 Molecular scattering

Similarly as described for unit cell scattering, the structure factor for molecular scattering is the sum of all structure factors originating from scattering of individual atoms given as:

$$\mathbf{G(S)} = \sum_{j=1}^{atoms} f_j \exp(2\pi i \mathbf{r}_j \bullet \mathbf{S})$$

Thus the scattering of X-rays by each atom gives structure factor of molecular scattering G (S), where $2\pi \mathbf{r}_j \bullet \mathbf{S}$ is the phase difference between scattered waves at position $\mathbf{r}_j$ and the origin in a molecule.

2.4.4 Crystal scattering

Protein crystals are ordered arrays; hence they can amplify the scattering factor to detectable levels, only if the unit cells in crystals scatter in phase with respect to one another. The phase difference between adjacent unit cells in one dimension is described by **2πa.S** where **a** represents the shift in position with respect to the origin between the unit cells. The ordered arrays of unit cells will only scatter in phase if **2πa.S** is equal to **2π*h***, in other words when **a.S = *h***, where ***h*** is an integer.

Thus in three dimensions, the following Laue conditions are met:

$$\mathbf{a.S} = \boldsymbol{h}$$

$$\mathbf{b.S} = \boldsymbol{k}$$

$$\mathbf{c.S} = \boldsymbol{l}$$

h, k and ***l*** being integers. The *h, k* and *l* represent discrete reflections from the set of planes and each atom in the unit cell. Hence they describe the reciprocal lattice of the diffraction pattern of the unit cell.

Taking the positional vector of $\mathbf{r}_j$ from molecular structure factor, the fractional coordinates in the unit cell can be expressed as (ax_j, by_j, cz_j). This gives the structure factors for the unit cell with regard to the Laue conditions as:

$$\mathbf{F}(hkl) = \sum_{j=1}^{N} f_j \exp[2\pi i(hx_j + ky_j + lz_j)]$$

Here it basically means x_j, y_j, and z_j are the fractional coordinates of the j^{th} atom in the unit cell. Thus the diffraction pattern of any crystal can be elucidated from this equation given the positions of all atoms in the unit cell. Indeed, the amplitude that is measured experimentally is the scattered X-ray of each of these reflections set, *h, k* and *l*. Since experimentally the measured reflections are *h*, *k* and *l* sets, the two most important pieces of information for protein structure determination (determining an electron

density map), the structure factor amplitudes and phases, can be represented in a single equation as;

$$\mathbf{F}\ (hkl) = |F(hkl)|e^{i\alpha(hkl)}$$

The structure factor amplitude is represented by |F(hkl)| and the phases by α(hkl). Thus, once the amplitudes and phases are determined, the electron density map of the protein can be obtained, but in X-ray crystallographic data collection the phases are lost. This is because the intensities of the diffraction pattern recorded on films or counters are proportional to the square of the amplitudes. Therefore all the phase angle information of the diffracted X-rays are lost. This is a major problem in X-ray crystallography.

2.5 Bragg's law

Bragg's Law, which won W. L. Bragg and W. H. Bragg with the Physics Nobel Prize of 1915, describes the X-ray diffraction of crystalline materials. As described in 2.4.2 a crystalline state is made up of ordered arrays of atoms periodically repeating themselves to form a lattice. The crystalline state can reflect incident X-rays as if it is a mirror. In order to understand Bragg's Law, let us consider two incident X-rays at angle θ on atoms Z and B separated by distance d (Figure 2.6). Assuming that the interference is constructive, and therefore the phase shift is a multiple of 2π, then the Bragg's Law is simply defined as $\mathbf{n\lambda = 2dsin\theta}$, where n is an integer, λ is the wavelength of the X-rays, d is the distance between the diffraction and θ is the angle of incidence of the incoming X-rays (Blow, 2002), this is shown in (Figure 2.6).

When the two incident waves are travelling in phase and parallel to each other, they will strike atoms Z and B. The first incident beam will strike atom Z in the top plane. The second beam will continue travelling and strike atom B on the bottom plane. Thus the second beam travels an extra distance AB + BC. For the two beams to be in phase, the distance AB + BC is the integer number of nλ:

Thus $AB + BC = n\lambda$

By taking the interatomic distance d, as the hypotenuse of right–angled triangle formed by AB and the diffracting beam, the distance AB becomes:

$$AB = d \sin\theta$$

Since AB = BC, AB+BC= 2AB, and

$$n\lambda = 2AB$$

Combining the two gives the Bragg's law, $\mathbf{n\lambda = 2dsin\theta}$.

Thus x-ray diffraction is observed when the interplanar distance (d), the incidence angle (θ) and the wavelength of the incident X ray ($\boldsymbol{\lambda}$) are satisfied as stated in the Bragg's Law.

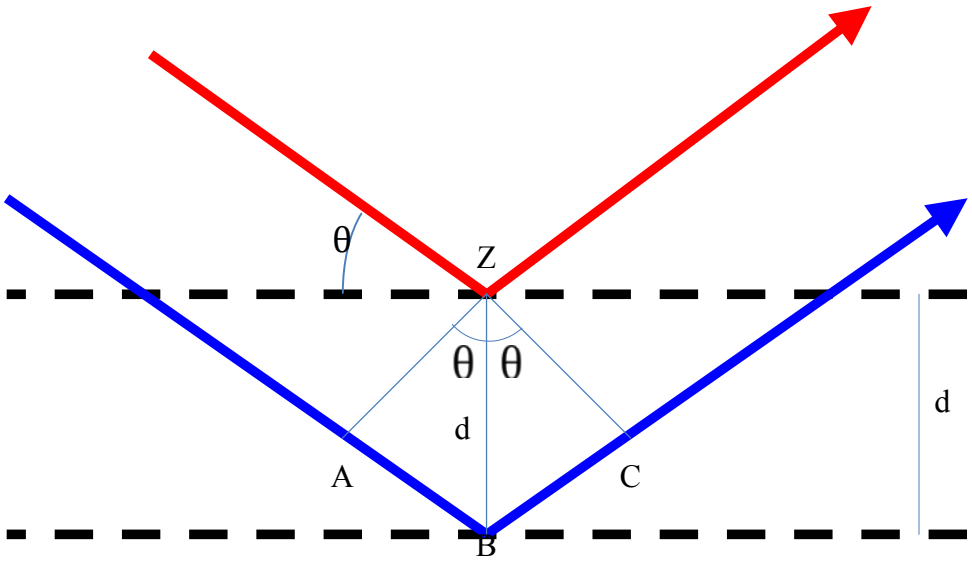


Figure 2.6: **The incident X-rays are scattered by atoms Z and B, which are in interplanar distance d.**
AB and BC are each equal to dsinθ. AB +BC = an integral number of wavelengths, the path difference between the red and blue lines is 2AB, which is equal to **2dsinθ.** Thus the diffraction takes place. i. e. when **2dsinθ= nλ** which is Bragg's law.

2.6 Data collection

The phase of diffracted X-rays cannot currently be measured but their position and intensity can. There are generally two main detectors used widely in X-ray crystallography at present: image plate detectors and Charged Coupled Device **(**CCD)-based detectors.

2.6.1 Image plate detectors

These are the most commonly found detectors in most X-ray laboratories. They are coated with a thin layer of inorganic phosphor (BaFBr) on a flat base. The top layer of the inorganic phosphor is doped with europium (Eu^{2+}). When the incident X-rays from the protein crystal strike the image plate, the Eu^{2+} is excited to a metastable Eu^{3+}. Reversing the metastable Eu^{3+} to Eu^{2+} using red laser light of 633 nm wavelength releases radiation of 390 nm wavelength. The intensity of the emission at 390 nm wavelength is proportional to the quantity of the X-rays absorbed by europium, which gives a good dynamic range and uniform sensitivity across the image plate. The emission of metastable europium is detected by a photomultiplier, which reads the diffraction patterns and forms the diffraction image. Image plate detectors have two major disadvantages of having a long readout dead time of up to two minutes and low sensitivity. Thus they are not ideally suited to synchrotron experiments, which require very quick readout time; otherwise the crystals will end up damaged from radiation.

2.6.2 CCD cameras

All synchrotron X-ray data collection throughout this thesis was conducted by using Charged Coupled Device **(**CCD)-based detectors. These are the commonest detectors found at synchrotrons, but are rarely available in the in-house X-ray laboratories. The CCD-based detectors are made up of a scintillator screen, a fibre optic taper and CCD chip. The scintillator screen

has a phosphor used to convert X-rays to visible light/photons. The light photons are then transferred by the fibre optics to the CCD chip, where they are digitised. CCD detectors have a major advantage of quick readout times; hence they are available for use immediately after the exposure.

2.7 Data processing

Once the diffraction patterns have been collected from the protein crystals, they are processed by various programs to extract information, including intensities and locations of reflections. The process of data processing from diffraction data involves space group assignment, unit cell parameters, refinement, integration of the intensity values, sorting, and scaling. There are many programs that are used to process the diffraction data such as MOSFLM/iMOSFLM, (Leslie, 1994), X-ray Detector Software (XDS) (Kabsch, 2010), HKL2000 (Otwinowski & Minor, 1997) and SCALA (CCP4, 1994). The XDS, iMOSFLM and CCP4 programme suites were used in this thesis.

2.7.1 Autoindexing

Indexing is the first step of processing any image data set, usually by autoindexing. During autoindexing, the space group of the crystal lattice and unit cell dimensions are determined. Autoindexing can be done with spots from one or multiple images. Two images that are 90^{o} apart are usually a good starting point.

2.7.2 Refining crystal parameters and dataset integration

After indexing the data, the final refined matrix is used by iMOSFLM to predict the reflection positions. Prerequisite to this prediction is accurate unit cell dimensions, which are obtained during indexing. The refinement is

normally done before integration, although if it's done after integration, then the data need to be reintegrated. The integration involves calculating estimates of the reflection intensities and determining the standard errors from the raw data. Integration is performed by either summation or profile fitting. The result of this process is a single *.mtz file that is used in subsequent data processing.

2.7.3 Sorting, scaling and data merging

The reflection indices (*h,k,l*) assigned by iMOSFLM during data integration are assorted in numerical ascending order using the programme SORTMTZ (CCP4, 1994). The file generated from SORTMTZ (CCP4, 1994) is analysed with SCALA in order to scale and merge the data. SCALA calculates the scale factor between each image with respect to a specific image of the data set. It tries to put all the images into a common observation scale. This scale is used as a diagnostic tool for analysis of data quality and to check if the data collection was satisfactory.

SCALA outputs a series of statistics, which are used to assess the data quality with increasing resolution. The main indicators are: R_{sym} or R_{merge}, which is the agreement between symmetry equivalent reflections, I/σI; which is the ratio of intensity to standard deviation of the intensity; and the completeness. Both R_{merge} and R_{sym} are given by:

$$R_{sym} = \frac{\sum_N \left[\sum_N \left(I_i(hkl) - \bar{I}(hkl) \right) / n \right]}{\sum_N \bar{I}(hkl)}$$

Where $I_i(hkl)$ is the i^{th} measurement of reflection *hkl* and $\bar{I}(hkl)$ is the mean value of the n equivalent reflections. The summation is come out over all N independent reflection in the data set.

I/σI provides details of the signal to noise ratio of the data, which should be greater than one for highest resolution data shell. Theoretically how much of the possible data to a given resolution has actually been collected is given by the completeness. This should be at least 90% of completeness to a given resolution.

In the higher resolution shells the data become weaker and I/σI becomes less. This means that the agreement between equivalent reflections becomes poores and therefore R_{sym} becomes bigger.

2.8 The phase problem and its solution in X-ray Crystallography

The biggest challenge in X-ray crystallography is to determine phases, which are needed together with the intensities of reflections to elucidate the electron density map. If a similar structure already exits then Molecular Replacement (MR) can be used. Otherwise the phases must be obtained experimentally using Multiple Isomorphous Replacement (MIR) or Single Isomorphous Replacement (SIR) or Multi-wavelength Anomalous Dispersion (MAD) or Single -wavelength Anomalous Dispersion (SAD). This can be one of the toughest stages during macromolecule structure determination. In this thesis, only MAD was used to elucidate the protein structure.

2.8.1 The Patterson function

The Patterson function, originally proposed by Arthur Lindo Patterson in 1934 at Massachusetts Institute of Technology, is used to determine atomic positions by utilising intensities as coefficients rather than structure factors. It is a Fourier transform (FT) of the reflection intensities assuming all phases are equal to zero. The Fourier transform used to calculate electron density maps, when all phases are set to zero, will be equal to:

$$P(uvw) = \frac{1}{V}\sum_{hkl}\left|F(hkl)\right|^2 \cos[2\pi(hu + kv + lw)]$$

where *u*, *v* and *w* are Patterson cell coordinates that locate a point on the Patterson map.

The Patterson function is a convolution of electron density with itself. The peaks of the Patterson function represent interatomic distance vectors both within the same molecule and between different molecules in the crystal. Thus if a molecule contains N atoms, then considering only the intra-molecular vectors, N^2 peaks will be observed in the Patterson function. The Patterson function is very useful for determining the structures of small molecules but not macromolecules like proteins because there are just too many atoms however it can be used to locate a small numbers of heavy atoms bound to a protein. The Patterson function is very useful in other phasing techniques such as MR to perform the rotation and translation functions, and in both MAD and MIR for locating heavy atom positions of derivative crystals.

2.8.2 Multiple isomorphous replacement (MIR)

Multiple isomorphous replacement experiments determine phases by comparing scattering from two almost identical protein crystals, except that one has one or more strong scattering centres. The strong scattering centres are made up of heavy atoms. These atoms could be naturally be part of the protein like its natural ligand/substrate or introduced into the protein through co-crystallisation or soaking of the crystal with heavy atoms. Typical heavy atoms that are used include Hg^{2+} ions which are used to bind thiol groups, uranyl salts which bind between carboxyl groups in Asp and Glu, Pb binds to Cys, $PtCl_4^{2-}$ anions bind to His and tryptophan binds to $Hg(OAc)_2$. When these heavy metal ions bind to the protein crystal's asymmetric unit, it is referred as a derivative crystal, while a metal free protein crystal is called a

native. There are several criteria that define heavy-atom derivative crystals such as *isomorphism,* which generally means that the presence of the heavy atom does not disturb the protein structure or crystal packing. Isomorphism further implies that the unit cell dimensions of the derivative and native crystal are the same. A second criterion is that there must be detectable and measureable changes of reflection intensities from which phases can be determined. Finally, it is important that the derivative crystal is able to diffract to a fairly high resolution, although not necessary as high as the native, but just enough to extract important information.

Considering that the reflection amplitude of native protein is given as $|\mathbf{F_P}|$ and that of the derivative as $|\mathbf{F_{PH}}|$, then the difference in amplitude ($|\mathbf{F_{PH}}| - |\mathbf{F_P}|$) approximate to the amplitude of a heavy atom alone ($|\mathbf{F_{PH}} - \mathbf{F_P}|$). Computation of the diffraction pattern of $(|\mathbf{F_{PH}} - \mathbf{F_P}|)^2$, will give the diffraction pattern of the heavy metal alone. The structure factor of the derivative can be defined as the sum of:

$$\mathbf{F_{PH}} = \mathbf{F_P} + \mathbf{F_H}$$

This equation allows to solve the phases of F_P and F_{PH} using the Harker construction. See Figure 2.7 and 2.8 for phase solutions using isomorphous replacement.

2.8.3 Multi-wavelength anomalous dispersion (MAD)

Multi-wavelength Anomalous Dispersion (MAD) (Rhodes, 2006; and Rupp, 2010) was used to solve the protein structure in this thesis. The reason for choosing this technique and not MIR is that in MIR several crystals are required, and in addition lack of isomorphism is usually very common between heavy metal ion derivatives and native crystals. MAD has the advantage that non-isomorphism is not a problem, and only one crystal is required.

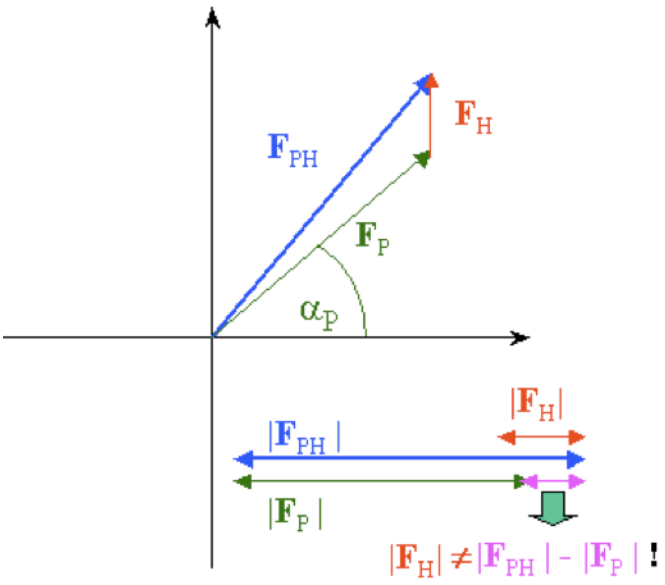


Figure 2.7: **The vector diagram illustrates the relationship between native and derivative scattering factors**.
The ultimate aim of phase determination is to determine the unknown phase α_P of each reflection. Colour key: the structure factor whose phase is lost (the native protein) is green. The structure factor of protein heavy atom derivative is blue and the structure factor of heavy atom is red.
Taken from http://www.ruppweb.org/ accessed on 20 September 2011. With permission from Bernhard Rupp.

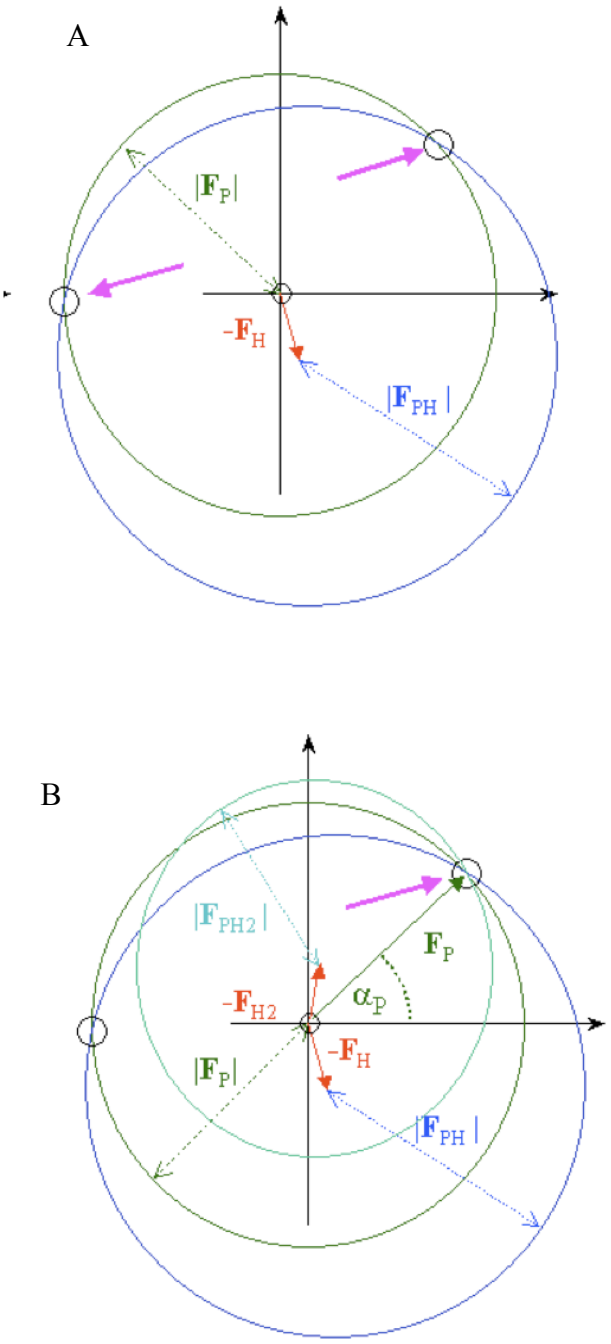


Figure 2.8: **Phase determination.**
(A) for single Isomorphous Replacement ; (B) for Multiple Isomorphous Replacement. Colour key: the structure factor whose phase is lost (the native protein) is green. The structure factor of protein heavy atom derivative is blue and the structure factor of the heavy atom is red. Purple arrows indicate phase solution. Taken from http://www.ruppweb.org/ accessed on 20 September 2011. With permission from Bernhard Rupp.

2.8.3.1 Anomalous scattering

Anomalous scattering can be defined as absorption of photons by heavy atom when inner shell electrons undergo transition to a higher energy level or get ejected totally from the atom. When the excited electron returns to the inner shell a new x-ray photon is emitted but with a different amplitude and phase to the incident x-ray. This is what is called *anomalous scattering*.

At the wavelength where the photon has enough energy to throw the electron out of its orbital, there is a quick change in the absorption factor, commonly referred to as the atom's *absorption edge*. All elements exhibit anomalous scattering when the X-ray wavelength is near the atom's absorption edge. However, the X-ray wavelength used in protein crystallography is not near the absorption edge of the majority of atoms that compose the proteins such as oxygen, nitrogen and carbon and so these atoms do not contribute to the anomalous scattering. Sulphur does anomalous scatter but selenium does so more strongly. That is why selenomethionine is used instead of the normal sulphur-containing methionine. X-rays can be tuned at the synchrotron to the specific absorption energy of selenium to collect anomalous datasets. In order to incorporate the phase and amplitude of the anomalously scattered wave, the atomic scattering factors are expressed with real and imaginary parts as follows:

$$f(\lambda) = f_0 - \delta f'(\lambda) + if''(\lambda)$$

where f_o is the usual atomic scattering factor for wavelengths far from the absorption edge, $\delta f'$ is the real number correction of the amount by which the normal scattering is reduced, f'' is the imaginary component.

f_0-$\delta f'$ is often known as f' and is called as the dispersive component while f'' is called the anomalous component.

These values can be determined experimentally by an X-ray fluorescence scan of the protein crystal itself on the synchrotron before data collection. This allows us to select wavelengths at which anomalous scattering will be

optimal for structure determination. The datasets are collected at the peak, the inflection point and remote wavelengths, with the peak providing the highest value of f'', which is also the maximum intensity (Figure 2.8).

2.8.3.2 Friedel's law

Friedel's law states that a pair of reflections *h,k,l* and *-h,-k,-l* have the same structure amplitudes and that their phases have opposite sign (Blow, 2002). This holds true when anomalous scattering is not involved because the heavy atom structure factors $F_H(h,k,l)$ and $F_H(-h,-k,-l)$ have the same magnitudes. The anomalous component of the scattered wave from the heavy atom is always at 90^0 to the advancing non-anomalously scattered wave (Figure 2.9). The law breaks down because the atomic scattering factors have to take into consideration the anomalous scattering factor (Figure 2.10).

2.8.3.3 Phasing

During anomalous scattering, the difference in scattering intensity (I_{hkl} and $I_{-h,\ -k,\ -l}$) and amplitudes ($|F_{hkl}|$ and $|F_{-h,\ -k,\ -l}|$) is called the Bijvoet difference. The Bijvoet difference can be used to determine anomalous scattering due to the heavy atom alone in F_{PH}. The position of the heavy atoms in the unit cell can be determined by using a Patterson map. The peaks obtained from the Patterson maps can be used to calculate the amplitudes and phases of the atomic structure factors. From these phases and amplitudes, by constructing the Harker diagram, the phases of the original reflections can be deduced geometrically (Figure 2.11A). The anomalous scattering depends on the wavelength of the incident beam. Using a single wavelength, it is possible to deduce two different reflection phases. More phases can be obtained by taking measurements at multiple wavelengths. Each of these wavelengths will provide more vectors on the Harker diagram. When circles are drawn around the Harker diagram, they will intersect at a point that allows

deduction of the F_p phase (Figure 2.11B). Normally, two wavelengths are enough to determine the phases, but in most cases three datasets are collected at different wavelengths. The first two datasets are collected at the peak and inflection point wavelengths. At these points the contribution of f'' and f' to structure factor scattering is at its maximum and minimum respectively (Figure 2.11C). The third dataset is collected at the remote point where both of f'' and f' are very small. Having three datasets provide the best possible phasing power.

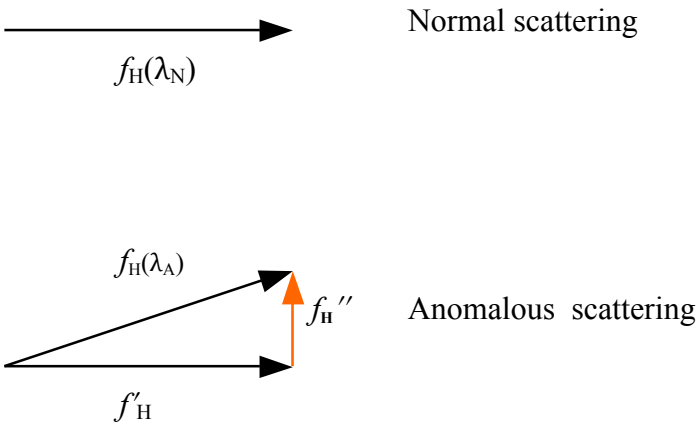


Figure 2.9: **A heavy atom scattering factor f_H at two different wavelengths**.
At wavelength λ_N it scatters normally. At wavelength λ_A there is an anomalous scattering component f_H'' which the phase is 90° in advance of the normal component, and also the normal component of the scattering is reduced to f'_H. Taken from Blow (2002).

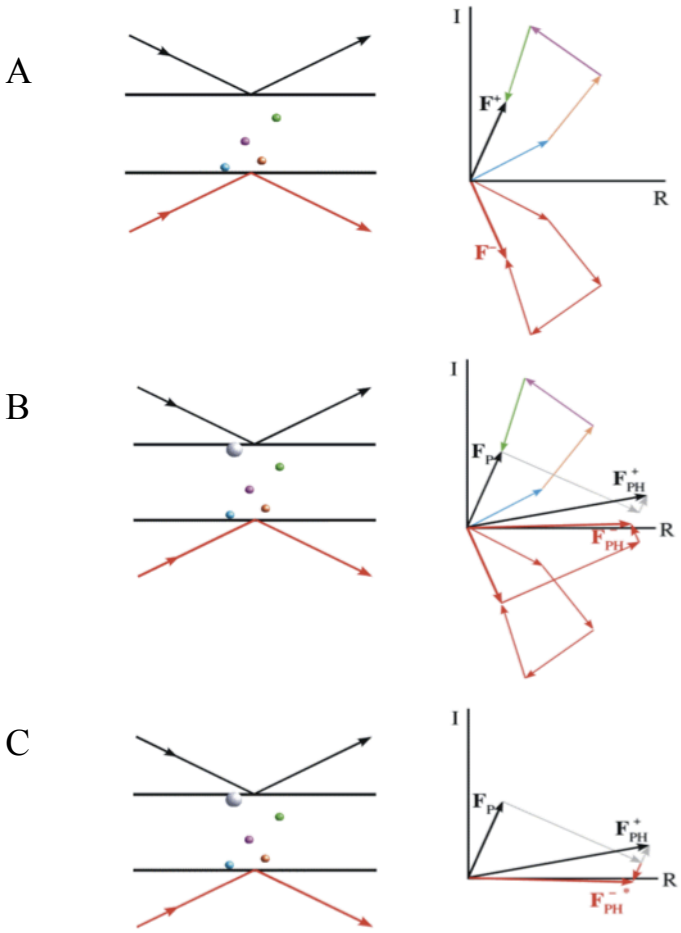


Figure 2.10: **Diagrammatic representation of Friedel's law and anomalous scattering.**
(A) is the normal scattering whereby the reflections along the Bragg plane have equal structure amplitude but opposite phases.
(B) is anomalous scattering due to the presence of heavy atom (grey). The anomalous scattering contributes to overall the structure factors (F_{PH}); hence the reflections are no longer in phase.
Similar in (C) the anomalous scattering has caused changes to the phases, with the reflections observed along the horizontal axis. Taken from http://www-structmed.cimr.cam.ac.uk/Course/Basic_phasing/Phasing.html accessed on 21st September 2011. With permission from Randy J. Read.

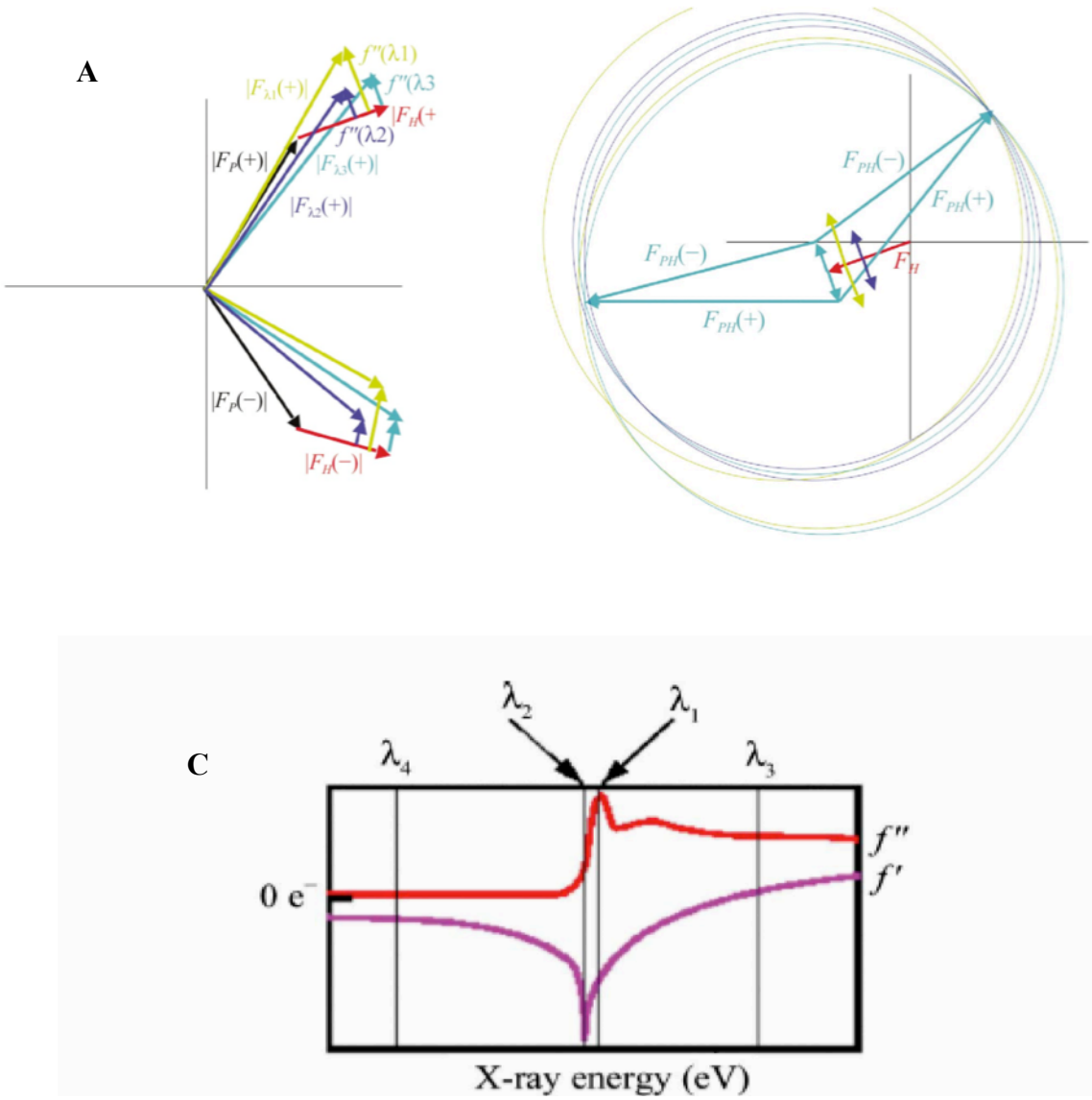


Figure 2.11: MAD phasing.
(A) Is the Argand diagram of three datasets collected at different wavelengths taken at inflection point λ1, peak λ2 and remote λ3. The scattering vector due to the protein is shown in black and is the same at all wavelengths. (B) Is the Harker construction of the three wavelengths from (C). (C) Is a typical absorption curve for an anomalous scatterer. The F_{PH} structure factor scattering is a combination of the protein crystal, dispersive and anomalous scattering from the remote point λ3, shown for clarity. Draw from the origin F_H, whose magnitude and phase was determined from Patterson functions then double circles are drawn at each wavelength of radius F_{PH} (+) and F_{PH} (-) centred at the end of their f'' contribution to the scattering.
The intersection of circle with one another can be seen at a single point which gives an estimate for the phase of F_p. Taken from Taylor (2003; (http://dx.doi.org/10.1107/S0907444903017815). With permission from IUCr.

2.8.4 Molecular replacement (MR)

Another method of determining phases is by using molecular replacement (MR). This method requires the existence of a previously solved protein structure that has some similarity to the unknown protein. MR can be done by using a program called PHASER (McCoy et al, 2007). MR requires first that the model protein to be rotated into the some orientation as the unknown protein in the unit cell. Generally, this rotation function is done by using Patterson maps of the known and unknown proteins. The correct positioning within the unit cell then requires a translational step.

The rotation function involves comparisons of the unknown Patterson map to the Patterson map of the known protein or phasing model in different orientations. The Patterson maps are overlapped and oriented in steps through three dimensions. At each orientation, the program calculates the structure factor amplitudes of the unknown protein and compares them with the known protein structure factor amplitudes until the maximum overlap is found. The maximum overlap or similar orientations are taken as rotation function solutions for the unknown protein.

The translational function is generally a positioning process of the correctly oriented protein in the asymmetric unit. The protein needs to be translated into the correct co-ordinates, x, y, z, axis to find the exact position where the protein lies in the unit cell. Normally this process is straightforward if there is single copy in the asymmetrical unit, but multiple copies of the target protein or flexible domains can take a very long time to process as each copy will need to be oriented. Otherwise, MR is the quickest method of determining phases than any other experimental methods

2.9 Calculation of electron density

When all is said and done and the phases have been determined to their very best, it's time to examine the image of the structure in the form of the electron density map. The relationship between the structure factors (F) for

reflection *hkl* and electron density (ρ) at the atomic position **r** is given by the following equation;

$$\mathbf{F}(hkl) = \int_{cell} \rho(\mathbf{r}) \exp(2\pi \mathbf{r} \bullet \mathbf{S}) dv$$

where **2πr.S** is the phase difference between the scattered wave at position **r** and the origin, and **S** is the diffraction vector of the refection indices *hkl*. Thus defining the position **r** with actual fractional cell coordinates x, y, and z, and substituting the **r.S** of the phase difference with (hx + ky + lz) for volume V of the unit cell, the above equation becomes

$$\mathbf{F}(hkl) = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(xyz) \exp[2\pi i(hx + ky + lz)] dx dy dz$$

F(hkl) is the structure factor with amplitude of |F(hkl)| and phase α(hkl). The relationship between the amplitude and the phase in three dimensions is given as **F**(hkl) = |F(hkl)| $e^{i\alpha(hkl)}$. Thus for every fractional coordinate, x, y and z in the unit cell, the electron density ρ (x,y,z) can be calculated from the structure factor **F** as follows:

$$\rho(xyz) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} \mathbf{F}(hkl) \exp[-2\pi i(hx + ky + lz)]$$

which is an inverse Fourier transform of the structure factor **F**(hkl) for the three-dimensional experimentally derived electron density map. The electron density is calculated in all combinations of *h*, *k*, and *l* as indicated by the summation sign.

2.10 Phase improvement

Regardless of which method is employed to determine phases, the initial electron density maps always contain some errors. This can make the first maps poor in quality and very difficult to interpret. This is the main reason

why programs have been developed to modify the electron density maps through improvement of the phases. These programs include solvent flattening, non-crystallographic symmetry (NCS) averaging and histogram matching (Rhodes, 2006; Rupp, 2010; Taylor 2003; and Taylor 2010).

2.10.1 Non-crystallographic symmetry (NCS) averaging

Non-Crystallographic Symmetry (NCS) arises from internal symmetry within a protein crystal that does not form part of the overall symmetry of the crystal. The NCS is characterised as local regions in the asymmetric unit with similar electron density. These local regions arise from multiple copies or subunits of a molecule. Thus, the asymmetric unit in NCS is made up of more than one identical molecule. These local regions of electron density can be used in NCS averaging to improve phases. The NCS operators can be obtained by a native Patterson function, heavy atom positions, real space correlation and self-rotation function, using both rotation matrices and translation vectors. The NCS averaging improves the signal to noise ratio and the overall quality of the electron density map. The more identical objects there are in the asymmetrical unit, the better is the improvement to the electron density map.

2.10.2 Solvent flattening

Solvent flattening is one of the most successful methods of improving the electron density map. In a protein crystal, more than 50% of its content can be made up of solvent. These solvent regions are highly disordered compared to the protein regions. Because of this lack of order in the solvent region, the electron density map shows fewer features. The solvent flattening procedure sets a mean electron density value for the solvent region while the protein regions are left unaltered. The resultant electron density distribution in the solvent areas is used to calculate structure factors, and the new phase angles can be used to improve the electron density of the map.

There is still uncertainly as to exactly which regions of the electron density map really represent the solvent so one needs to be very cautions not to assign protein regions as solvent areas; therefore a generous boundary must be placed around molecules. By flattening the solvent areas, the phases are improved and the protein regions density map is improved. This processing can be repeated a number of times to improve the maps.

2.10.3 Histogram matching

In histogram matching the electron density as observed (from lowest to highest recorded density) is used to construct the histogram, which is then compared to the ideal histogram of previously refined structures. This method assigns every electron density value within the asymmetric unit and compares it to the predicted distribution of electron densities when the protein is observed at this resolution. This takes into account the B-factors, the overall scale, the resolution and all other important parameters that can be expected to be observed at that resolution. Hence, deviations from the expected features on the map are attributed to poor phasing angles, which have to be adjusted to modify the electron density map.

2.11 Structure refinement

The refinement of the protein structure is necessary to improve the phases and the geometry of the model. This is achieved by modifying the molecule to improve the correlation between the observed ($|F_{obs}|$) and calculated ($|F_{calc}|$) structure factor amplitudes. During the initial stages of phase improvement, there are many details of the structure that are missing or assigned incorrectly. These include missing atoms, disordered regions, solvent molecules and poor coordinates and B-factors.

The key in structure refinement is to maximise the ratio of unique observations recorded to the number of the refined parameters. The higher is

the resolution, the more are parameters that will be needed to be refined, because there are many reflections at higher resolution.

2.11.1 Monitoring the progress of refinement

Refinement is a cyclic process and can go on forever. Therefore it is important to monitor the fitness of the refined model along the way so that one can know when to call it. Monitoring the refinement progress can help to detect some inherited or systematic errors in the model. The R factor measures the fitness of the model and it is defined as the ratio of the sum of differences between observed and calculated structure amplitudes to the sum of observed structure amplitudes. Every stage of the refinement should lead to a decrease in the R factor. Its values depend on the resolution, the order in the crystal, the diffraction data, B-factors etc. The R factor is expressed as:

$$R = \frac{\sum_{hkl} \left\| F_{obs} \right| - k \left| F_{calc} \right\|}{\sum_{hkl} \left| F_{calc} \right|}$$

The major problem with the R factor is that is does not take into consideration the biases in the refinement data. For example, over-fitting or addition of water molecules can reduce the R factor, while the overall model is not improved. The a second free R factor is introduced to provide unbiased structure refinement data therefore, the data is divide into two part, the working set which is used in refinement and is about 95% of the data and the free R set 5% of the data. These are used to calculate R_{work} and R_{free} respectively. The free R factor is calculated from the *test set* or *free R flag* that was set aside from the beginning i.e. these reflections were not used during the refinement process. This test set is a randomly chosen 5% intensities. They can be randomly generated using UNIQUE (CCP4, 1994) programme. The free R factor measures how the model predicts a subset of

measured intensities that were not included in the refinement. The free R factor is expressed as;

$$R_{free} = \frac{\sum_{hkl \subset T} \left\| F_{obs} \right| - k \left| F_{calc} \right\|}{\sum_{hkl \subset T} \left| F_{obs} \right|}$$

The $hkl \subset T$ is the reflections from the free R set. R factor and free R factor should decrease at approximately the same rate towards end of the refinement, although it is normal at the beginning of structure refinement for the R_{free} to be 2-5% higher than the R factor, depending on the resolution. It is common to find most crystallographers using both Rs during structure refinement progress. In the end of any refinement with an improved R factors, the model must be stereochemically and conformationally reasonable and should be validated before depositing it to the databanks.

2.11.2 Model validation

The working R factor and free R factor cross validate the model and show the fitness of the model. The model needs to be stereochemically and conformationally valid. It must not contain inverted amino acids i.e. D-amino acids. The model also needs to show that all peptide bonds are planar and nonproline peptides bonds are *trans,* unless a *cis* conformational is followed by proline due to a specific local conformation constrains. The backbone conformational dihedral angles phi (ϕ), psi (ψ) and omega (ω) should be within the allowed regions as determined from the Ramachandran plots in the case of ϕ and ψ. Normally these parameters are expected to improve during the refinement progress. As they are not restrained like bond angles and bond lengths during refinement, they are very good indicator of the model quality. The MOLPROBITY program (Chen et al, 2010) and PROCHECK (Laskowski et al, 1993) analyses were used to validate the protein structure in this research project.

References

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res. *Nucleic Acids Research* **28:** 235-242

Blow D (2002) *Outline of Crystallography for Biologists*, New York: Oxford University Press.

Blundell TL, Johnson LN (1976) *Protein Crystallography*, New York: Academic Press.

CCP4 (1994) The CCP4 suite- programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50:** 760–763

Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66:** 12-21

Drenth J (1999) *Principles of Protein X-ray Crystallography*, 2nd edn. Berlin: Springer-Verlag.

Hammond C (2009) *The Basics of Crystallography and Diffraction*, 3rd edn.: International Union of Crystallisation, Oxford University Press, UK. .

Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66:** 125-132

Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) Prochecka program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26:** 283-291

Leslie AGW (1994) MOSFLM. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26**

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Cryst* **40:** 658-674

Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A* **276:** 307-326

Rayment I (2002) Small-scale batch crystallization of proteins revisited: An underutilized way to grow large protein crystals. *Structure* **10:** 147-151

Rhodes G (2006) *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models* London: Academic Press.

Rupp B (2010) *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, New York: Garland Science.

Taylor G (2003) The phase problem. *Acta Crystallogr D Biol Crystallogr* **59:** 1881-1890

Taylor G (2010) Introduction to phasing. *Acta Crystallogr D Biol Crystallogr* **66:** 325–338

Chapter 3: Materials and Methods

This chapter will describe the experimental procedures used during the course of this project.

3.1 Strains, plasmids and antibodies

Bacterial strains, plasmids, and primary antibodies used throughout this study are listed in the (Table 3.1).

3.2 Chemical suppliers and equipments

The majority of chemicals and equipment used in this thesis were general products and are commonly available from a variety of companies. Chemicals were purchased from BDH Laboratory Supplies, Poole, England; Fisher Scientific UK Ltd., Leicestershire, UK; Bio-Rad Laboratories, Inc., CA, USA; Sigma-Aldrich, St. Louis, MO, USA; Invitrogen Corporation, CA, USA; Qiagen, Hilden, Germany; and GE Healthcare, Freiburg, Germany.

All lipids were purchased from Avanti Polar Lipids, Inc., Alabama, USA and detergents from Glycon Bioch. GmbH, Germany.

3.3 Microbiological Methods

3.3.1 Growth media and conditions: Luria-Bertani (LB)

Bacterial strains were routinely grown in sterilised Luria-Bertani (LB) medium (Sambrook & Russell, 2006) .

	LB medium	LB-agar
Tryptone	10 g /L	10 g /L
Yeast extract	5 g /L	5 g /L
NaCl	10 g /L	10 g /L
Bacteriological agar	-	15 g /L

All media were sterilised by autoclaving. The appropriate antibiotics were added to LB-agar after cooling to 40-45 °C and LB agar was poured into Petri dishes and allowed to set. Plates were kept at 4 °C and used within one month.

3.3.2 M9 Minimal medium

Bacterial strains were grown in M9 minimal medium when this medium was required (Larsson 2009).

	Stock concentration	Volume added
5x M9 salts*	-	200 ml
$MgSO_4$	1 M	2 ml
Glucose	20% (w/v)	20 ml
$CaCl_2$	1 M	0.1 ml
Sterile deionised H_2O	-	up to 980 ml

*5x M9 salts have been added to deionised H_2O to a final volume of 1 litre:

$Na_2HPO_4.7H_2O$	64 g
KH_2PO_4	15 g
NaCl	2.5 g
NH_4Cl	5.0 g

Table 3.1: The bacterial strains, plasmids and primary antibodies that used in this thesis.

Bacterial Strain, plasmid, or antibody	Source or reference
***E. coli* Bl21 (DE3)** (Novagen)	Lab collection
B. subtilis strain JH642	D. Phung and Prof. P. E. Granum (Our collaborators in Norwegian School of Veterinary Science, Laboratory Animal Unit, Oslo, Norway)
pGS2134 pET28a derivative with the APEC *clyA* coding region and kanamycin resistance . APEC *clyA* was inserted between EcoR1 and Sal1 restriction sites with a thrombin cleavage site.	Prof. J. Green (Department of MBB/University of Sheffield/United Kingdom)
PEXP5-CT/TOPO-NheA PEXP5-CT/TOPO derivative with the *nheA* coding region and ampcillin resistance. *nheA* was cloned between Topoisomerase site 1 and Topoisomerase site 2.	D. Phung and Prof. P. E. Granum (Our collaborators in Norwegian School of Veterinary Science, Laboratory Animal Unit, Oslo, Norway)
pDG148-NheB pDG148-StuVector derivative with the *nheB* coding region and kanamycin resistance.	D. Phung and P. E. Granum (Our collaborators in Norwegian School of Veterinary Science, Laboratory Animal Unit, Oslo, Norway)
MAb 1A8 Primary antibody reactive with NheA	Dr. Erwin Märtlbauer (Ludwig-Maximilians-Universitat, Munich, Germany), Our collaborators in Germany.
MAb 1E11 Primary antibody reactive with NheB	
MAb 2B11 Primary antibody reactive with NheB	
MAb 1C2 Primary antibody reactive with NheB	D. Phung and P. E. Granum (Our collaborators in Norwegian School of Veterinary Science, Laboratory Animal Unit, Oslo, Norway)

3.3.3 Storage of strains and plasmids

Strains were stored for up to one month on solid media at 4 °C. For long-term storage, strains were kept in the form of glycerol stocks in 80 % (v/v) sterile glycerol and stored at -80 °C. Plasmids were stored at -20 °C.

3.3.4 Transformation of chemically competent cells

The transformation is based on (Hanahan, 1983). Chemically competent bacteria were thawed on ice before plasmid DNA was added as recommended in the manufacturer's instructions. Briefly, a 50 µl aliquot of competent cells was incubated on ice with ~100 ng of DNA for 30 minutes after being mixed gently. The cells were then heat shocked for 60 seconds at 42 °C and a further 2 minutes on ice. 950 µl of LB medium was added and the transformation mixture incubated at 37 °C for 1 hour. The cells were pelleted on LB-agar plates, that containing the appropriate selective antibiotic. Plates were growth overnight at 37 °C.

3.3.5 Agarose gel electrophoresis

Agarose gel electrophoresis was used for confirming the presence of the APEC *clyA* insert gene. Agarose 1% (w/v) was added to 1x TAE and dissolved by heating in a microwave oven. Once cooled to 50 °C, ethidium bromide (0.6 µg/ml, final concentration) was added to the molten agarose and the gel was poured. Electrophoretic fractionation was carried out using electrophoresis tanks at 100 V for up to 2 hours. DNA inserts were visualised using a UV Itech photodocumentation system.

50x TAE buffer:

Trizma® base	242 g /L
Glacial acetic acid	57.1 g /L
$Na_2EDTA.2H_2O$, pH 8.5	37.2 g /L

3.3.6 Plasmid sequencing

Insert gene sequencing was done by DNA Sequencing and Bioinformatics GATC Biotech Ltd, Cambridge.

3.4 Protein Methods

3.4.1 Concentration of proteins

Where proteins of increased concentration were required for crystallization, the concentration of the protein was performed in a 10 kDa or 100 kDa molecular weight cut off Centricon centrifugal ultrafilteration device (Vivascience) as required. Centrifugation at 4000 *g* was carried out at 4 °C until the protein solution reached the desired concentration.

3.4.2 Measurement of protein concentration

Protein concentration was estimated with the Bio-Rad protein assay protocol Bradford method (Bradford, 1976) with reagents purchased from BioRad. Absorbance was measured at 595 nm using a spectrophotometer.

3.4.3 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used routinely throughout this thesis for analysis of proteins. SDS-PAGE analysis was performed according to the method described by (Laemmli, 1970) using the BioRad Mini Protean 3 system according to the manufacturer's instructions. Gel and buffer recipes used in this thesis were taken from 'Manual of Molecular Cloning (Sambrook & Russell, 2006). All SDS-PAGE gels reported in this thesis were 12% (acrylamide concentration).

12% SDS-PAGE separating gel:

Acrylamide/ Bis solution	5 ml 30% (w/v)
1 M Tris-HCl, pH 8.8	4.69 ml
10% SDS (w/v)	125 µl
Milli-Q water	2.56 ml
Ammonium persulphate	125 µl 10% (w/v)
N,*N*,*N*′,*N*′-tetramethyl-ethane-1,2-diamine (TEMED)	12.5 µl

6% SDS-PAGE stacking gel:

Acrylamide /Bis solution	1.5 ml 30% (w/v)
1 M Tris-HCl, pH 6.8	0.94 ml
Milli-Q water	2.6 ml
Ammonium persulphate	75 µl 10% (w/v)
10% SDS (w/v)	75 µl
TEME	7.5 µl

1x SDS-running buffer:

Glycine	14.4 g /L
Tris-HCL	3 g /L
SDS	1 g /L
*Adjust to pH 8.8	

2x SDS-loading buffer:

Glycerol	20% (v/v)
Tris-HCl	100 mM
SDS	4% (v/v)
Bromophenol blue	0.02%(v/v)
β-mercaptoethanol	200 mM

1x Coomassie blue stain:

Methanol	20% (v/v)
Acetic acid	7.5% (v/v)
Coomassie Blue	0.1% (w/v)

Destain:

Methanol	20% (v/v)
Acetic acid	7.5% (v/v)

3.4.4 Western immunoblotting

3.4.4.1 Chemiluminescence method

Western immunoblotting was carried out to identify and locate Nhe proteins (NheA and NheB). 12% SDS-PAGE analysis for Nhe was removed to a nitrocellulose membrane at 80 V for 1 hour. Then, the membrane was blocked with 5% non-fat dried milk in TBST buffer for 16 hours at 4˚C with shaking.

The target protein was revealed by primary antibody at 4˚C for 1 hour with shaking. After five washes of the membrane by TBST buffer, Secondary antibody anti mouse HRP was added. Bound antibodies were detected by Chemiluminescence using LumiGLO reagent according to company's procedure.

3.4.4.2 Colorimetric kit method

Western immunoblotting was carried out to check for the presence of NheB. 12% SDS-PAGE analysis of NheB was transferred to a nitrocellulose membrane at 80 V for 1 hour. Then, the membrane was blocked with 5% non-fat dried milk in TBST buffer for 16 hours at 4 ˚C with shaking.

NheB was revealed by primary antibody 1C2 at 4 ˚C for 1hour with shaking. After five washes of the membrane by TBST buffer, Secondary antibody biotin-conjugated anti-mouse antibodies from goat (Amersham Biosciences) was added. The membrane was washed five times with TBST buffer then transferred into mixture solution of streptavidin (Bio-Rad) and biotinylated alakaline phosphatase (Bio-Rad). The membrane was developed using colorimetric kits AP Color Development Reagent [5-bromo-4-chloro-3-indolyl phosphate (BCIP) and nitro blue tetrazolium (NBT)] from Bio-Rad according to company's procedure.

1x TBST buffer

1 M Tris HCL pH 7.4	20 ml
NaCl	8 g /L
Tween20	1 ml

3.4.5 Protein expression

In general for the expression of APEC ClyA, NheA and NheB (more details of each protein in chapters 4, 5, and 7), aerobic growths of bacteria strains were typically cultured in 2L flasks with appropriate antibiotics and shaking at 250 or 200 rpm as required. All flasks and LB media were capped with foam bungs and aluminium foil then autoclaved before use. Bacterial growth was measured by reading the optical density (OD) at 600 nm of 1 ml culture samples. IPTG was added (as required) at point 0.6 to 0.8 then the culture was incubated at 37 °C for a further hours.

In case of APEC ClyA, NheA, the cultures of bacteria were harvested at 4˚C (15,344 g, 20 minutes) and lysed by sonication. The sonication was performed in 3 cycles for 29 seconds at 16-micron amplitude, with cooling of the samples between sonication cycles. The lysed bacteria were centrifuged at 70000 g for 10 minutes.

The protocol of (Larsson 2009) it was used for expression of Selenomethionine (SeMet) labelled NheA protein in *E. coli* . The amino acids inhibition methionine biosynthesis, L-Isoleucine (50 mg /L), L-Leucine (50 mg /L), L-Lysine (100 mg /L), L–Phenylalanine (100 mg /L), L –Threonine (100 mg /L) and L–Valine (50 mg /L), were used together with L-SeMet (60 mg /L) then IPTG (200 mg /L) was added to induce SeMet-NheA in minimal media.

In case of NheB, it was assumed that the protein would be released from the cells into the media, and the NheB was precipitated from the media using 60% $(NH_4)_2SO_4$.

3.4.6 Protein purification

As we are working on three proteins APEC ClyA, NheA and NheB, several different types of column chromatography were used throughout this project including ion exchange, hydrophobic interaction, gel filtration, and affinity chromatography:

3.4.6.1 Ion exchange chromatography

In ion exchange chromatography, the proteins separate on the basis of ionic attraction between molecules of an opposite charge. In this project it was used several types of ion exchange chromatography were used (as required): HiPrep ^{TM} 16/10 DEAE FF (anion exchange column DEAE-Sepharose Fast Flow weak anion exchanger) from GE Helthcare, Resource Q (HPLC) (strong anion exchanger) from GE Healthcare, and hydroxylapatite (Bio-rad Laboratories, Bio-Gel HT gel).

Washing the column with an increasing the gradient of salt can elute the protein. So the weakly bound proteins will be eluted first, followed by strongly bound components.

3.4.6.2 Hydrophobic interaction chromatography

In this kind of chromatography, the proteins separate based on hydrophobic groups on their surface, like ethyl, butyl or phenyl group.

In the lab, Phenyl-Toyopearl 630S (hydrophobic matrix) from GE Healthcare (formerly Amersham) was used. Ammonium sulphate is used to elute hydrophobic binding.

3.4.6.3 Gel filtration chromatography

Gel filtration chromatography is also known as size exclusion chromatography or molecular sieve chromatography. Gel filtration separates proteins by size and shape. Beads with different sizes of pores are available, allowing different size of molecules to be effectively separated.

In the lab Hi-Load Superdex 200 1.6x60 cm columns GE Healthcare (formerly Amersham) it was used.

3.4.6.4 Affinity chromatography: Nickel Nitrilotriacetic acid (Ni-NTA) Sepharose

In this kind of chromatography exploits a protein binding to another molecule. In Ni-NTA, which was used in this thesis, histidines bind to the nickel sites, allowing separation based on the surface concentration of this amino acid. The bound protein is released from immobilized ligand by increasing the gradient of Imidazole.

3.4.7 N-terminal sequencing of the proteins

N-terminal sequences were done to confirm the amino acid sequences of our proteins that we were working on. N-terminal sequences were determined by Dr. Arthur Moir (Krebs Sequencing and Synthesis Facility, Department of Molecular Biology and Biotechnology, The University of Sheffield, UK) using an applied biosystems protein sequencer.

3. 4. 8 Liquid chromatography–mass spectrometry (LC-MS)

LC-MS was used to measure the level of incorporation of selenomethionine (SeMet) in NheA. LC-MS of NheA and SeMet NheA were done by Mr.

Simon Thorpe (Department of Chemistry, The University of Sheffield, Dainton Building, Brookhill, Sheffield, UK).

3.5 Three dimensional crystallization trials

Initial 3D crystallization trials for the proteins (APEC ClyA, NheA, and NheB) were performed either in an automated procedure using the Matrix Hydra II Plus One crystallization robot (Figure 3.1) or manually via vapour diffusion methods (hanging drop and sitting drop) using different crystallization screens from Hampton Research (Crystal Screen 1, Crystal Screen 2, PEG/Ion, and PEG/Ion 2); Jena Bioscience (JBScreen Membrane HTS, JBScreen Classic 1, JBScreen Classic 6), and QIAGEN screens (JCSG+ Suite, PACT Suite, $(NH_3)_2SO_4$ Suite, Classics Suite, pH Clear) were utilized at 7 ºC and 17 ºC.

Optimization of promising conditions was performed in 24-well Linbro plates (Hampton Research) using varying concentrations of proteins, precipitant concentration, temperature, pH, or by adding additives.

3. 6 Cryoprotection of crystals

Crystals of the proteins (ClyA, NheA, and NheB) that were suitable for diffraction were picked and soaked in a cryoprotectant solution, before freezing in a nitrogen stream at 100 K. Several different cryoprotectants were initially screened: PEG-400, ethylene glycol, sucrose and glycerol. Diffraction of cryoprotectant solution was assessed on our in-house source at 100 K using the Rigaku Micromax 007 copper rotating anode generator,

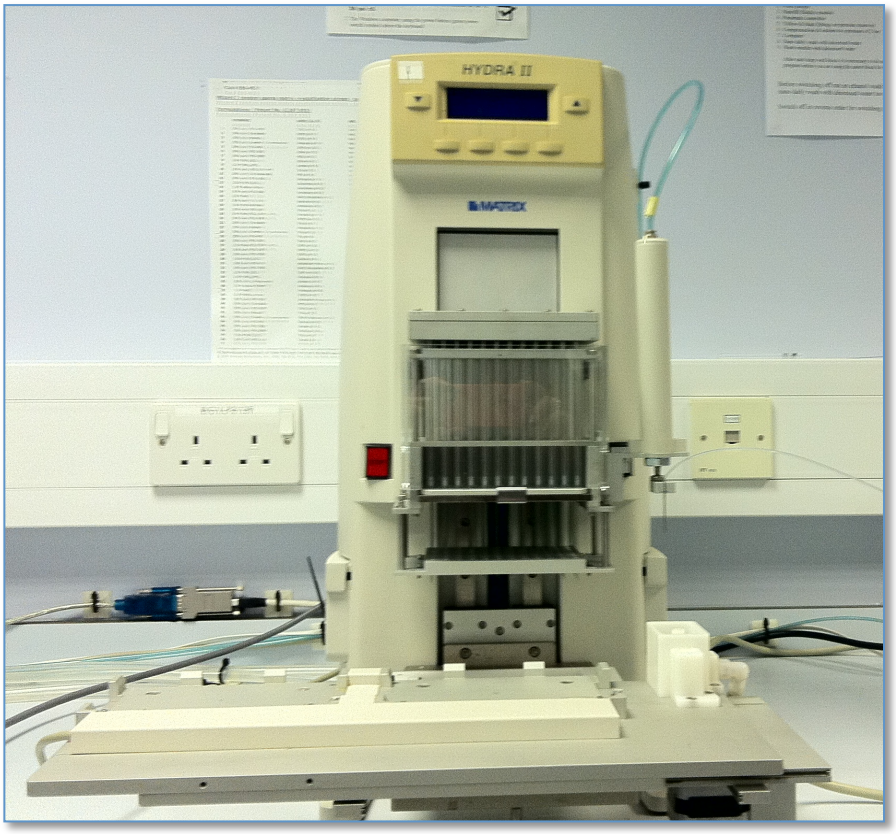


Figure 3.1: The Matrix Hydra II Plus One crystallization robot used in our lab.

which produces X-rays at a wavelength of 1.542 Å to identify the most suitable cryoprotectant.

3. 7 Data collection and processing

For producing high quality data collection, synchrotron X-ray beams at the Diamond Light Source in Oxford, United Kingdom were used to collect the data of the protein native crystals. Diffraction data were indexed and integrated using the programme IMOSFLM (Leslie, 1994) in home or XDS (Kabsch, 2010) at the Diamond Light Source in Oxford, United Kingdom. Scaled using SCALA (Evans, 1997) in CCP4 (CCP4, 1994). This gives an *.mtz file containing indices representing intensities (I) and their associated statistical errors σ (I), and final data statistics.

Fluorescence scans of SeMet NheA crystals were collected and anomalous data (MAD data) collected using synchrotron X-ray beams at the Diamond Light Source in Oxford, United.

Data were collected for anomalous scattering at three X-ray wavelengths of high remote 0.9784 Å, peak 0.9794 Å, and inflections 0.9797 Å. These wavelengths were selected because they determined by a fluorescence scan at wavelengths around the theoretical Selenium absorption edge. Anomalous data were processed using the same program as native data (see above).

3. 8 Structure Determination for NheA

In this thesis, we able to solve one structure of the protein that we are working on, which is NheA.

The NheA structure was solved by Multi-wavelength Anomalous Dispersion (MAD), more details in chapter 6. To carry out phase improvement we used

a combination of the SHELX program to locate heavy atoms (Sheldrick, 2010) and PHENIX to carry out phase improvement (Adams et al, 2010). Water molecules and un-modelled regions of electron density were interpreted using the program COOT (Emsley & Cowtan, 2004). Refinement of the model against the electron density map was performed using the program REFMAC (Murshudov et al, 1997). Between rounds of refinement the model was rebuilt using the program COOT (Emsley & Cowtan, 2004). More details and any other programs used for data analysis or structural validation are described in chapter 6 and chapter 7.

3.9 Biochemical protocol

3.9.1 Hemolysis Assay

The protocol of (Rowe & Welch, 1994) was used to measure the hemolytic activity of avian ClyA. Defibrinated horse and sheep blood was obtained from TCS Microbiology (UK). Briefly, purified protein was incubated with defibrinated blood in saline Tris- HCL pH 7 (50 mM Tris HCl, pH 7.0. containing 150 mM NaCl) for different tome (15 minutes, 30 minutes and overnight) at 37 ºC. The mixture was centrifuged at 1000 g for 1 minute. The amount of hemolysis was determined spectrophotometrically by measuring hemoglobin release at 543 nm in the supernatants and readings were repeated at least three times.

3.10 Electron microscopy studies

Pore formation was examined using a Philips CM100 transmission electron microscope. Micrographs were recorded using a Gatan MultiScan 794 charge-coupled device camera.

3.10. 1 Negative staining protocols

The protocol of negative staining (Ohi et al, 2004) was used to stain the protein samples. Briefly, 3 to 5 µl of sample was loaded on glow discharged carbon-coated EM grids for 1 minute. This was then washed with two 50 µl drops of deionized water and subsequently stained with two drops of 0.75% of Uranyl Formate. 0.75% (w/v) Uranyl formate solution was prepared in boiling dH_2O, then filter sterilised, and stored at 4 ^oC, at pH 7.0.

3.10.2 Immunogold labelling

3-8 µl of sample from each trial was loaded onto microscope grids for 1 minute. Excess sample was removed by blotting on filter paper. Grids were placed face down onto 50 µl drops of BSA (Bovine Serum Albumin) buffer. Grids were placed sample side down onto primary antibody drops and then were removed and washed three times in 50 µl drops of BSA buffer. The secondary antibody (Anti-Mouse IgG–Gold antibody produced in goat from SIGMA ALDRICH) was diluted 1/25 in BSA buffer and grids were again added sample side down onto drops. The grids were washed three further times in BSA buffer, and prepared for negative staining as normal.

BSA buffer

Tris HCL pH 7.5	25 mM
NaCl	0.5 M
Ca_2Cl	2.5 mM
Na_3N	0.005%
BSA	5%

3.10.3 Cross-linking experiment

The protocol was modified from (Huysmans et al, 2012). 3-8 µl of sample from each trial was loaded onto microscope grids for 1 minutes then cross-linking was performed using 50 mM glutaraldehyde for 30 min at 25 °C and then immunogold labelling was done as described in section (3.10.2).

3.10.4 Lipid prepration

Method 1: Brain total lipid extract (BL) and *E. coli* total lipid extract (EL) were used which were prepared from lipid film (Avanti Polar lipids). In this method it detergent was used to solublize the lipid. The glass vials were washed with chloroform/methanol in ratio 1:1 (v/v), and then with pure chloroform. Lipid was added to glass vials and the organic solvent was evaporated at room temperature. Finally, the lipid was solubilized by buffer 100 mM NaCl, 2.5 mM $CaCl_2$, 0.005 % Na_3N, in the presence of detergent, either ß-OG or DDM.

Method 2: Brain Total Lipid Extract from Avanti Polar Lipids (Stock solution of lipid in chloroform (10 mg/ml) was used. 2 ml of chloroform lipid solution was pipetted into a clean glass vial. Then the organic solvent was evaporated and the lipid was dried on to the glass walls using a nitrogen stream at room temperature. The lipid was re-solubilised in 5ml of buffer (25 mM Tris HCL pH 7.5; 0.5 M NaCl; 2.5 mM Ca2Cl and 0.005% Na3N); thus the final solution is 4mg/ml lipid in buffer. And then lipid extrusion by suspension strongly through pore (we used filter unit 20 µm) with syringe about 20 times to encourage liposomes to form.

3.10.5 SEM & TEM for *B. cereus* strain NHV 0075/95

Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) were used to carry out the first experiments to observe *B. cereus* strain NHV 0075/95.

The specimen was prepared according to techniques of (Bozzola & Russell 1999) by Dr. Christopher J Hill using the Philips XL-20 SEM and the FEI Tecnai G2 Spirit TEM in the Department of Biomedical Science, University of Sheffield, UK. Electron micrographs were taken using a Gatan digital camera.

References

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66:** 213-221

Bozzola J, Russell LD (1999) *Electron microscopy*, 2nd edn.: Jones & Bartlett Publisher.

Bradford MM (1976) Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72:** 248–254

CCP4 (1994) The CCP4 suite- programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50:** 760–763

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta crystallographica* **60:** 2126-2132

Evans PR (1997) SCALA. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **33:** 22-24

Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166:** 557-580

Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66:** 125-132

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227:** 680-685

Larsson AM (2009) Preparation and crystallization of Selenomethionine protein. In *IUL Biotechnology Series: protein crystallization*, Bergfors T (ed), 2nd edn. La Jolla California: International University Line

Leslie AGW (1994) MOSFLM. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26**

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53:** 240-255

Ohi M, Li Y, Cheng Y, Walz T (2004) Negative staining and image classification - powerful tools in modern electron microscopy. *Mol Microbiol* **20:** 191-199

Rowe GE, Welch RA (1994) Assays of hemolytic toxins. *Bacterial Pathogenesis, Pt A* **235:** 657-667

Sambrook JW, Russell DW (2006) *The condensed protocols from molecular cloning: a Laboratory Manual*, New York, USA: Cold Spring Harbor Press.

Sheldrick GM (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66:** 479-485

Chapter 4: APEC ClyA Expression, Purification, and Crystallization

This chapter will describe the work on Avian Pathogenic *E. coli* (APEC) ClyA that was carried out in a project in my first year and half of my second year. ClyA was introduced in chapter 1, section 1.4. Attempts were made to crystallize the protein both in its soluble and its pore form.

4.1 Plasmid sequence and agarose gel electrophoresis

The APEC ClyA insert was confirmed firstly by restriction analysis using EcoR1 and Sal1 restriction enzymes. Gel electrophoresis showed the presence of His Tag-APEC ClyA fragment on Figure 4.1 with 950 bp, which was the expected size of the His Tag-APEC ClyA insert.

The insert was then analyzed using the Basic local alignment search tool (BLAST) (Altschul et al, 1990) at the National Centre for Biotechnology Information http://blast.ncbi.nlm.nih.gov/Blast.cgi. It was clear the APEC ClyA insert sequence was 98% identical with the sequence of APEC ClyA from avian *E.coli* (Figure 4.2) which had been deposited in nucleotide database with accession number: AF052225. The discrepancies were due to undetermined nucleotides and to an erroneous G-R mistranslation in the original deposition as pointed out by Wyborn et al, (2004).

4.2 Expression of recombinant His Tag- APEC ClyA

E. coli strain BL21 (λDE3) was used as the host for His tagged-APEC ClyA expression vector, pGS2134.

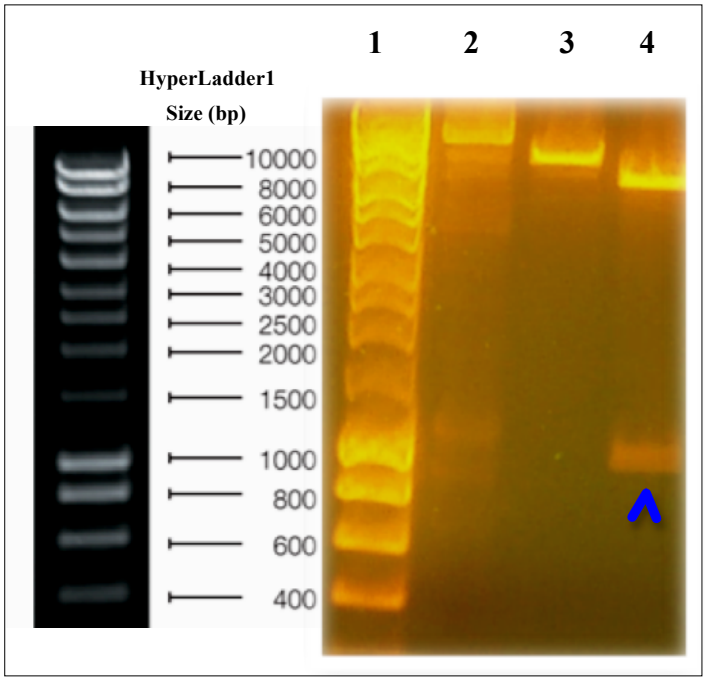


Figure 4.1: Single and double digests of APEC ClyA expression vector. Visualised by agarose gel (1%) electrophoresis. Lane 1: hyperladder I; Lane 2: non-digested plasmid; Lane 3: single digest with EcoR1; Lane 4: double digest with EcoR1 and Sal1. Arrowhead shows the expected size of His Tag-APEC ClyA insert, which about is 950 bp.

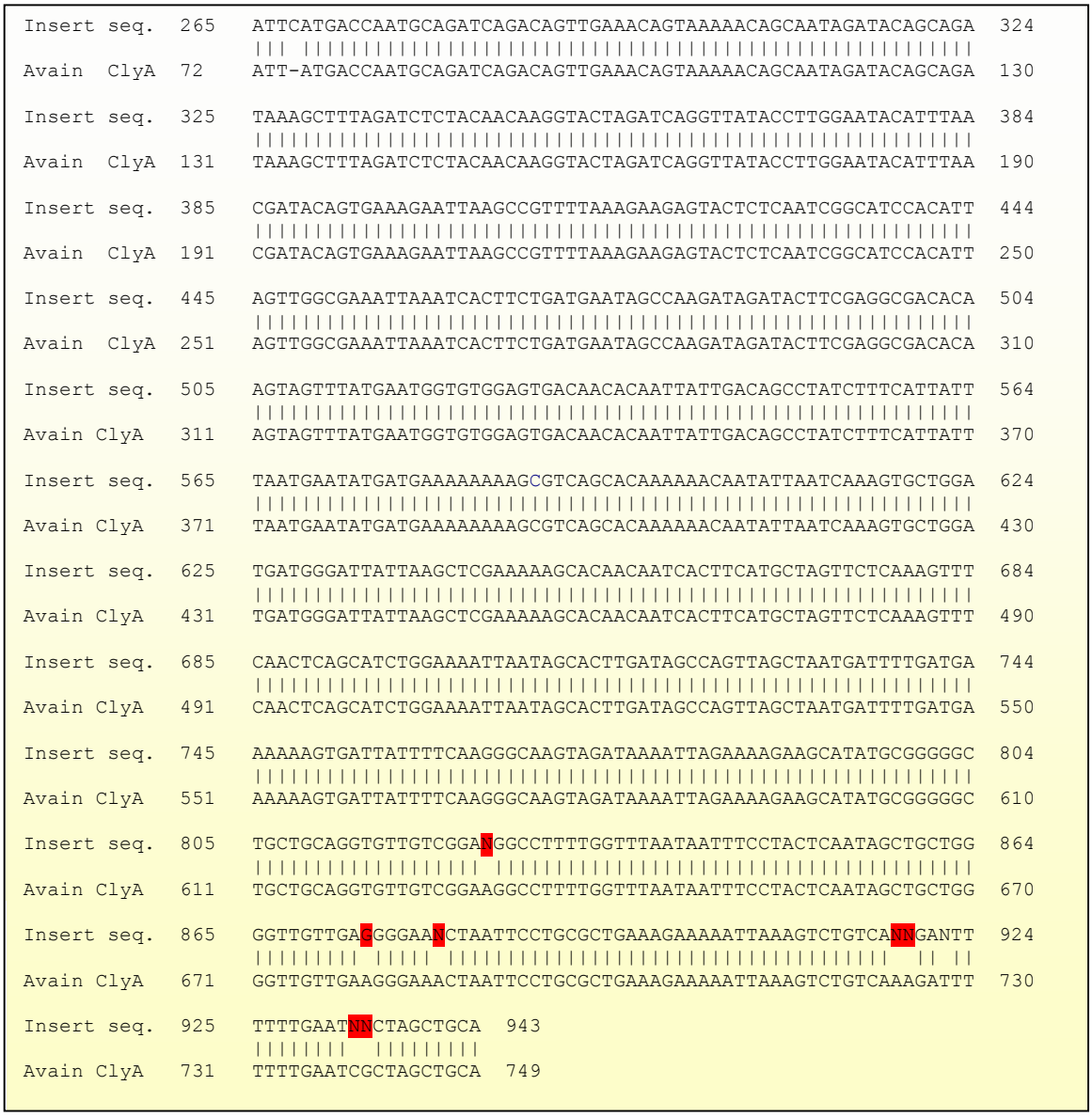


Figure 4.2: The APEC ClyA insert was sequenced and then analysed using BLAST, (Altschul et al, 1990) http://blast.ncbi.nlm.nih.gov/Blast.cgi. BLAST results showed that the APEC ClyA sequence is 98% identical to avian *E. coli* ClyA with accession number AF052225.

Expression of His Tag-APEC ClyA was performed aerobically in *E. coli* strain BL21 (λDE3) in LB supplemented with kanamycin (100 µg/mL) with shaking (250 rpm) at 37 ºC. In order to optimize the concentration of IPTG sufficient to induce the His Tag-ClyA, the following concentrations: 100 µg/ml, 500 µg/ml, and 1000 µg/ml of IPTG were used when the bacterial growth reached an OD_{600} of 0.7.

As a result, it can be observed from Figure 4.3 that all concentrations of IPTG were useful for expression of the protein after 5 hours. The molecular weight of expressed protein is the same as expected, which is about 34-35 kDa. Thus, we decided to use 100 µg/ml IPTG for expression of the recombinant protein after 5 hours at 37 ºC in the future.

Atkins et al, (2000) reported that 3 hours after induction by 100 µg/ml IPTG was enough for *E.coli* K12 strain JM109, which was used as the host for GST-ClyA, and we suppose the hosting cells may be responsible for slower expression of the protein.

4.3 Purification of His Tag- APEC ClyA

Cells were harvested by centrifugation for 20 minutes at 4 ºC at 15,344 g. The cell pellet was re-suspended in the lysis buffer (30 mM HEPES at pH 7.4, 300 mM NaCl, 20 mM imidazole at pH 7.4) followed by disruption on ice with a sonicator.

After sonication, the supernatant was loaded on a 5 ml Nickel-Nitrotriacetic acid (Ni-NTA) sepharose affinity column chromatography at room temperature and equilibrated with lysis buffer. The column was washed with lysis buffer to remove any unbound protein. The recombinant protein was then eluted from the column with elution buffer containing 30 mM HEPES, pH at 7.4, 300 mM NaCl, 500 mM imidazole at pH 7.4 (Figure 4.4).

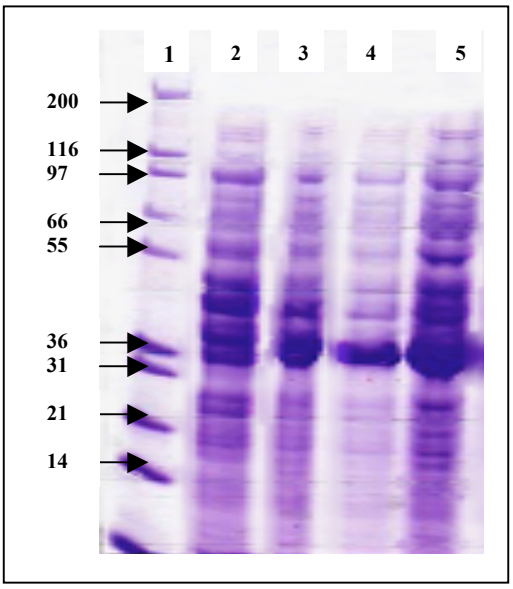


Figure 4.3: 12% SDS-PAGE analysis of *E. coli* strain BL21 (λDE3), induced for 5 hours at 37 °C, carrying plasmids for the production of His tagged APEC ClyA protein.
Visualised by staining with Coomassie blue. Lane1: Marker 12 from top to bottom: 200, 116, 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: supernatant (control); Lanes 3, 4, 5: supernatant induction by IPTG (100 µg/ml, 500 µg/ml, and 1000 µg/ml) respectively.

The fusion protein was then digested with thrombin at 25 °C for 16 hours to remove the tag. In order to determine how many units of thrombin are enough to release the APEC ClyA from its His-tag, a sequence of concentrations of thrombin was used (1 unit/mg, 3 units/mg, 5 units/mg, 7 units/mg and 10 units/mg). We found one unit of thrombin per one mg of the protein was sufficient to release APEC ClyA from its His-tag (Figure 4.5). It was supposed that thrombin recognises the consensus sequence L-V-P-R-G-S in the tag (Figure 4.6), cleaving the peptide bond between R and G. However, N-terminal sequencing analysis of APEC ClyA digested with thrombin protease revealed that thrombin also cleaved at site M-G-R-G-S. As a consequence of this cleavage site, APEC ClyA has an additional four residues (GSEF) fused to the N-terminus of the protein (Figure 4.6) rather than 19 as originally expected.

Cleaved APEC ClyA was applied to a calibrated superdex200 gel filtration column pre-equilibrated with 50 mM Tris HCl, at pH8, 0.5 M NaCl (Figure 4.7). Purified APEC ClyA had a band molecular mass of 33-34 kDa and was approximated to be ≥95% pure and homogeneous by SDS-PAGE (Figure 4.8) and LC-MS mass spectrum analysis (Figure 4.9). Data from LC-MS show the molecule weight of the protein is 35405 Da. This MW is in agreement with the expected MW from cleaving APEC ClyA at the first position (marked as red underline in the APEC ClyA sequence in Figure 4.6).

N-terminal sequencing of the sample confirmed it was APEC ClyA. Protein concentration was estimated with the Bio-Rad protein assay protocol (Bradford method) and concentration of the protein was performed in a 10 kDa molecular-weight cutoff of the Centricon centrifugal ultrafilteration device. APEC ClyA was concentrated to different protein concentrations (5 mg/ml, 7 mg/ml, 10 mg/ml and 18 mg/ml) before crystallization trials.

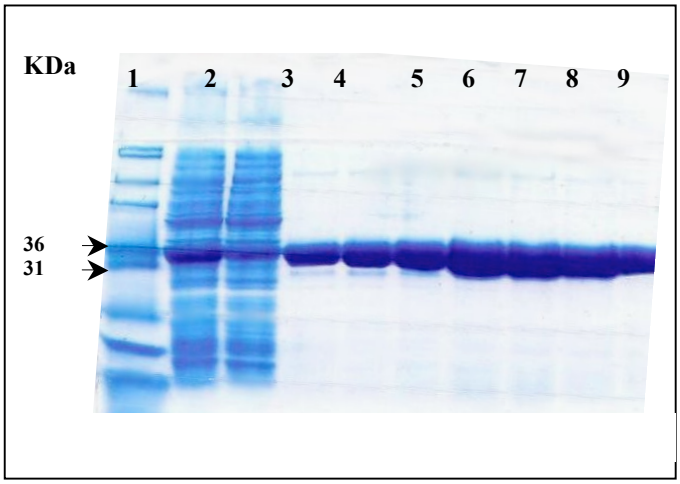


Figure 4. 4: **12% SDS-PAGE analysis of Ni-NTA purified APEC ClyA**. Visualised by staining with Coomassie blue stain. Lane 1: Marker12 from top to bottom: 200, 116, 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: cell extract after induction (100 µg/ml of IPTG); Lane 3: unbound proteins; Lanes 4, 5,6,7,8 and 9 the fractions collected from the Ni-NTA column.

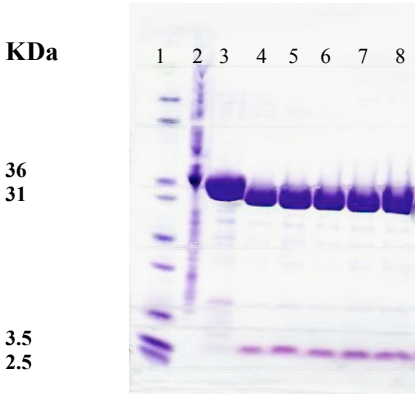


Figure 4.5: 12% SDS-PAGE analysis for APEC ClyA protein cleaved using different concentrations of thrombin.
Visualised by staining with Coomassie blue stain. Lane1: Marker12 from top to bottom: 66, 55, 36, 31, 21, 14, 10, 3.5, and 2.5 kDa; Lane 2: cell extract after induction (100 µg/ml of IPTG); Lane 3: loading sample Ni-NTA; Lanes 4, 5, 6, 7, and 8 a series units of thrombin (1 unit/mg, 3 units/mg, 5 units/mg, 7 units/mg and 10 units/mg), respectively.

```
        10         20         30         40         50         60
FCLTLRRRYT MGSSHHHHHH SSGLVPRGSH MASMTGGQQM GRGSEFMTNA DQTVETVKTA

        70         80         90        100        110        120
IDTADKALDL YNKVLDQVIP WNTFNDTVKE LSRFKEEYSQ SASTLVGEIK SLLMNSQDRY

       130        140        150        160        170        180
FEATQVVYEW CGVTTQLLTA YLSLFNEYDE KKASAQKTIL IKVLDDGIIK LEKAQQSLHA

       190        200        210        220        230        240
SSQSFNSASG KLIALDSQLA NDFDEKSDYF QGQVDKIRKE AYAGAAAGVV GGPFGLIISY

       250        260        270        280        290        300
SIAAGVVEGK LIPALKEKLK SVKDFFESLA ATVKSANTDI DKAKSKLKDE ISVIGDLKTE

       310        320        330        340
TETTRFFVDY DDLMLKQLQD SATKLILSCN EYQKRHGKKN EQPLWY
```

Figure 4.6: **Primary sequence of insert APEC ClyA** (accession ID: AF052225_1) blue colour, His-tag (green colour), Cleaving site (red colour). The thrombin could be has two cleaving sites as confirmed by N-terminal sequence; shown in red colour (the first position of cleavage site marked by red underline).

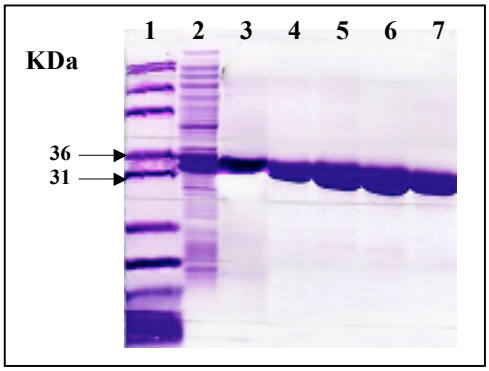


Figure 4.7: 12% SDS-PAGE analysis of APEC ClyA, released from His tag by thrombin cleavage after gel filtration.
Visualised by staining with Coomassie blue stain. Lane1: Marker12 from top to bottom: 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: cell extract after induction (100 µg/ml of IPTG); Lane 3: loading sample an Ni-NTA column; Lanes 4, 5, 6, and 7 the fractions collected from gel filtration after thrombin cleavage.

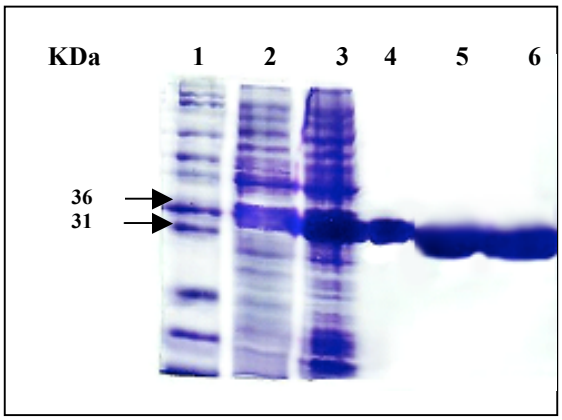


Figure 4.8: 12% SDS-PAGE analysis of the purification of APEC ClyA protein. Visualised by staining with Coomassie blue stain. Lane1: Marker12 from top to bottom: 200, 116, 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: cell extract before induction; Lane 3: cell extract after induction (100 µg/ml of IPTG); Lane 4: loading sample on Ni-NTA; Lane 5: APEC ClyA cleaved by 1 unit/mg of thrombin; Lane 6: loading sample on gel filtration.

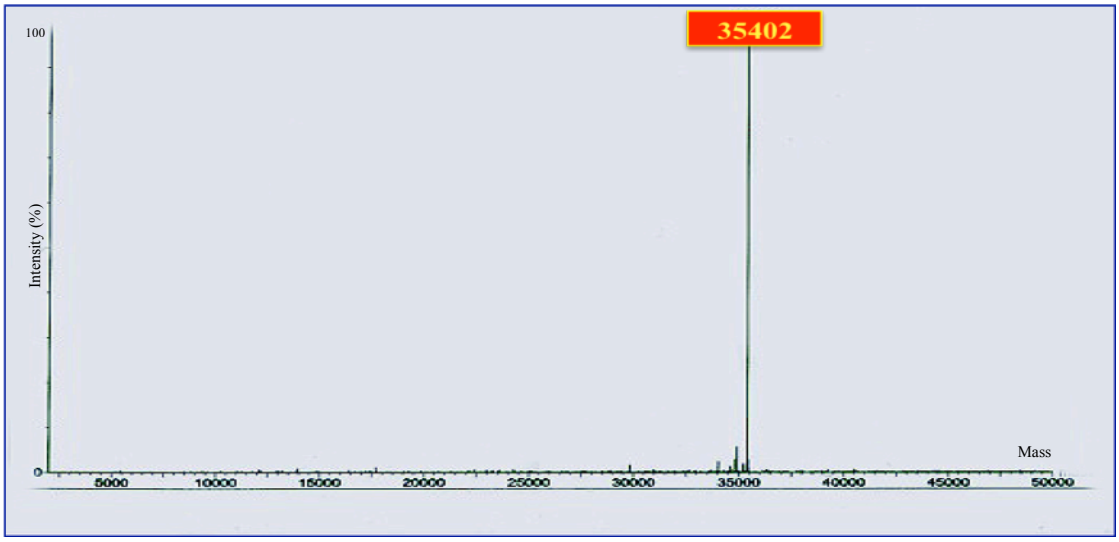


Chart 4.9: LC- MS peptide analysis of APEC ClyA (20 µg).
Highlighted peaks indicate mass weight of APEC ClyA.

4.4 3D Crystallization trials of cleaved monomeric APEC ClyA

More than 5000 crystallization trials for monomeric Cleaved APEC ClyA were preformed manually or by automated setup via vapour diffusion methods (hanging drop and sitting drop, respectively). Different crystallization screens were utilized at 7 ºC and 17 ºC with different protein concentrations (5 mg/ml, 7 mg/ml, 10 mg/ml and 18 mg/ml).

No usable crystals were observed from 3D crystallization trials of monomeric APEC ClyA. However, microcrystals were seen in some conditions: precipitated discs with tiny needles projecting from them; a very heavy circular drop with some small rectangular crystalline material as well growth of small crystals in 0.1 M Succinic acid at pH 7; and fused plates-like crystals in 0.1 M Bicine at pH 9, 0.3 M $MgCl_2$, 25% PEG2000, 15% Glycerol. There is a summary of these results in Figure 4.10. Attempts were made to optimize these conditions to obtain large crystals but no crystals formed.

Chemical modification of APEC ClyA by lysine methylation was attempted using an alkylating agent, formaldehyde, and the reducing agent, dimethylamine borane complex, as described in a previous study (Walter et al, 2006). Needle microcrystals were obtained (Figure 4.11) in 70% MPD and 0.1 M HEPES pH 7.5. Methylation is expected to add +14 Da to the molecular mass per methyl group added. The difficulty here was that mass spectrum analysis of methylated APEC ClyA (Figure 4.12) shows several different states of lysine methylation and some evidence of lower molecular weight species than the 35402 Da of unmodified APEC ClyA. Accordingly, the solution becomes heterogeneous, which may have an effect on crystallization trials, since one of the most significant factors in crystallization is homogeneity of the sample (Qiu & Janson, 2009). Microcrystals were obtained but could not be optimized.

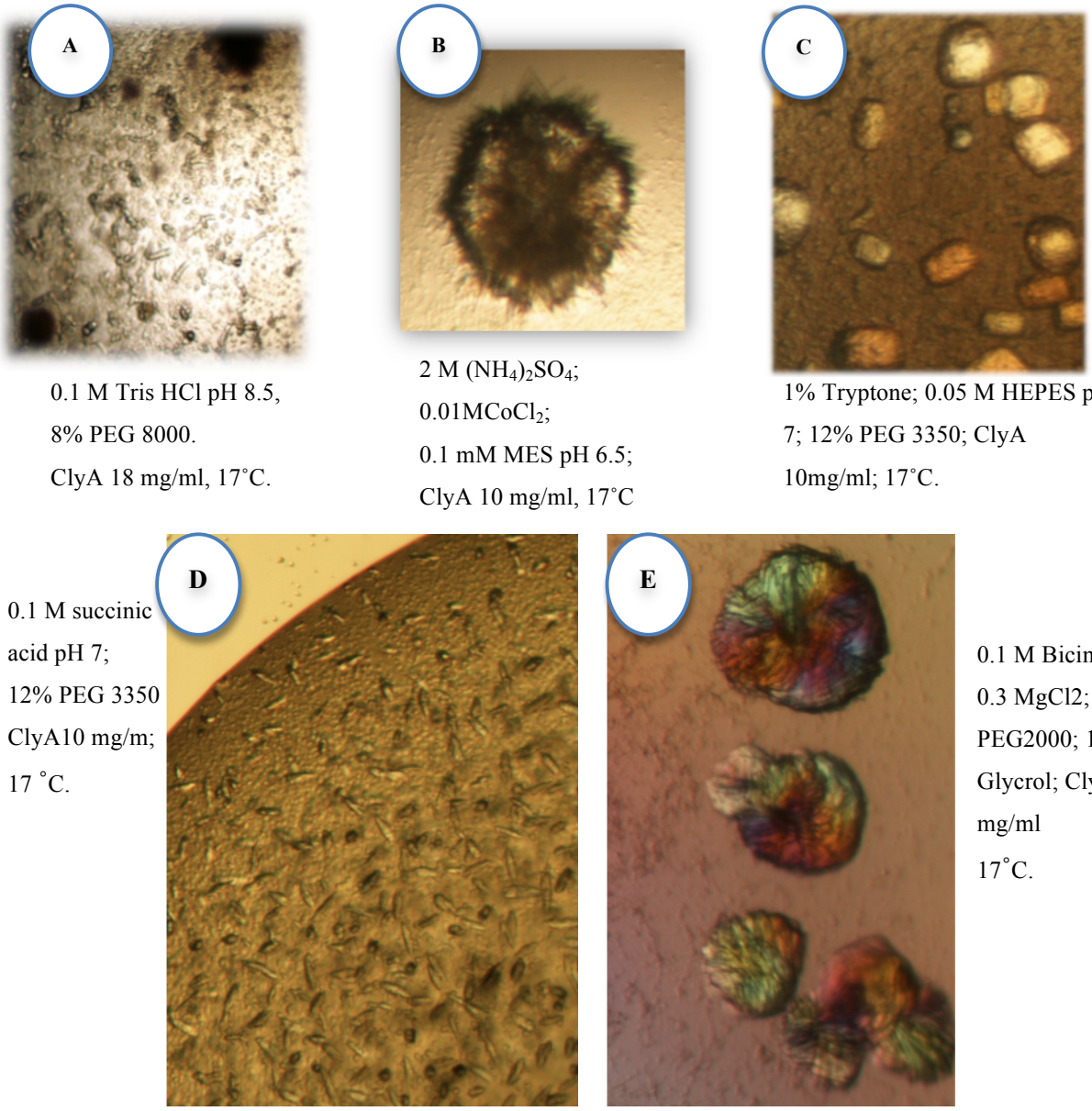


Figure 4.10: Crystallization hits for cleaved monomeric APEC ClyA. (A) Microcrystals; (B) Discs of precipitate which have tiny needles projecting from them; (C) Heavy circular drops with some small rectangular crystals, which did not diffract; (D) Small crystals; (E) Fused plate-like crystals.

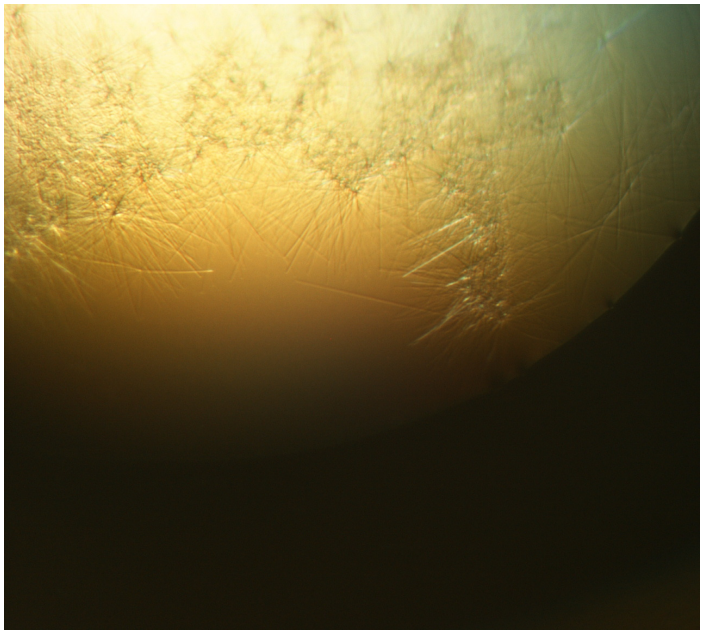


Figure 4.11: Growth of needles after chemical modification by Lysine methylation of APEC ClyA in 70% MPD, 0.1 M HEPES AT pH 7.5. ClyA 10 mg/ml, 17 ˚C.

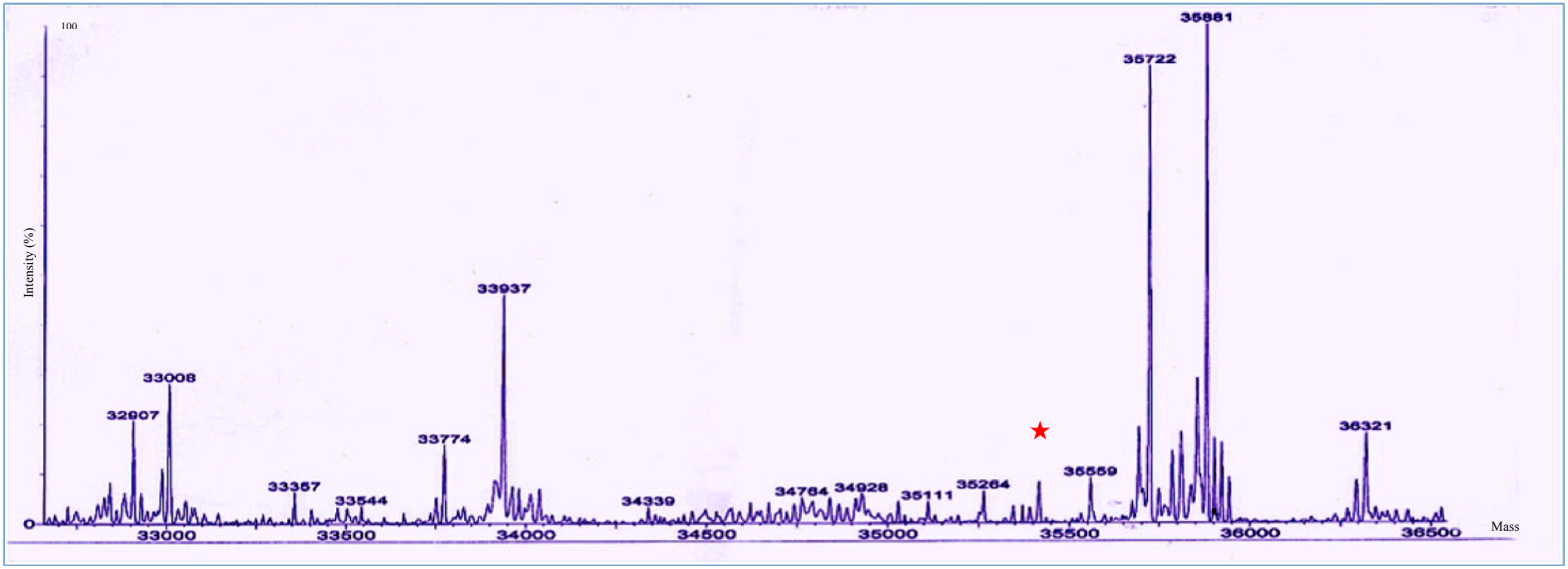


Figure 4.12: LC- MS peptide analysis of APEC ClyA (20 µg) after chemical modification by Lysine methylation. MW of unmodified APEC ClyA would be around 35402 (Red star).

4.5 3D Crystallization trials of non-cleaved monomeric APEC ClyA

Tags in recombinant proteins change the solubility of the protein or increase the aggregation of the pure protein (Carson et al, 2007). Therefore many crystallographers suppose that tags hinder crystallization and therefore cleaving and removing the tag before screening for crystallization conditions can be very important (Carson et al, 2007). On the other hand, the tag may be sometimes helpful or even required for crystallization. (Tajika et al, 2004) determined the structure of PH0828, a hypothetical protein from *Pyrococcus horikoshii OT3*, where the His tag on a long C-terminal helix protrudes outside the molecule and packs with another molecule about the crystallographic two-fold axis.

The considerations discussed above suggested making some crystallization trials of the His tagged-APEC ClyA. The same protocol of expression and purification was used except thrombin was not used for cleaving. Thus the purified APEC ClyA contained a 46-residue linker (FCLTLRRRY TMGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSEF) at the N terminus (see Figure 4.6).

Non-Cleaved monomeric APEC ClyA crystallized in the initial crystallization screen (pH clear) in condition C3 [0.1 M MES at pH 6 and 5% PEG 6000] using a protein concentration of 8.5 mg/ml at 17 ºC, after 3 weeks. This condition was optimized and the best rectangular shaped crystal were obtained in condition 0.1 M MES at pH 6 and 12% PEG 6000 and 0.1 M MES at pH 7.5 and 12% PEG 6000 after 7 weeks at 17 ºC by using protein concentration 10 mg/ml (Figure 4.13).

Crystals were cryo cooled at 100 K using mother liquor plus 20% ethyleneglycol as the cryoprotectant. The crystals were tested in house on a MAR345 image plate using 1.542 Å wavelength X-rays generated by a Rigaku Micromax 007 copper rotating anode generator.

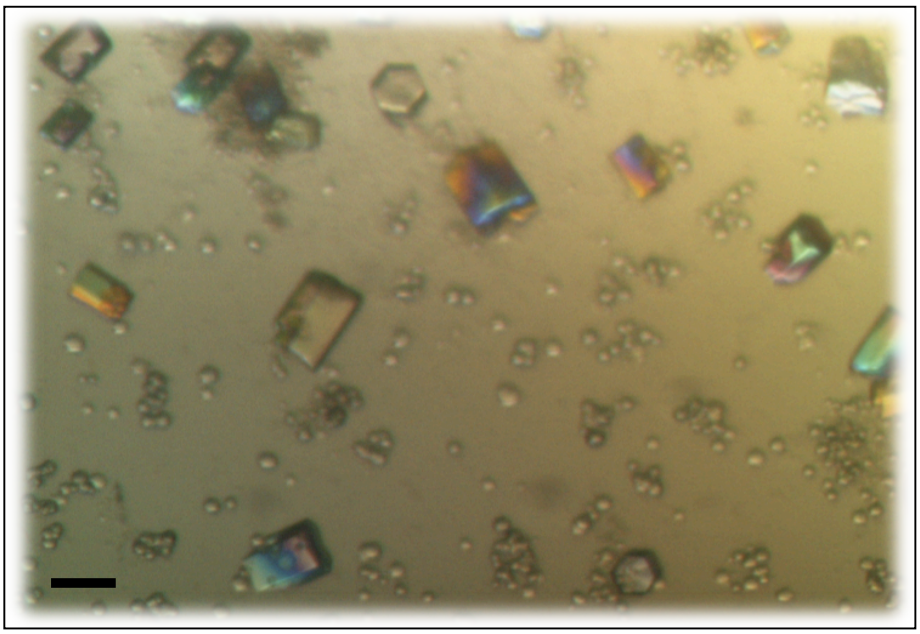


Figure 4.13: Rectangular shaped crystals for non-cleaved monomeric APEC ClyA.
The crystals were growth in 0.1 M MES at pH 6 and 12% PEG6000 using 10 mg/ml of protein 17 ˚C. Scale bar: 100 µm.

Two initial diffraction images separated by 90° were collected, with each image a 1° oscillation over a 30 minutes exposure but only diffracted to very low resolution. The crystals were kept then tested by synchrotron radiation but the diffraction did not go beyond 15 Å.

4.6 Crystallization of membrane protein

It was decided to attempt to crystallize APEC ClyA in its membrane-associated pore form. The crystallization of membrane proteins is classified into three common methodologies (Smyth et al, 2003):

- **2D crystallization:** this is used in electron microscopy (more details in chapter 9). In 2D crystals the membrane protein molecules are ordered in a single lipid bilayer.

- **Type I 3D crystallization:** this type of crystal is built up of the layers of 2D crystals in the third dimension.

- **Type II 3D crystallization:** in this type, membrane proteins are crystallized within a detergent micelle. Detergent-solubilized protein molecules interact together by the hydrophilic part of the protein (protruding from the detergent-covered hydrophobic part of the protein transmembrane regions).

In this project, for crystallization of APEC ClyA pore form (transmembrane protein), 2D crystallization trials for EM (as described later in chapter 9) were attempted, as well type II 3D crystallization trials for X-ray crystallography as described below.

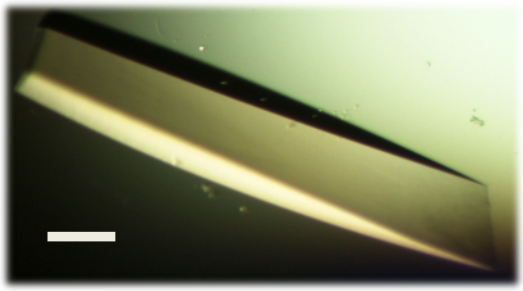
4.7 3D Crystallization trials of APEC ClyA pore form (transmembrane protein)

3D crystallization trials for the APEC ClyA pore form were set up using three types of detergents separately: dodecyl maltoside (β-DDM), n-Decyl-maltoside (DM), octyl β-D-glucopyranoside (β-OG).

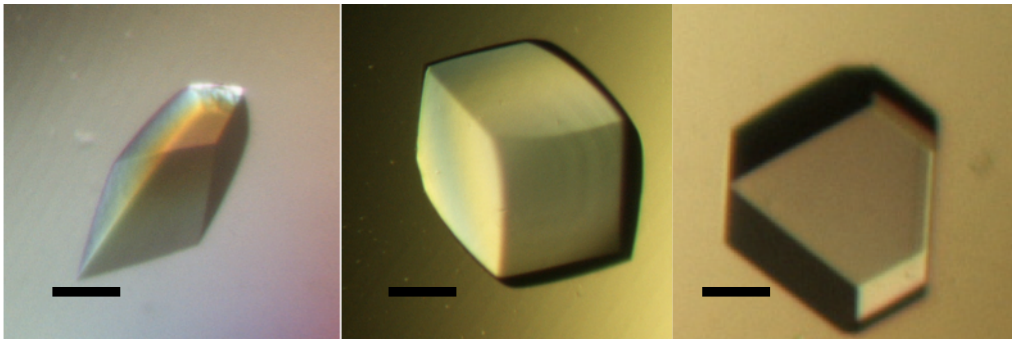
To produce APEC ClyA pores the protein was diluted to 0.2 mg/ml and incubated separately with 0.1% of detergent (DM, β-DDM, OG) and 2 mM dithiothreitol (DDT) for 16 hours at 4 ºC. The complex was concentrated to 6-7mg/ml and then it was set up for crystallization trials.

Commercial membrane protein screens from Jena Bioscience were used to set up initial crystallization conditions. Small protein crystals were observed in A7 [20% PEG500, 50 mM Tris HCl at pH 8.5 and 100 mM $MgCl_2$], and B1 [10% PEG2000, 50 mM Tris HCl at pH 8.5 and 300 mM $MgCl_2$] using DDM-protein complex at 17 ºC.

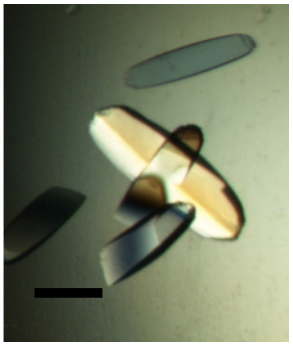
Promising crystallization trials were optimized. The best conditions were found in different concentrations of PEG 400 (11-15%), 50 mM Tris HCl at pH 8.5 and 100 mM $MgCl_2$. Crystals were also seen at different concentrations of PEG 2000 (8-10%), 50 mM Tris HCl at pH 8.5 and 300 mM $MgCl_2$ by mixing 5 µl of protein complex solution with 5 µl of reservoir solution and equilibrating against 1ml reservoir solution. Most of the crystal growths took two months. It was noticed that the crystals have a variety of forms (tetragonal shaped crystals, hexagonal shaped crystal with dimensions ~ (80 µm x100 µm x1000 µm), cubic shaped crystal with dimensions ~ (200 µm x 200 µm x 200 µm), wedge-shaped crystal with dimensions ~ (200 µm x100 µm x 300 µm), pentagonal shaped crystal with dimensions ~ (50 µm x 200 µm x 300µm); and plates look like crystal (30 µm x 40 µm x 300 µm). these crystals have a different qualities (Figure 4.14).



Hexagonal shaped crystal



Wedge-shaped crystal Cubic shaped crystal Pentagonal shaped crystal



Plates

Figure 4.14: **APEC ClyA pore form crystals.**
There were multiple crystalline forms, with a variety of quality and dimensions in different concentration of PEG 400 (11- 15%), 50 mM Tris HCl at pH 8.5 and 100 mM $MgCl_2$. Scale bar: 100 µm.

4.7.1 Cryoprotection

In order to optimize the cryoprotectant conditions, different concentrations of glycerol, ethylene glycol, Sucrose, and PEG400 were mixed separately with reservoir solution. It was found that all the crystal forms started dissolving immediately after soaking in most cryoprotectant agents and the best result was obtained by mixing PEG400 with reservoir solution. As a result, PEG400 was used as the cryo-protectant agent, and for X-ray diffraction testing the crystals were immersed in cryoprotectant solution continuing 30% PEG400, 20 mM Tris HCl at pH 8.5, 20 mM $MgCl_2$, 0.1% DDM for two hours then flash cooled in a stream of liquid nitrogen cooled to 100 K.

4.7.2 X-ray diffraction

Diffraction was tested for all crystal forms on the in-house source at 100 K using a Rigaku X-ray generator with a Mar345 detector. All crystal forms diffracted to very low resolution (around 20 Å) and only the hexagonal shaped crystals diffracted to beyond 15 Å. These crystals were kept and later they were tested with synchrotron X-ray beams at the Diamond Light Source in Oxford, United Kingdom but most of them diffracted to 12 Å resolution.

4.8 Techniques for maximizing diffraction

To test whether the cryo-protectant was damaging the crystals it was found that all APEC ClyA pore crystal forms also did not diffract when tested at room temperature without freezing. (For more details see chapter 2, section 2.2.1). As well, there was no improvement in the diffraction when very freshly grown crystals were freshly cooled.

Two methods were used to attempt to improve the diffraction of APEC ClyA pore form crystals: crystal annealing and crystal dehydration.

4.8.1 Crystal annealing

Crystal annealing can be considered as a technique that lies between room temperature mounting and cryo-crystallography. This technique involves rounds of room temperature mounting, cryogenic data collection, transfer to cryoprotected mother liquor at room temperature and then again data collection under cryogenic conditions. Thus, crystal annealing is generally a freeze, thaw, freeze processes; and during its inauguration it was greeted with a lot of skepticism (Newman, 2006). Nevertheless, this technique has been shown sometimes to improve crystal packing and resolution (El Omari et al, 2011; Harp et al, 1998). Alternatively, instead of transferring the crystal into the cryoprotected mother liquor, the cold stream of gas coming from liquid nitrogen can be blocked for few seconds to thaw the crystal and then allowed to flow again during data collection until the quality of the data improves. This method has a major advantage over dipping the crystal in the mother liquor at room temperature because the latter can change the crystal orientation during remounting while the former does not suffer from this problem.

Some of the APEC ClyA pore form crystals were frozen in a stream of liquid nitrogen, and then warmed them up for 2-3 seconds, and finally re-frozen again. No improvement was observed in the diffraction.

4.8.2 Crystal dehydration

The main purpose of protein crystal dehydration is to reduce the solvent content of the crystal. The gradual increase in precipitant concentration technique was used (Heras & Martin, 2005). This technique can be considered as a continuation of the hanging or sitting drop techniques once the protein crystals have been observed. After the crystals have been observed, they are transferred from one crystallisation reservoir to another with a gradual increase in precipitant concentration using the same crystallisation technique. This allows the solvent content of the crystal to

decrease as water vapour diffuses out of the crystal to the highly concentrated reservoir solution, which can improve crystal packing quality. The advantage of this method is the crystals can be reproduced since the humidity is controlled during the dehydration process. Even through dehydration alone has been observed to improve the crystal quality (Sam et al, 2006) though improvement in crystal packing and shrinkage of the lattices, a combination of crystal annealing and dehydration has also been shown to produce very good data (Newman, 2006).

Here, the dehydration technique was used by transferring APEC ClyA pore form crystals gradually in to a higher concentration of PEG400 (15%, 17%, 19%, 21%) for one to two minutes at each concentration prior to immersing in cryo-protectant. However the crystals deteriorated (Figure 4.15).

Generally, was found that annealing and dehydration methods did not improve the quality of the data of APEC ClyA pore form crystals but it is unclear why some protein crystals respond to these techniques but others do not.

4.9 Heavy metal soaking

The initial strategy to solve the phase problem prior to the SeMet expression was heavy metal soaking to produce heavy metal derivatives for multiple isomorphous replacement. Normally, Heavy-metal derivative crystals which are produced in this way diffract less well than native crystals. However, in some cases the heavy-metal derivative crystals have an improved resolution limit. For example the resolution limit of MscL protein crystals, a mechanosensitive ion channel from *Mycobacterium tuberculosis*, was improved from 7 to 3.5 Å after soaking in heavy-metal (Chang et al, 1998; Heras & Martin, 2005). It was believed that binding of heavy atoms to MscL protein stabilized a unique orientation of the crystals (Chang et al, 1998).

Some native APEC ClyA pore form crystals were transferred into drops of cryoprotectant solution continuing 30% PEG400, 20 mM Tris HCl pH 8.5,

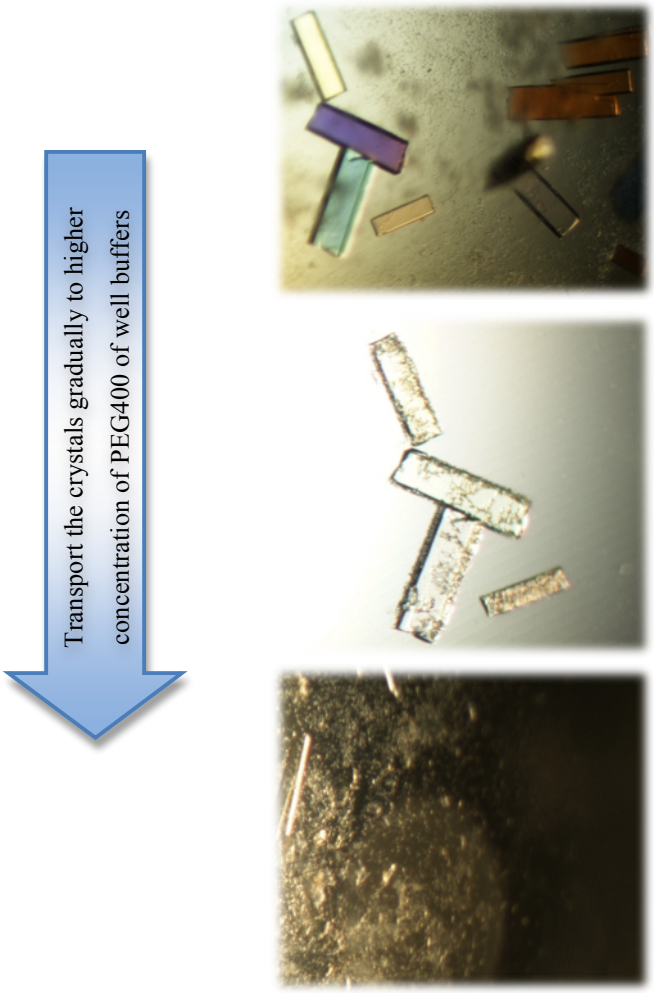


Figure 4.15: APEC ClyA pore form crystals deteriorate after dehydration.
These crystals were obtained from condition 14% PEG 400, 50 mM Tris HCl at pH 8.5 and 100 mM $MgCl_2$ then transferred gradually in to higher concentration of PEG400 (15%, 17%, 19%, 21%) prior to immersing in cryoprotectant. The crystals became worse.

20 mM $MgCl_2$, 0.1% DDM supplemented with 2 mM of Ethyl Mercuryphosphate (EMP) and 2 mM Ethylmercurithiosalicylic acid, sodium salt ($C_9H_9HgNaO_2S$), separately for three hours then frozen in a stream of liquid nitrogen cooled to 100 K. The Mercury compounds helped to improve the resolution limit of the APEC ClyA pore form crystals from 15-10 Å to 8-7 Å using synchrotron radiation (Figure 4.16). The beamline fluorescence detector showed that the heavy atom was present in the crystal.

Mercury interacts strongly with free sulfhydryl groups of cysteine residues (the native ClyA sequence has two cysteine residues) to form covalent S-Hg bonds. It is assumed that this covalent modification could have improved the crystal packing quality.

4.9.1 X-ray diffraction of APEC ClyA pore derivative crystals

The X-ray diffraction patterns of the best APEC ClyA pore form crystal diffracted to 8-9 Å (Figure 4.16 A) but the data were fielded for processing due to low -resolution limit. The exposure time was 1 minute per image and the crystal-to-detector distance was 779.2 mm at wavelength 0.9795 Å.
The second best APEC ClyA pore form crystal data was collected on the Diamond synchrotron light source to 7 Å resolution (Figure 4.16 B). A total of 900 images were collected with an oscillation angle of 0.2°, the exposure time was 0.5 second per image and the crystal-to-detector distance was 446.9 mm at wavelength 1.009 Å. The collection of the data strategy was optimized using MOSFLM (Leslie, 1994), then were processed by XDS (Kabsch, 2010). APEC ClyA pore form crystal belonged to space group *C222*, with unit-cell parameters a = 195.8 Å, b = 358.3 Å, c = 194.4 Å. The data-collection and processing statistics are listed in Table 1.

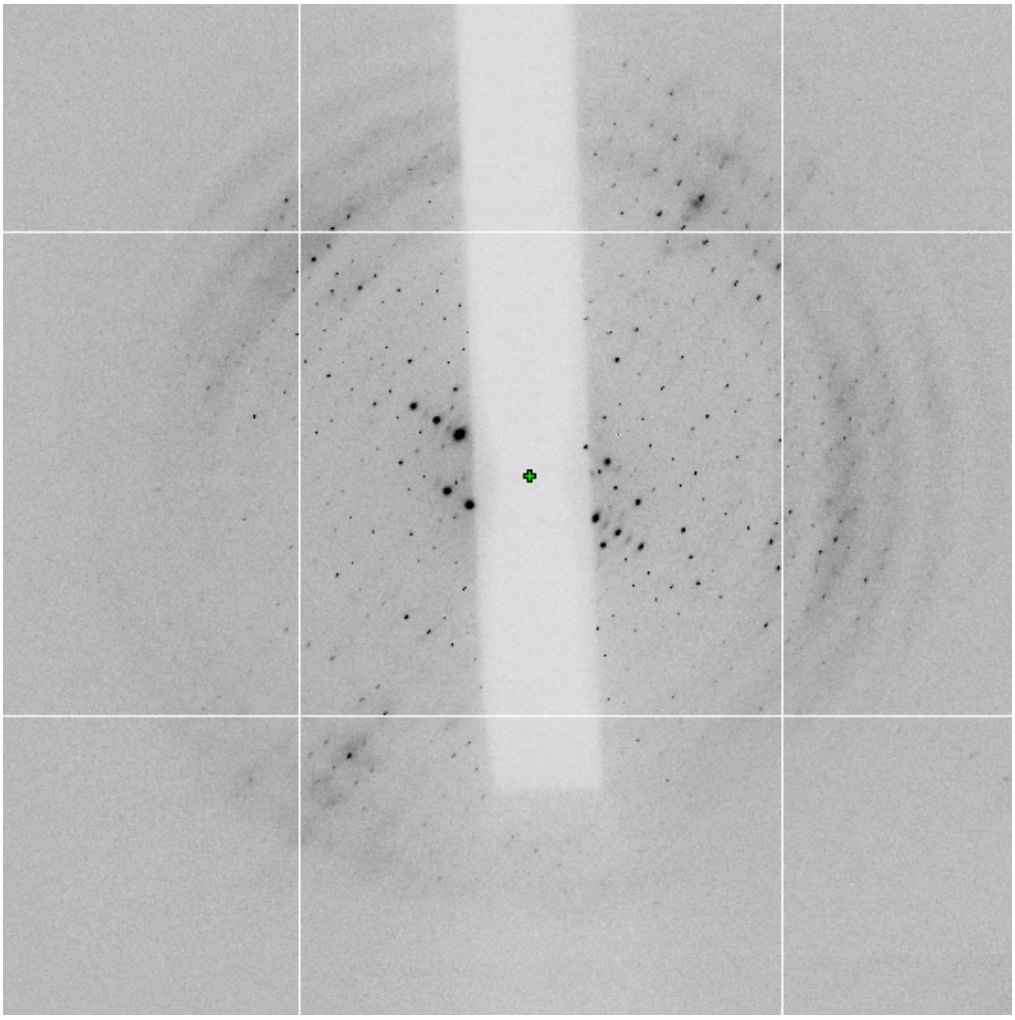


Figure 4.16 A: The X-ray diffraction patterns of an APEC ClyA pore form crystal.
X-ray diffraction pattern of an APEC ClyA crystal that had been soaked in 2 mM EMP collected to 8- 9 Å resolution using beamline I02 at Diamond light source. The exposure time was 0.60 s, the crystal-to-detector distance was 779.2 mm, wavelength 0.9795 Å and the oscillation range per image was 1°. The diffraction image was collected on an ADSC Q315 CCD detector.

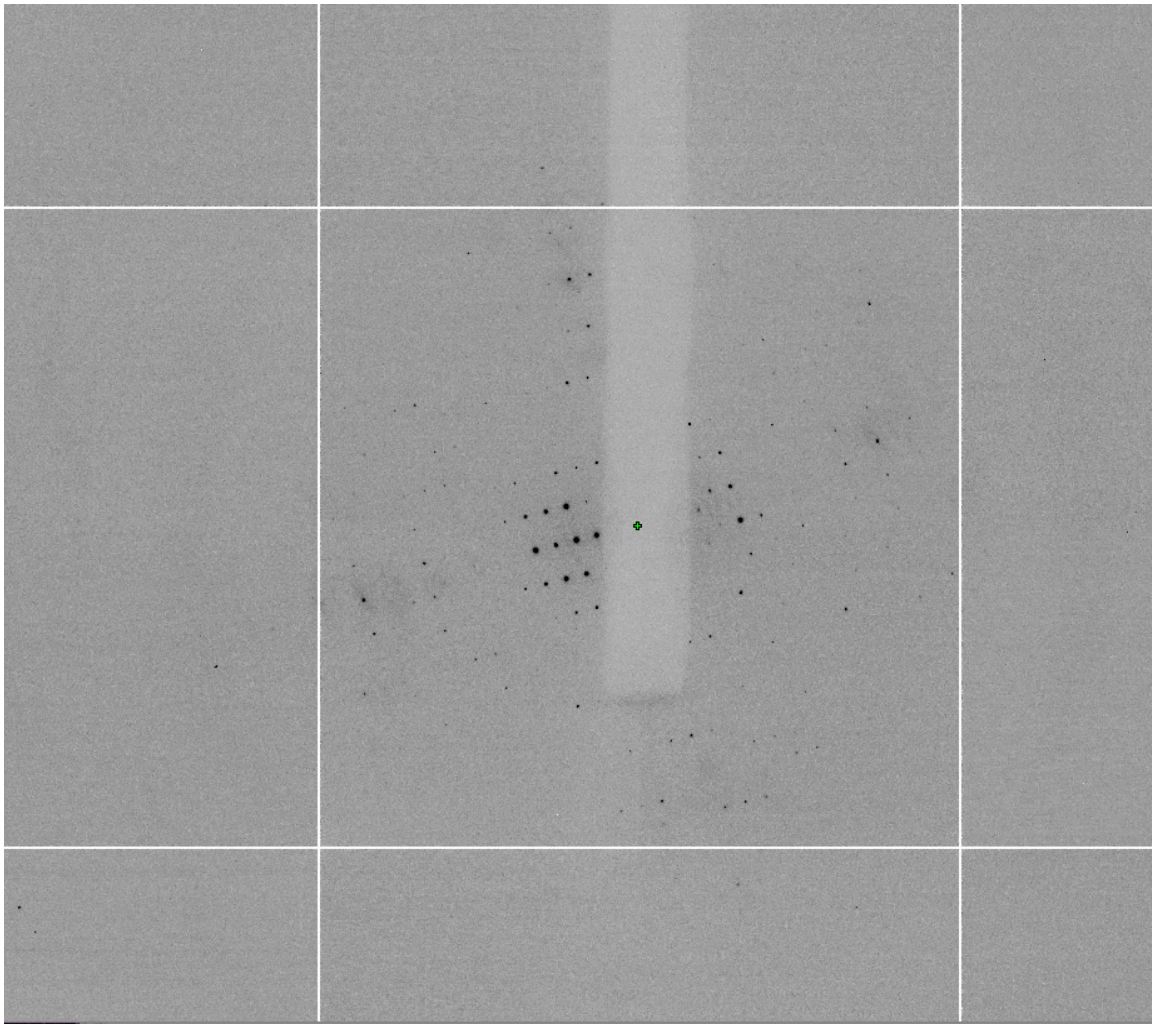


Figure 4.16 B: X-ray diffraction patterns of an APEC ClyA pore form crystal.

X-ray diffraction pattern of an APEC ClyA crystal that had been soaked in 2 mM EMP collected to 7 Å resolution using beamline I02 at Diamond light source. The exposure time was 0.5 s, the crystal-to-detector distance was 446.9 mm, wavelength 1.009 Å and the oscillation range per image was 0.2°. The diffraction image was collected on an ADSC Q315 CCD detector.

Table 4.1: Data collection statistics from the APEC ClyA crystal processed in space group C222.

Space group	*C222*		
Unit-cell parameters (Å, °)	*a*= 196.16 *b*= 358.96 *c*= 194.78 α=90 β=90 γ=90		
Crystal system	Orthorhombic		
	Overall	**Inner Shell**	**Outer Shell**
Low resolution limit (Å)	81.51	81.51	6.62
High resolution limit (Å)	6.45	28.87	6.45
R_{merge}	0.081	0.025	1.540
Total No. of observations	100070	692	6832
Total No. Unique	14016	164	992
Mean ((I)/sd(I))	14.7	27.5	1.2
Completeness	98.8	87.9	97.1

4.10 Matthews number (*Vm*) calculation

The Matthews number (*Vm*) (Matthews, 1968) was calculated using the web page http://www.ruppweb.org/Mattprob/ (Kantardjieff & Rupp, 2003), which suggested there are approximately 20 ClyA monomers in the asymmetric unit. Since the structure of the pore form of *E. coli* K12 has 12 molecules in the pore (Mueller et al, 2009) (2WCD), this may indicate that there are two 12 meric pores in the asymmetric unit. Alternatively there may be two 11 mers or two 10 mers. Assuming 20 meric oligomer the crystals would have a solvent content of 48.9% (the Matthews coefficient was 2.4 $Å^3$/Da). In order to investigate the possible rotational non-crystallographic symmetry in the APEC ClyA pore form crystal the POLARRFN self-rotation function (CCP4: Supported Program) analysis was used in the resolution range 7 Å. For two-fold rotation, corresponding to kappa= 180°, there are 5 peaks (Figure 4.17) which are the expected crystallographic 2-fold symmetry axes of the space group *C222*, however there is an extra peak (marked X in Figure 4.17) which shows evidence of an additional 2-fold axes of non-crystallographic symmetry. In 4-fold, corresponding to kappa= 90° there is a further peak (marked Y in Figure 4.17) that may indicate non-crystallographic 4-fold symmetry. For 12 fold rotation, corresponding to kappa = 30° there is some elevated density but no clear evidence for 12-fold symmetry. A self-Patterson synthesis was calculated with 50–7 Å resolution data to look for pseudo- translational symmetry in the APEC ClyA crystal. However this showed only the origin peaks at (0, 0, 0,) and (½, ½, 0) which are expected from the *C222* space group symmetry.

While our effort was under way to find high quality pore form crystals to acquire higher resolution data, the structure of *E. coli* K-12 ClyA, which has 75% amino acid similarity to APEC ClyA, was determined to 3.3 Å resolution by the group of Prof. N. Ban in Switzerland (Mueller et al, 2009). This structure had a 12-meric pore. Because of the similarity between K-12 ClyA and APEC ClyA and the very slow progress on APEC ClyA project, it was decided to move onto the investigation of the Nhe toxin.

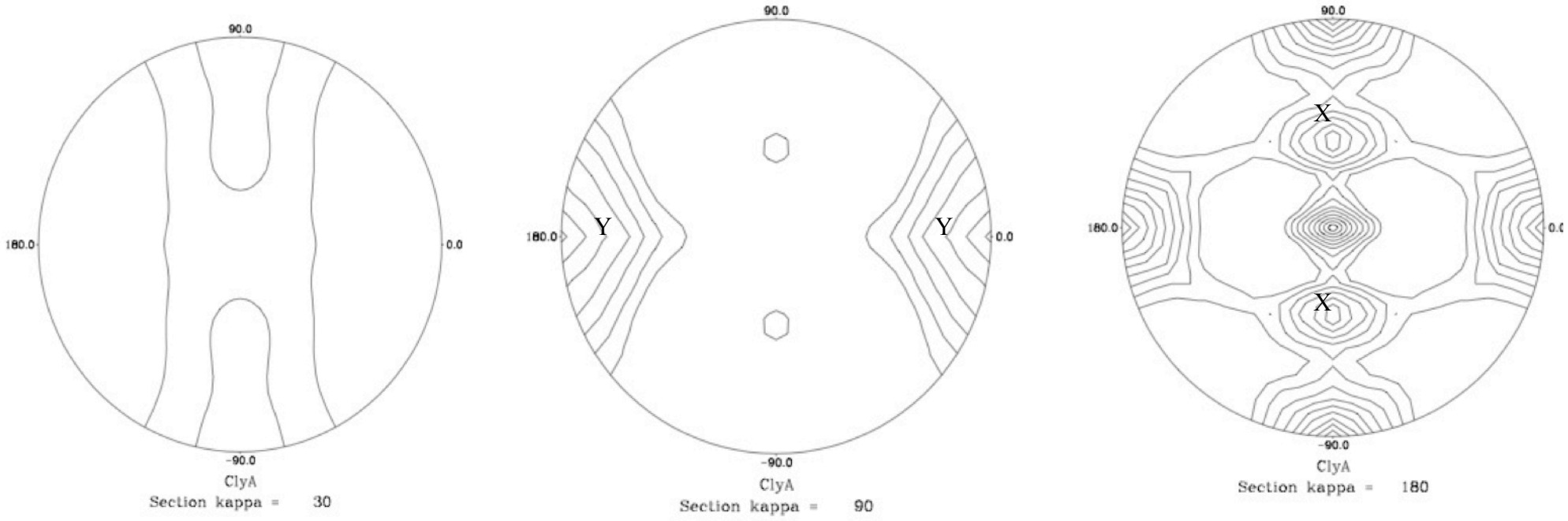


Figure 4. 17: Self-rotation function of APEC ClyA that was calculated using the program POLARRFN in the resolution range 50 Å – 7 Å. X labelling indicates that additional 2-fold axes of non-crystallographic symmetry. Y labelling indicates that non-crystallographic 4-fold symmetry.

4.11 Biochemical studies on APEC ClyA

4.11.1 Effect of APEC ClyA on mammalian erythrocytes

The effect of APEC ClyA on red blood cells (RBC) was assayed using the hemolysis assay described by (Rowe & Welch, 1994). Known concentrations (20 μg/ml) of purified APEC ClyA protein were incubated for 15 minutes, 30 minutes and 1 hour with defibrinated horse and sheep erythrocytes in saline Tris-HCl at pH 7. The amount of hemolysis was determined spectrophotometrically by measuring hemoglobin release at 543 nm in cell-free supernatants. The hemolytic activity (100%) was determined by incubation of red blood cells in water, in place of buffer and measuring the A_{543} of the supernatant. APEC ClyA has high hemolytic activity against both defibrinated horse and sheep RBCs, which was unsurprising since the blood comes from similar class of vertebrate (mammalian organisms). Hemolytic activity increased with time and 100 % of cell lysis appeared just after one-hour incubation with APEC ClyA (Figure 4.18 and Figure 4.19). The cell lysis of APEC ClyA is comparable with other groups of pore forming toxins, such as *E. coli* hemolysin (Eberspacher et al, 1989).

4.11.2 Effect of pH on erythrocyte lysis by APEC ClyA

Certain cytolysins show a dependence on a specific pH for activity such as listeriolysin O (LLO), a cholesterol-dependent cytolysin, which has pH-dependent activity (Schuerch et al, 2005). LLO is active at acidic media pH (<6), however the toxin activity is decreased at neutral pH, due to rapid and irreversible denaturation of the domain 3 of LLO, this domains usually insert into the membrane to form the transmembrane β-barrel pore (Schuerch et al, 2005).

To investigate APEC ClyA activity at different pH values, samples of horse blood in saline Tris- HCl at pH 8, pH 7 and pH 5 were incubated with purified APEC ClyA (20 μg/ml) for 30 minutes at 37 °C. From Figure 4.20, it can be seen that APEC ClyA has high hemolytic activity at both pH 5 and

7 with 87% lysis of cells. However, at alkaline pH 8 there was slightly less hemolytic activity, which means that altering the pH had little effect on the efficiency of APEC ClyA-mediated RBC lysis. This result suggests that APEC ClyA is not as dependent on a specific pH for activity as LLO. No significant change in APEC ClyA activity was noticed at pH 5, 7, and 8. It was not possible to examine pHs outside of this range due to the instability of RBCs. This result suggests that APEC ClyA is cytolytically active across a range of pHs around physiological pH. It seems that APEC ClyA could be effective in numerous locations in mammalian cells, whether it be exposed to acidic pH (e.g. the digestive system) or alkaline pH (e.g. the blood, as the pH of blood is slightly alkaline).

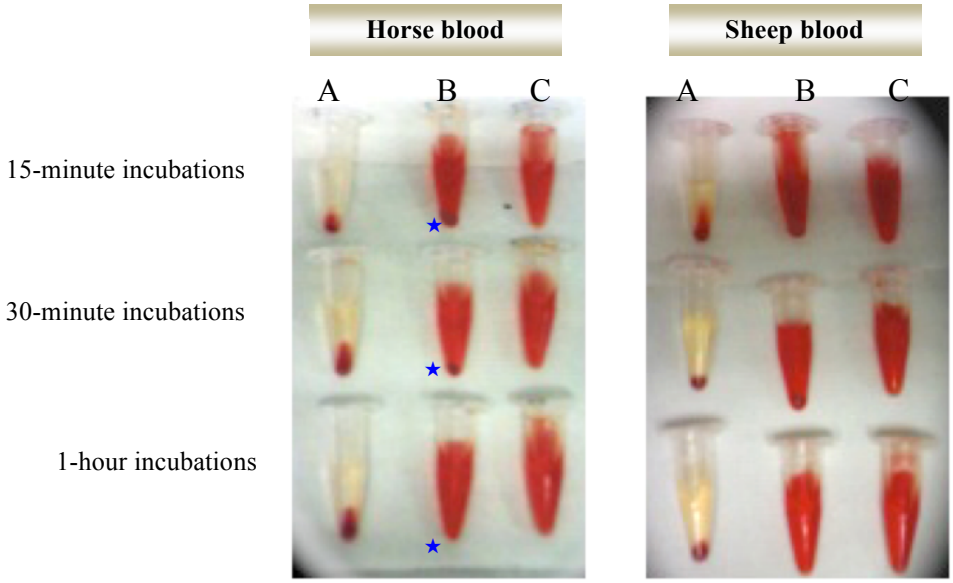


Figure 4.18: Hemolytic activity of APEC ClyA on RBCs with increasing incubation time.

(A) Negative control: RBC with saline Tris-HCl at pH 7.
Saline Tris-HCl at pH 7 buffer acts as an isotonic solution that keeps the RBCs as a pellet after centrifuging.
(B) RBC with saline Tris-HCl at pH 7 and 20 µg/ml of APEC ClyA: it is clear APEC ClyA has high hemolytic activity. An increase in the incubation time with APEC ClyA caused an increase in cell lysis shown by the cell pellet decreasing in size (★).
(C) Positive control: RBC with water: It is obvious that water causes full cell lysis and hemoglobin was released completely.

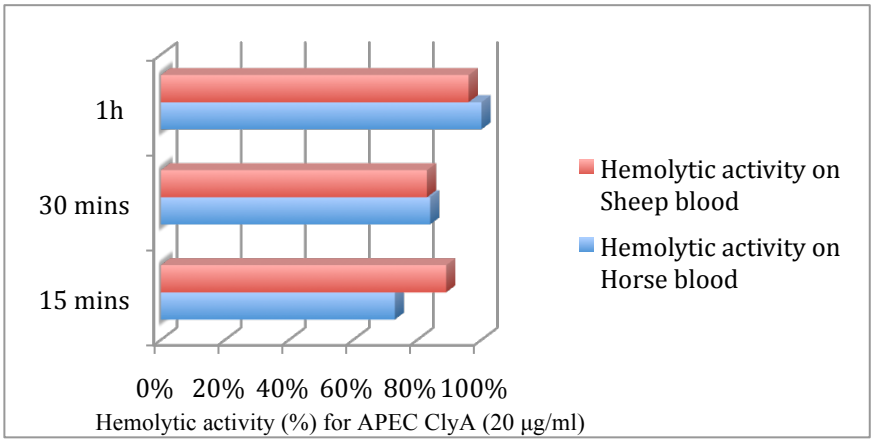


Figure 4.19: Hemolytic activity of defibrinated horse blood and sheep blood at different times 15 minutes, 30 minutes and 1 hour in saline Tris-HCl at pH 7 using 20 µg/ml of purified APEC ClyA protein.

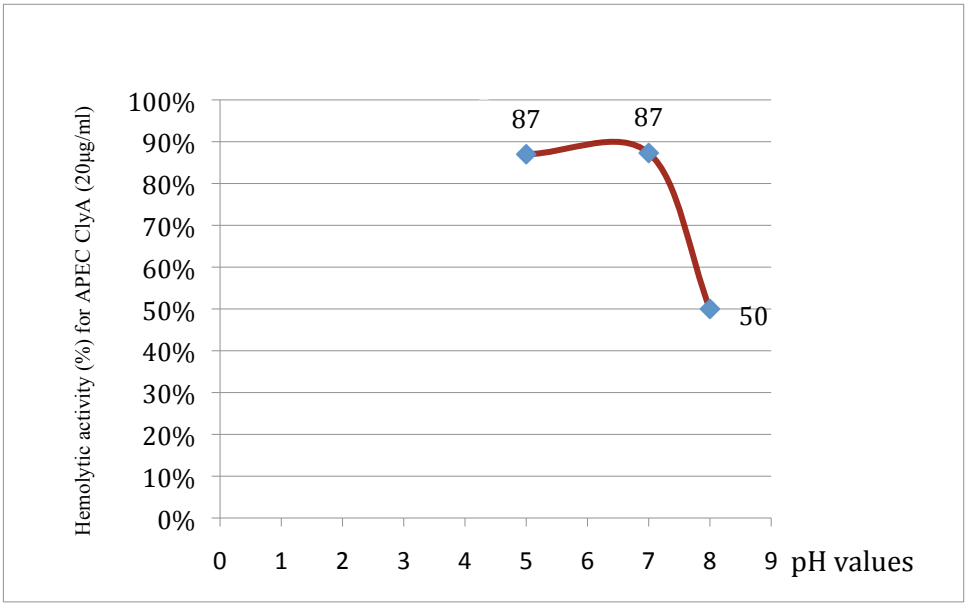


Figure 4.20: Hemolytic activity of APEC ClyA (20 µg/ml) at different pHs on horse blood cells.
Saline Tris-HCl at pH 7 was used at pH 8, pH 7 and pH 5 for 30 minutes at 37 °C.

Reference

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* **215:** 403-410

Atkins A, Wyborn NR, Wallace AJ, Stillman TJ, Black LK, Fielding AB, Hisakado M, Artymiuk PJ, Green J (2000) Structure-function relationships of a novel bacterial toxin, hemolysin E - The role of alpha(G). *J Biol Chem* **275:** 41150-41155

Carson M, Johnson DH, McDonald H, Brouillette C, DeLucas LJ (2007) His-tag impact on structure. *Acta Crystallogr D Biol Crystallogr* **63:** 295-301

Chang G, Spencer RH, Lee AT, Barclay MT, Rees DC (1998) Structure of the MscL homolog from Mycobacterium tuberculosis: A gated mechanosensitive ion channel. *Science* **282:** 2220-2226

Eberspacher B, Hugo F, Bhakdi S (1989) Quantative study of the binding and hemolytic efficiency of *Escherichia coli* hemolysin. *Infect Immun* **57:** 983-988

El Omari K, Hoosdally SJ, Tuladhar K, Karia D, Vyas P, Patient R, Porcher C, Mancini EJ (2011) Structure of the leukemia oncogene LMO2: implications for the assembly of a hematopoietic transcription factor complex. *Blood* **117:** 2146-2156

Harp JM, Timm DE, Bunick GJ (1998) Macromolecular crystal annealing: Overcoming increased mosaicity associated with cryocrystallography. *Acta Crystallogr D Biol Crystallogr* **54:** 622-628

Heras B, Martin JL (2005) Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr D Biol Crystallogr* **61:** 1173-1180

Mueller M, Grauschopf U, Maier T, Glockshuber R, Ban N (2009) The structure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism. *Nature* **459:** 726-730

Newman J (2006) A review of techniques for maximizing diffraction from a protein crystal in stilla. *Acta Crystallogr D Biol Crystallogr* **62:** 27-31

Qiu X, Janson CA (2009) Improving protein crystallizability by modifications and engineering. In *IUL Biotechnology Series: protein cystallization*, T. B (ed), 2nd edn. La Jolla California: International university line

Rowe GE, Welch RA (1994) Assays of hemolytic toxins. *Bacterial Pathogenesis, Pt A* **235:** 657-667

Sam MD, Abbani MA, Cascio D, Johnson RC, Clubb RT (2006) Crystallization, dehydration and preliminary X-ray analysis of excisionase (Xis) proteins cooperatively bound to DNA. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **62:** 825-828

Schuerch DW, Wilson-Kubalek EM, Tweten RK (2005) Molecular basis of listeriolysin O pH dependence. *Proc Natl Acad Sci U S A* **102:** 12537-12542

Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B (2003) Crystal structures of fusion proteins with large-affinity tags. *Protein Sci* **12:** 1313-1322

Tajika Y, Sakai N, Tamura T, Yao M, Watanabe N, Tanaka I (2004) Crystal structure of hypothetical protein PH0828 from *Pyrococcus horikoshii*. *Proteins-Structure Function and Bioinformatics* **57:** 862-865

Walter TS, Meier C, Assenberg R, Au KF, Ren JS, Verma A, Nettleship JE, Owens RJ, Stuart DI, Grimes JM (2006) Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14:** 1617-1622

Wyborn NR, Clark A, Roberts RE, Jamieson SJ, Tzokov S, Bullough PA, Stillman TJ, Artymiuk PJ, Galen JE, Zhao L, Levine MM, Green J (2004) Properties of haemolysin E (HlyE) from a pathogenic *Escherichia coli* avian isolate and studies of HlyE export. *Microbiology* **150:** 1495-1505

Chapter 5: NheA Preparation and Crystallization

This chapter will describe the work involved in the preparation and crystallization of native NheA, which including preparation and crystallization of selenomethionine incorporated protein for experimental phasing (see chapters 6 and 7).

5.1 Expression of NheA and SeMet NheA

5.1.1 Expression of native NheA

NheA was expressed in *E. coli* strain BL21(λDE3) transformed with PEXP5-CT/TOPO-NheA vector (see section 3.3.4 for transformation protocol).

The primary culture was grown at 37 ˚C in LB media containing Ampicillin (100 μg/ml) with shaking (250 rpm) overnight.

Six litres of LB liquid media in six flasks was prepared, then ampicillin was added to a final concentration of 100 μg/ml. The primary culture was added to the secondary culture medium in a ratio of 2:100 (primary: secondary). Therefore 20 ml of primary culture was added to each 1000ml of secondary culture medium. The secondary cultures were grown at 37 ˚C with shaking (250 rpm) until the bacterial growth reached the logarithmic growth phase. This was measured when the cultures reached an OD_{600} of 0.6 to 0.8. NheA expression was induced by the addition of IPTG to final concentration of 200 μg/ml for 20 hour. The *nheA* gene is under the control of the lac operon promoter genes, as is the whole vector inserted into the *E. coli.* IPTG was used to promote the Lac operon promoter site, and therefore induce expression of NheA protein (Figure 5.1).

5.1.2 Expression of SeMet NheA

SeMet NheA was expressed in the same *E. coli* strain as native protein. The primary culture was grown at 37 ˚C in LB media containing Ampicillin (100 μg/ml) with shaking (250 rpm) overnight. Two litres of LB liquid media in two litre flasks was prepared, and then ampicillin was added to a final concentration of 100 μg/ml. The primary culture was added to the secondary culture medium in a ratio of 2 ml primary: 100 ml secondary.

The cultures of *E. coli* BL21 (λDE3) were harvested at 4 ˚C (15,344 g, 20 minutes) and then the pellet was transferred to two litres of minimal media containing Ampicillin (100 μg/ml) and supplemented with amino acids that inhibit synthesis of the methionine biosynthetic pathway to encourage SeMet incorporation (section 3.4.5) and with the bases Adenine (0.50 g/l), Guanine (0.50 g/l), Thymine (0.50 g/l), Uracil (0.50 g/l). The media also included $MgSO_4.7H_2O$ (0.01 mg/ml) and Thiamine (0.004 mg/ml). Then the culture was warmed to 37 ˚C for 30 minutes before induction by 200 μg/ml IPTG. The culture containing IPTG was left to grow for 24 hours with shaking (250 rpm) at 37 °C before being centrifuged (15,344 g, 20 minutes). The supernatant containing the liquid media was removed and the pellets containing the cells were retained and kept for purification (Figure 5.1).

5. 2 Purification of NheA and SeMet NheA

5.2.1 Purification of NheA

The cultures of *E. coli* BL21 (λDE3) were harvested at 4 ˚C (15,344 g, 20 minutes) and lysed in 50 mM Tris HCl at pH 8 for sonication. The sonication was performed in 3 cycles for 20 seconds at 16-micron amplitude, with cooling of the samples between sonication cycles. The lysed bacteria were centrifuged at 70000 g for 10 minutes.

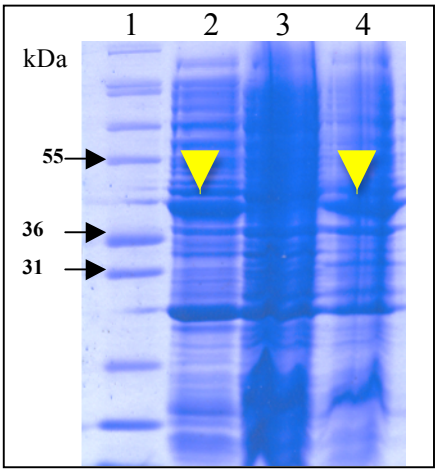


Figure 5.1: 12% SDS-PAGE analysis of *E. coli* strain BL21 (λDE3), carrying plasmids for the production of NheA protein.
Visualised by staining with Coomassie blue stain. Lane1: Marker 12 from top to bottom: 200, 116, 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: induction of NheA by 200 µg/ml IPTG in LB media for 20 h; Lane 3: supernatant (control); Lane 4: induction of SeMet NheA by 200 µg/ml IPTG in minimal media for 24 h.
The molecular weight of the protein should be around 41 kDa (yellow pointer).

For the purification of NheA the following protocol was used for both the native NheA and the SeMet-derivative, with some amendments of the protocol for NheA SeMet as explained in section 5.2.2.

The cell extract was applied onto a 20 ml DEAE sepharose fast flow ion exchange column pre-equilibrated with loading buffer (25 mM Bis Tris at pH 5.8). Un-bound proteins were washed from the column using the same buffer. The fractions were run on a 12% SDS polyacrylamide electrophoresis gel for analysis (Figure 5.2). Western immunoblotting was carried out to identify and locate NheA (as described on section 3.4.4.1). Consequently, the western immunoblotting result (Figure 5.3) confirms the presence of NheA in the fractions.

Fractions were chosen for the least amount of contaminants and loaded onto a Hydroxylapatite Chromatography (HA) Column followed by elution with a linear gradient of 0 to 0.25 M Na_3PO_4 at pH 6.8 (Figure 5.4).

The eluted protein concentration was measured using a spectrophotometer to measure the OD at 595nm. The fractions with highest protein content were then loaded onto a Hi load Superdex 200 gel filtration column, which was pre-equilibrated with 50 mM Tris-HCL at pH 8.5 and 0.5 M NaCl (Figure 5.5). The fractions were run on a SDS PAGE for analysis along with a sample from each step in the purification to allow comparisons. N-terminal sequencing of the sample confirmed it was NheA.

5.2.2 Purification of SeMet NheA

For purification of the selenomethionine labeled protein the protocol was essentially the same, but with the following amendments:

- The eluted proteins from the anion exchange chromatography were cut with 2 M - 2.5 M of $(NH_4)_2SO_4$ for 2 hours at 4 ˚C, after spinning down the solution at 70000 g for 10 minutes.

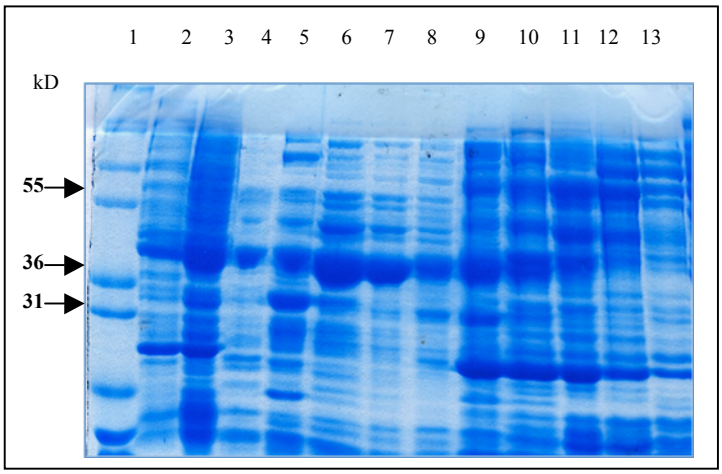


Figure 5. 2: **12% SDS-PAGE analysis of DEAE sepharose column purification for NheA**.
Visualised by staining with Coomassie blue stain. Lane1: Marker12 from top to bottom: 200, 116, 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: pellet; Lane 3: cell extract; Lane 4: to lane 13 the fractions collected from the DEAE sepharose column.

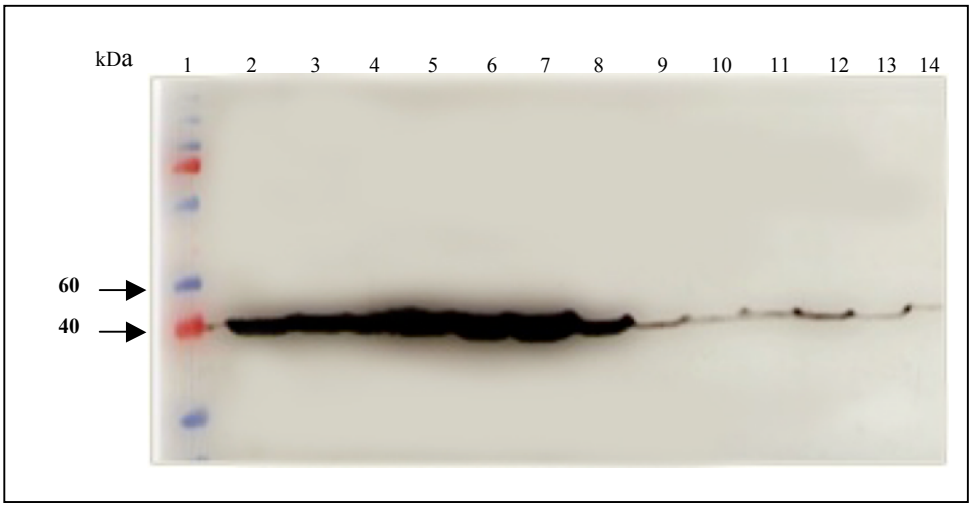


Figure 5. 3: **Western immunoblot of 12% SDS-PAGE analysis of DEAE sepharose for NheA**.
Visualised by chemiluminescence method. The primary antibody MAb 1A8 was used.
Lane 1: Marker Nexus View 20 kDa Dual Color from top to bottom: 140, 120, 100, 80, 60, 40, and 30 kDa; Lane 2 cell extract; Lane 3 to 14 fractions from DEAE sepharose for NheA.

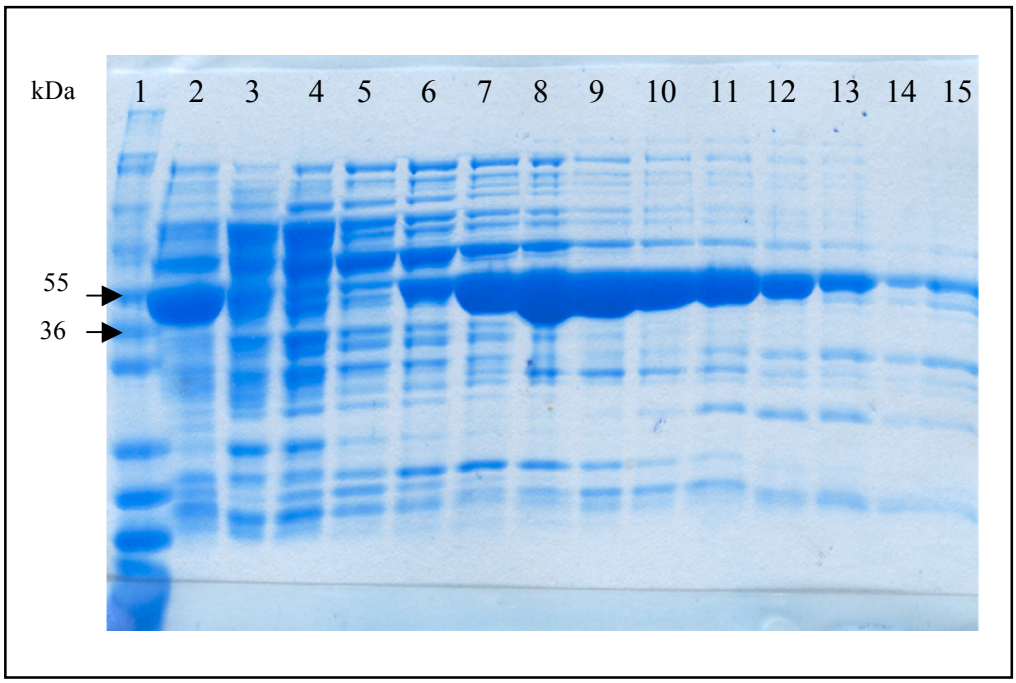


Figure 5.4: **12% SDS-PAGE analysis of a Hydroxylapatite Chromatography (HA) purification for NheA**.
Visualised by staining with Coomassie blue stain. Lane 1: Marker12 from top to bottom: 200, 116, 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: HA load (after DEAE sepharose); Lane 3: to lane 15 the fractions collected from the HA column.

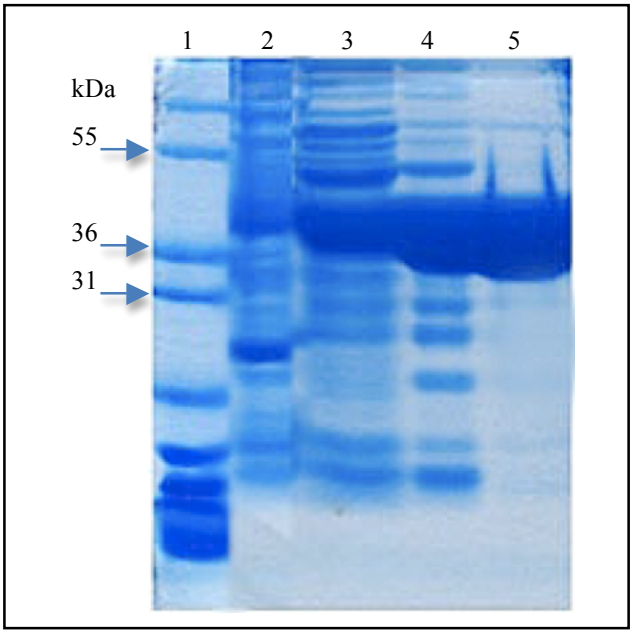


Figure 5. 5: **12% SDS-PAGE analysis of purification for NheA**.
Visualised by staining with Coomassie blue stain. Lane 1: Marker12 from top to bottom: 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: Cell extract; Lane 3: After DEAE sepharose column; Lane 4: After HA column; Lane 5: After gel filtration column

- Four chromatographic steps were used: the first step was DEAE sepharose column fast flow ion exchange (Figure 5.6). Then gel filtration and hydroxylapatite chromatography were applied in the reverse order compared to native protein (Figure 5.7, and Figure 5.8), and an additional anion exchange chromatography step on a Resource Q column HPLC was used as a final polishing step for purification of the SeMet-derivative NheA to remove small contaminant proteins (Figure 5.9).

5.3 Properties of pure NheA protein

The purified NheA was homogeneous with molecular weight approximately 41 kDa shown by SDS–PAGE, which is near to the expected weight.

To prepare an NheA sample for crystallization, the buffer in the sample was exchanged with 10 mM sodium phosphate at pH 6.8 using a 0.5 ml Zeba^{TM} Desalting Column (Thermo Sciencific). Protein concentration was determined by the Bradford method. Both native NheA and SeMet-derivative NheA were concentrated in a Centricon centrifugal ultrafilteration device with a 10 kDa molecular-weight cutoff membrane to 10.5 and 9 mg/ml, respectively, before setting up crystallization trials.

5.4 Mass spectrometry analysis for NheA and SeMet NheA

The amino acid sequence of NheA (Figure 5.10) shows the existence of seven Methionine residues. Theoretically, when replacing Methionine by Selenomethionine the molecular weight of NheA should rise by approximately 47 Dalton per Selenomethionine. Native NheA and SeMet NheA were examined by liquid chromatography–mass spectrometry (LC-MS) to investigate the incorporation of SeMet in NheA. LC-MS analysis of SeMet NheA confirmed approximately, 100% incorporation of the 7 Selenomethionines in NheA and no evidence for oxidation to selenoxide as the mass peaks are the same breadth as native (Figure 5.11).

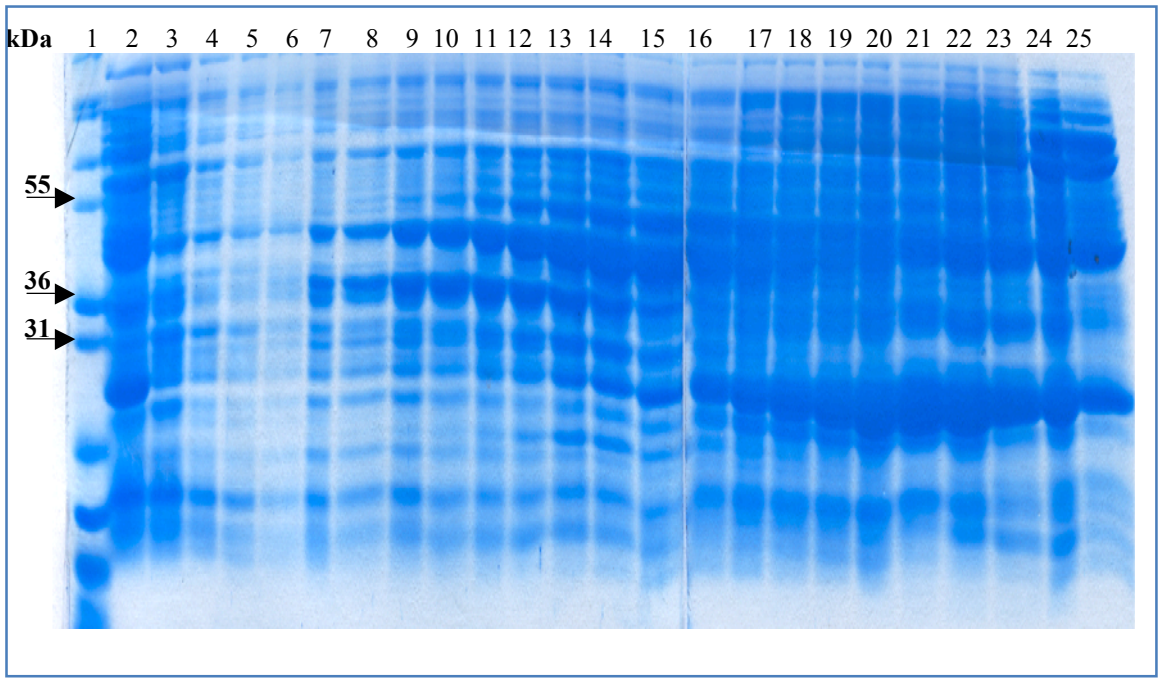


Figure 5. 6: **12% SDS-PAGE analysis of DEAE sepharose column purification for SeMet NheA**.
Visualised by staining with Coomassie blue stain. Lane 1: Marker12 from top to bottom: 200, 116, 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: cell extract; Lane 3: lane 25 the fractions collected from DEAE sepharose column.

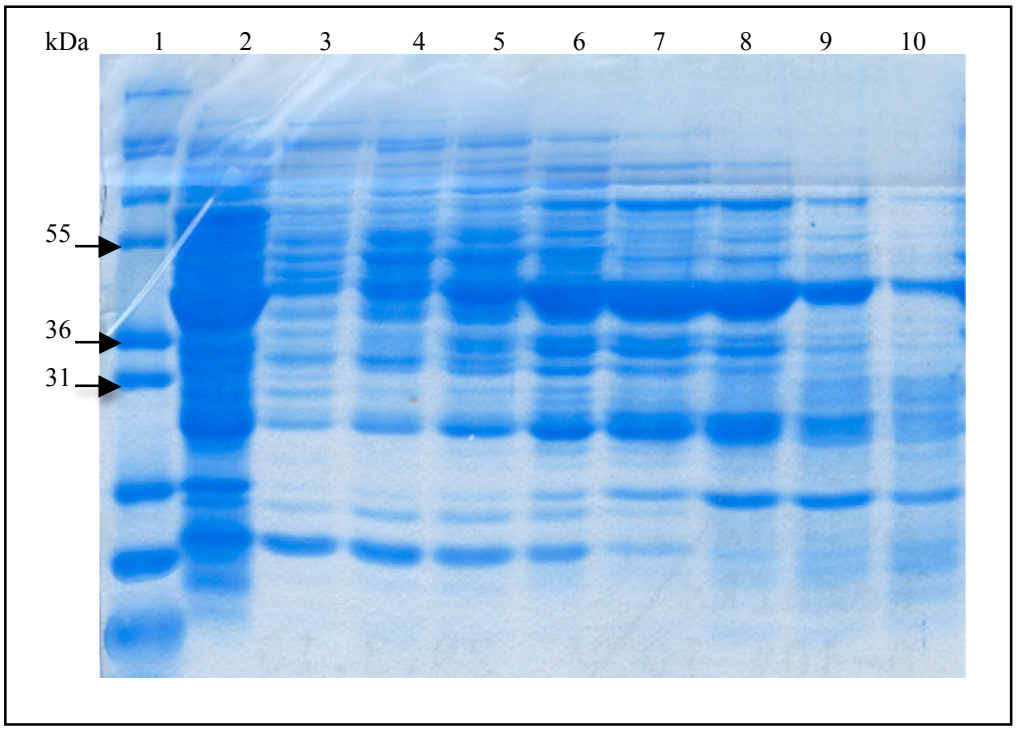


Figure 5. 7: **12% SDS-PAGE analysis of gel filtration of purification for SeMet NheA**. Visualised by staining with Coomassie blue stain. Lane 1: Marker12 from top to bottom: 200, 116, 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: gel filtration load; Lane 3 to Lane10: gel filtration column fractions.

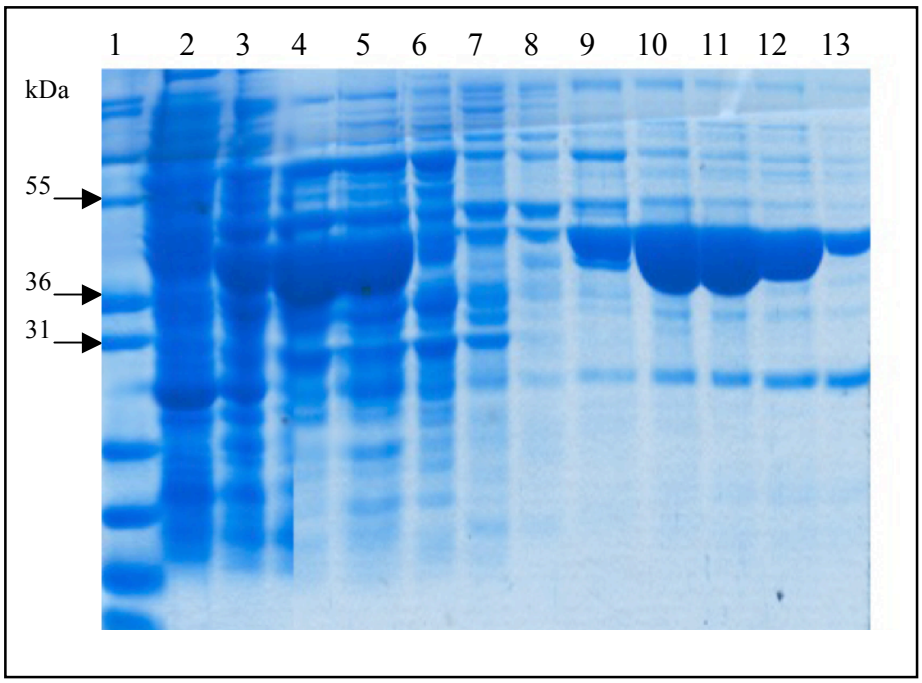


Figure 5. 8: **12% SDS-PAGE analysis of HA of purification for SeMet NheA**. Visualised by staining with Coomassie blue stain: Lane 1: Marker12 from top to bottom: 116, 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: cell extract; Lane 3: after DEAE sepharose column; Lane 4: after gel filtration column; Lane 5: HA load; Lane 7 to lane 13: HA fractions.

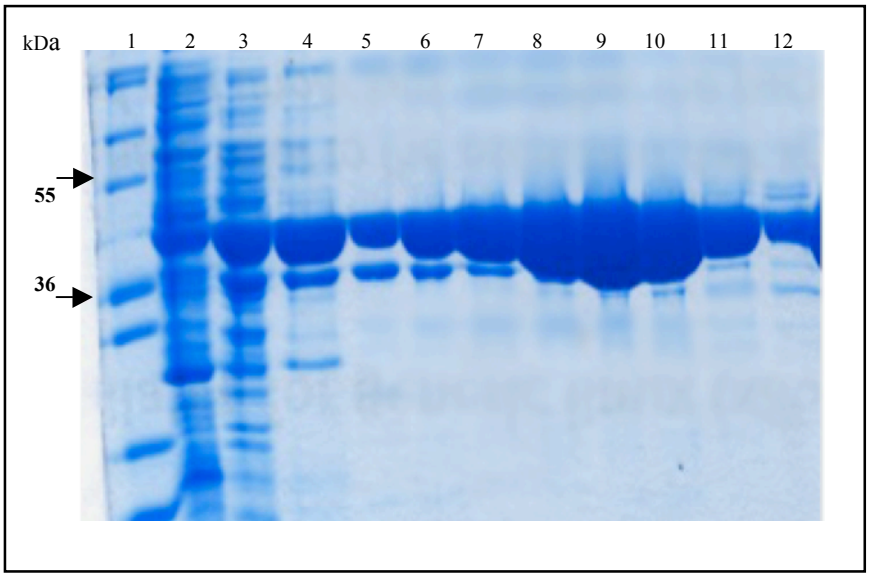


Figure 5. 9: **12% SDS-PAGE analysis of Resource Q for SeMet NheA purification.** Visualised by staining with Coomassie blue stain. Lane 1: Marker12 from top to bottom: 116, 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: cell extract; Lane 3: after HA column; Lane 4 to 12: the fractions from the Resource Q column. The less contaminated fractions (8-11) were collected for further experiments.

```
        10         20         30         40         50         60
KEGQTEVKTV YAQNVIAPNT LSNSIRMLGS QSPLIQAYGL VILQQPDIKV NAMSSLTNHQ

        70         80         90        100        110        120
KFAKANVREW IDEYNPKLID LNQEMMRYSI RFNSYYSKLY ELAGNINEDE QSKADFTNAY

       130        140        150        160        170        180
GKLQLQVQSI QENMEQDLLE LNRFKTVLDK DSNNLSIKAD EAIKTLQGSS GDIVKLREDI

       190        200        210        220        230        240
KRIQGEIQAE LTTILNRPQE IIKGSINIGK QVFTITNQTA QTKTIDFVSI GTLSNEIVNA

       250        260        270        280        290        300
ADSQTREAAL RIQQKQKELL PLIQKLSQTE AEATQITFVE DQVSSFTELI DRQITTLETL

       310        320        330        340        350        360
LTDWKVLNNN MIQIQKNVEE GTYTDSSLLQ KHFNQIKKVS DEMNKQTNQF EDYVTNVEVH
```

Figure 5. 10: **NheA amino acid sequence**.
The blue colour shows Methionine positions.

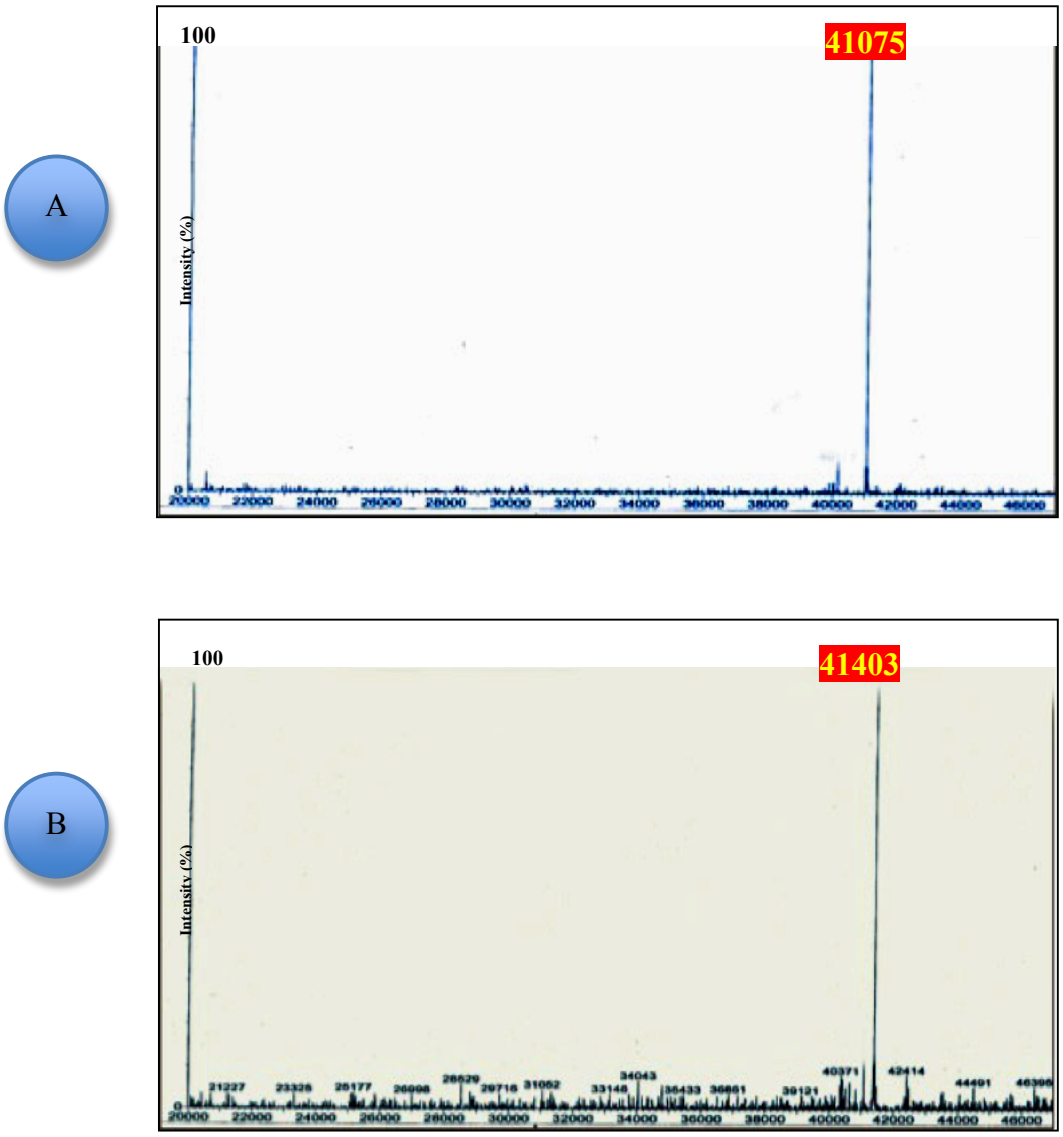


Chart 5.11: LC-MS analysis for NheA and SeMet NheA.

(A) Mass spectrum analysis of native NheA

(B) Mass spectrum analysis of SeMet NheA. It is clear the molecular weight of NheA was increased by roughly 328 Dalton, corresponding to changing all 7 Methionines to Selenomethionine. The result also indicates the high level of incorporation of SeMet.

5.5 3D Crystallization trials for NheA and SeMet NheA

5.5.1 Initial crystallization trials for NheA

Initial protein crystallization trials for NheA were done at 17 ˚C using the Matrix Hydra II Plus One crystallization robot, using crystallization screens JSCG+ Suite and the PACT screen (Qiagen) and vapour-diffusion (sitting-drop method). The initial indication was needles produced in one condition, H8 from the JSCG+ Suite [0.2 M $(NH_4)_2SO_4$, 0.1 M Bis-Tris at pH 5, and 25% PEG 3350] (Figure 5.12).

5.5.2 Optimization of crystallization conditions for NheA

Optimization of the initial hit was performed by reproducing the condition in 24-well plate format (Hampton Research) using the sitting drop vapour diffusion method (as described in section 2.1.2.2). Crystals could be reproduced in this format and so further optimisation grids were set up using different protein concentrations, PEG-concentrations, PEG molecular weights, salt concentrations, pH from 5 to 8, using different buffers. Further optimization conditions involved lowering the incubation temperature to 7 ºC and screening additional buffer components (additive screen).

The droplet was prepared by mixing 2 µl protein solution and reservoir 2 µl solution and then equilibrating against 1 ml of reservoir solution at 17 ˚C and 7 ˚C. Long plate-like crystals of native NheA were grown from 0.2 M $(NH_4)_2SO_4$, 0.1 M Bis-Tris at pH 5.5, and different PEG3350 concentration (20-25 %) after 3 days.

In order to reduce nucleation and create a single crystal for native NheA, further optimization of crystallization conditions was attempted. The best single crystal (Figure 5.13A) was obtained from the condition 0.2 M $(NH_4)_2SO_4$, 0.1 M Bis-Tris at pH 7, and 22 % PEG3350 after 20 days. The crystals had maximum dimensions of 80 µm x 200 µm x 400 µm.

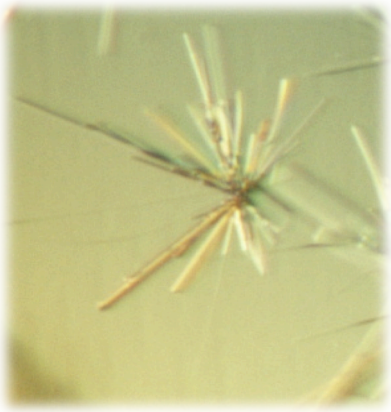


Figure 5.12: Growth of needles as the first hit for native NheA crystals in 0.2 M $(NH_4)_2SO_4$, 0.1 M Bis- tris pH 5, 25% PEG3350, NheA 10 mg/ml, 17˚C. Scale bar: 100 µm.

5.5.3 Search for a cryo-condition for the NheA crystals

Searching for a cryogenic condition for the crystals of a protein is a very significant point as this allows the crystals to be cooled to a very low temperature, which helps reduce radiation damage. When the crystal is exposed to the X-ray radiation, free radicals are generated and easily propagate throughout the entire crystal and damage the protein. To overcome this problem, a cryoprotectant solution should be chosen very carefully.

Initial searching for cryoprotectants focused on the use of ethelenglycol, glycerol, sugar, and low molecular weight PEG solutions. The best cryoprotectant solution is one that forms a glass, with a clear appearance and no visible ice ring on diffraction.

All the cryo-tests were carried out using freshly prepared solutions, which were flash-cooled to 100K in a cryo-stream, one-minute exposure, in the absence of the crystal. It was found that the best cryoprotectant for the NheA crystals was 0.2 M $(NH_4)_2SO_4$, 0.1 M Bis-Tris at pH 5.5, 22% PEG3350, and 20% ethyleneglycol.

The crystal was flash-cooled in liquid nitrogen at 100 K on the in-house source using the Rigaku X-ray generator with a Mar345 detector. The crystal showed excellent diffraction in-house to better than 2.4 Å resolution (Figure 5. 13 B). The crystal was then stored in liquid nitrogen for data collection at the Diamond Light Source in Oxford, United Kingdom for high resolution data collection.

5.5.4 3D Crystallization trials of SeMet NheA

Using a slight variation of the native NheA crystal conditions for SeMet-NheA, clusters of crystals (Figure 5. 14) were grown in 5 days from 0.2 M $(NH_4)_2SO_4$, 0.1 M Bis-Tris pH at 6, and 27% PEG 3350.

One crystal of SeMet-NheA was broken apart from a cluster, and then the crystal was mounted in a loop and soaked for a few seconds in 0.2 M $(NH_4)_2SO_4$, 0.1 M Bis-Tris at pH 6, and 27% PEG 3350 with 20% of ethylene glycol. X-ray diffraction was observed in-house to beyond 2.9Å resolution using the Mar435 detector. This crystal was saved in liquid nitrogen to send to the Diamond Light Source for high resolution data collection and to perform fluorescence scans on the crystal and to collect anomalous diffraction data.

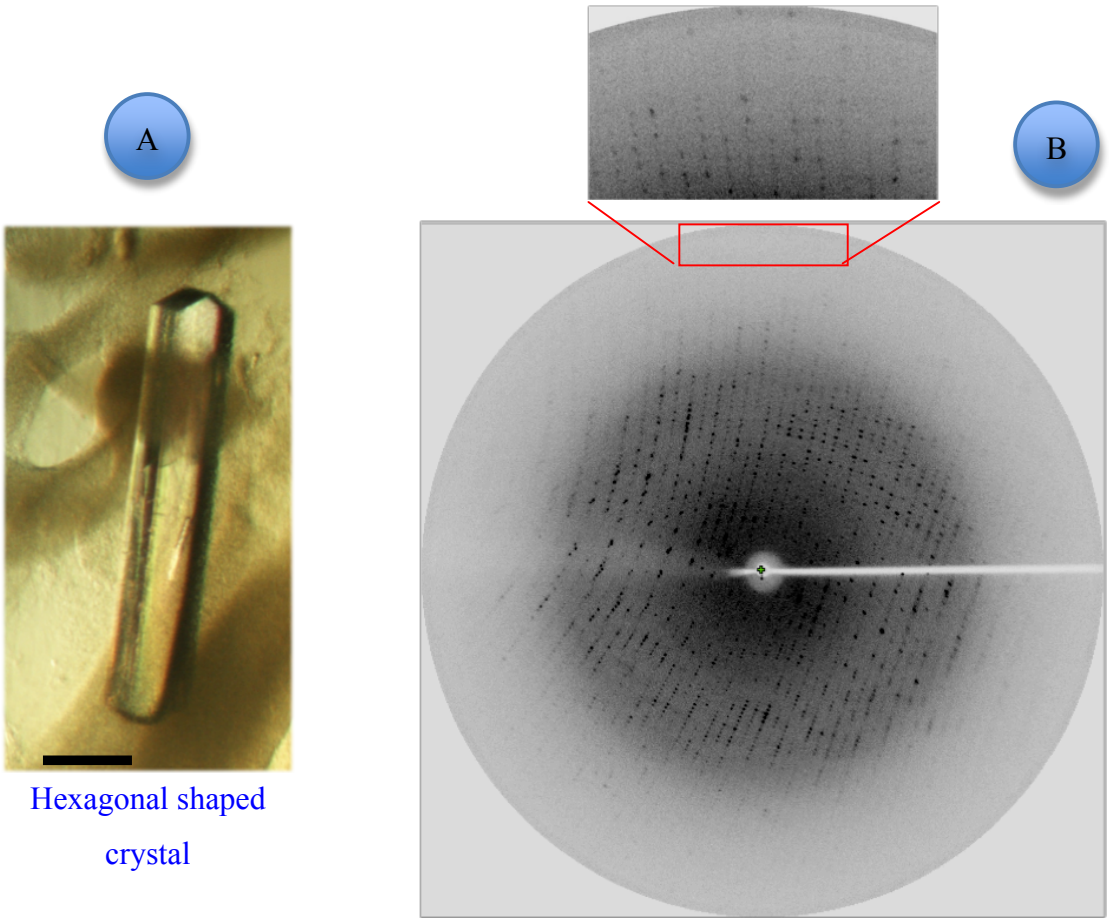


Figure 5.13: (A) A single crystal of native NheA grown in 0.2M $(NH_4)_2SO_4$, 0.1 M Bis- tris pH 7, 22% PEG3350 , NheA 10 mg/ml, 17 ˚C. Scale bar: 100 µm.
(B) Initial X-ray diffraction to 2.4 Å resolution from NheA crystal using beamline of Mar345 detector in house.
An enlarged portion of the image is revealed on the top.

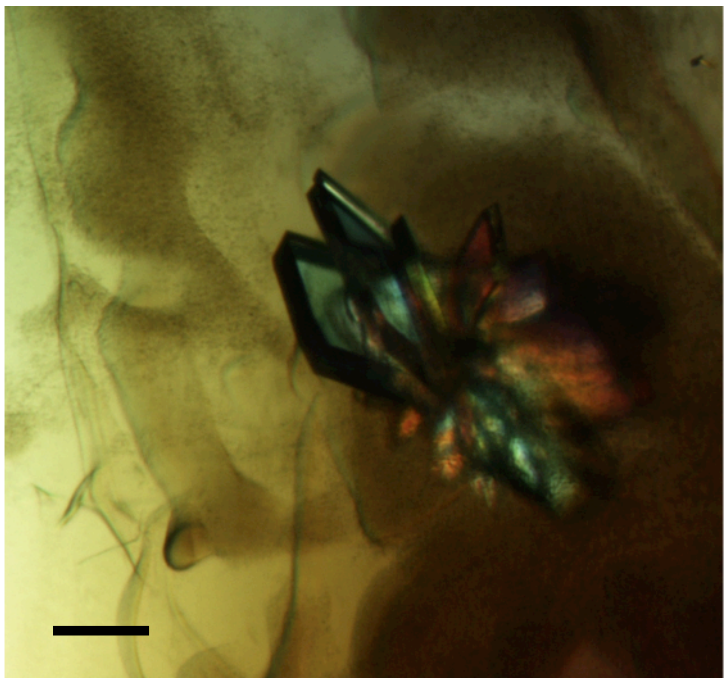


Figure 5. 14: A cluster of crystals of SeMet NheA grown in 0.2 M $(NH_4)_2SO_4$, 0.1 M Bis- tris pH 6, 27% PEG3350 , SeMet NheA 9.5 mg/ml, 17 ˚C. Scale bar: 100 µm.

Chapter 6: Experimental Structure Determination of NheA by SeMet MAD to 2.05Å

Attempts were made to solve the structure of NheA by molecular replacement using *E. coli* K-12 ClyA and *B. cereus* HblB as search models but they failed because there was no structure available with sufficient (>30%) sequence identity to NheA. Therefore, the structure had to be solved independently by multiple isomorphous replacement (MIR) or multiple anomalous dispersion (MAD) phasing. This chapter will describe the structure solution of *B. cereus* NheA at 2.05 Å resolution using SeMet MAD phases.

6.1 Space group determination and indexing

Native NheA and SeMet NheA crystals were initially evaluated on the MAR345 image plate using 1.542 Å wavelength X-rays generated by a Rigaku Micromax 007 copper rotating anode generator in Sheffield. Two 1° oscillations separated by 90° with a 30 minute exposure were used to initially determine the space group by autoindexing using MOSFLM (Leslie, 1994). The space group and autoindexing were confirmed at the Diamond Light Source in Oxford, United Kingdom using XDS (Kabsch, 2010) by testing two images separated by 90° with each image a 0.25° oscillation.

Both native NheA and the SeMet derivative belong to the monoclinic crystal system, which is characterized by unequal lengths of the three axes, *a*, *b*, and *c*, and the interaxial angles $\alpha = \gamma = 90°$, $\beta \neq 90°$.

The space group of native NheA and NheA SeMet crystals are C2. The native NheA crystal had a cell dimensions *a*=308.67 Å, *b*=58.17 Å, *c*=172.93 Å, $\alpha=90°$, $\beta=110.55°$, $\gamma=90°$. The cell dimensions of the NheA

SeMet crystal were *a* = 307.91 Å, *b*= 58.73 Å, *c*= 172.73 Å α=90°, β=110.40°, γ =90°.

6.2 Data collection and processing

As described in Section 3.7, the native NheA crystal was flash-frozen in liquid nitrogen at 100K and data were collected on the Diamond synchrotron light source to 1.79 Å resolution (Figure 6.1).

A total of 450 images were collected each with an oscillation angle of 0.5° but this data was interrupted two times due to technical problems of the synchrotron beam so that 199 images were discarded then the same 225° of data were recollected completely without interruptions. The exposure time was 0.25 second per image and the crystal-to-detector distance was 259.0 mm at wavelength 0.9686 Å. The strategy for collection of the native data was optimized using MOSFLM (Leslie, 1994), then the data were processed by XDS (Kabsch, 2010).

Native data analysis indicated that the data had poor quality at above 2.05 Å resolution. The processing of native data at 1.79 Å resolution indicates overall "R_{merge}" was 0.1 and 1.27 for the outer shell, and "Mean ((I)/ sd (I))" was 0.7 for the outer shell, and 5.3 overall. Thus the data was cut off to 2.05Å resolution. Final data collection statistics are shown in Table 6.1.

The SeMet crystal was also flash-frozen in liquid nitrogen at 100 K and the MAD data collected on the Diamond synchrotron light source. A fluorescence scan of the SeMet crystal was taken by Vortex fluorescence detector over an energy range of 12.62-12.67 keV, and revealed an absorption edge at 12.66 keV that confirmed the presence of selenium in the sample (See Figure 6.2) as already indicated by MS (See Figure 5.11). The fluorescence data were processed firstly in CHOOCH (Evans & Pettifer, 2001) to produce a chart illustrating the values of the anomalous scattering factors, f' and f'', which indicated the X-ray energies at points corresponding to maximum f' and f'' minimum.

Anomalous data (MAD data) were collected to 2.05 Å resolution (Figure 6.3) from one NheA SeMet crystal at three wavelengths: high remote 0.9784 Å, peak 0.9794 Å, and inflection 0.9797 Å. These were chosen near the selenium-absorption edge based on the fluorescence absorption spectrum of the crystal in order to maximize the f'' component (λ_1 peak), to minimize the f' component (λ_2 inflection) and to maximize Δf (λ_3 remote) in order to optimize the power of phasing from the incorporated Selenium atoms.

Anomalous data (MAD data) were collected on the I02 beamline at the Diamond Light Source in Oxford, United Kingdom. The diffraction images were collected on an ADSC Q315 CCD detector. A total of 540 images, with a 0.5º rotation per image, exposure time of 0.75 second per image and with the detector distance set at 225.0 mm, were collected at all three wavelengths. The collection of the anomalous data strategy was optimized using MOSFLM (Leslie, 1994) and then the data were processed in the same manner as the native data. The data collection statistics are shown in Table 6.1.

The result of data collection and processing for the anomalous data overall "R_{merge}" was below 0.1, and around 0.6 for the outer shell. The "Mean ((I)/ sd (I))" was between 2.5 and 2 for the outer shell, and over 7 overall. The Completeness was over 98% overall. In general these represent reasonable indicators for data quality. As well, there was no indication of crystal twinning and no indication of radiation damage.

These data processing statistics are good by the statistics of (Wlodawer et al, 2008). The data collection and processing statistics listed in Table 6.1.

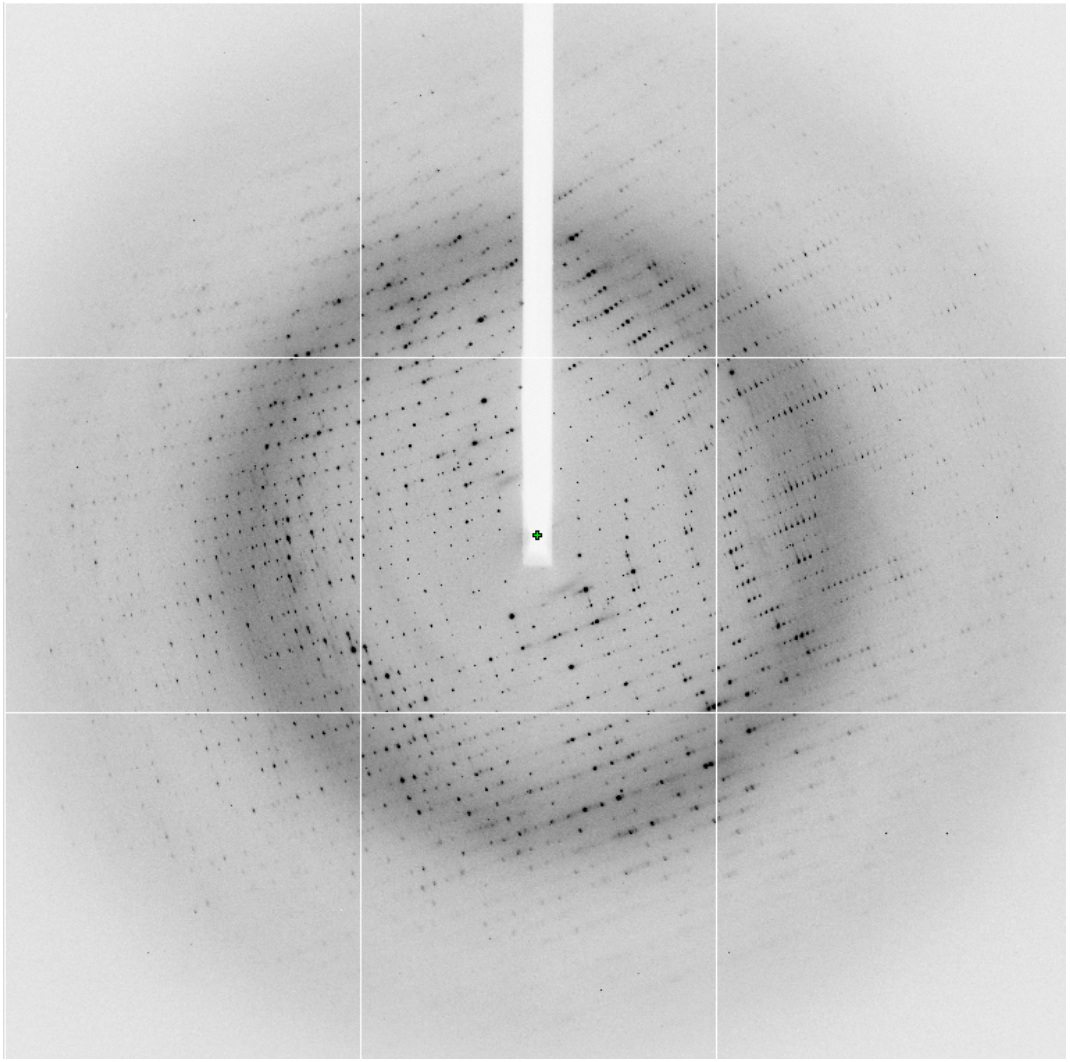


Figure 6.1: X-ray diffraction pattern of NheA crystal using beamline I02 of Diamond light source.

Diffraction intensities were collected to 1.79 Å resolution but the data werecut off to 2.05 Å resolution. The exposure time was 0.25 s, the crystal-to-detector distance was 259.0 mm, wavelength 0.9686 Å and the oscillation range per image was 0.5°. The diffraction image was collected on an ADSC Q315 CCD detector.

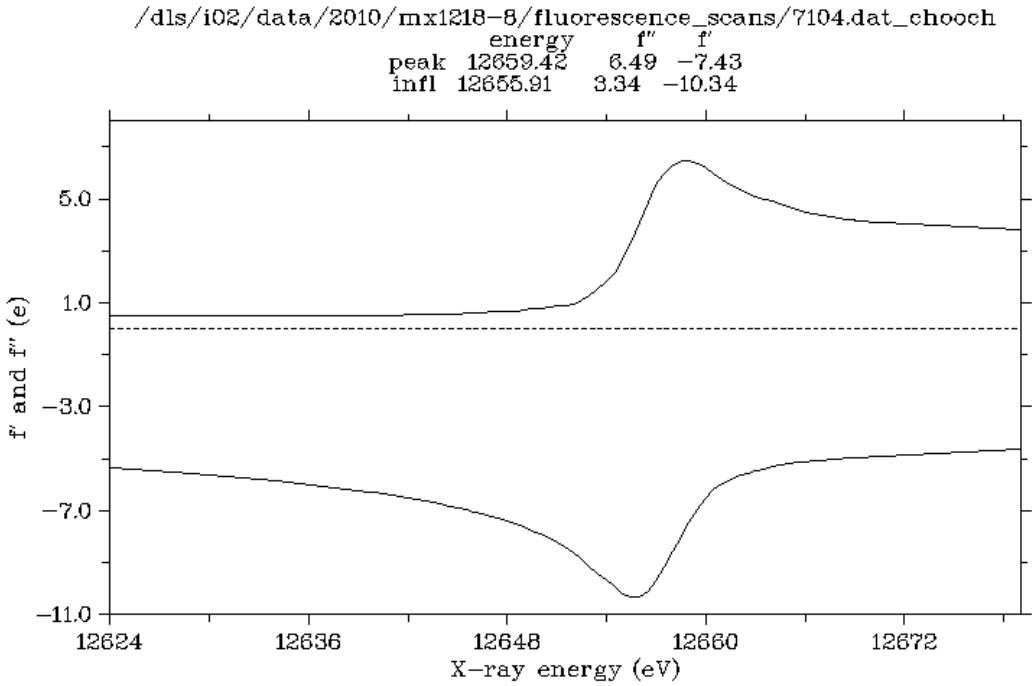


Figure 6.2: Fluorescence scan of anomalous scattering of SeMet NheA at three X-ray wavelengths of high remote 0.9784 Å, peak 0.9794 Å, and inflection 0.9797 Å. Taken at beams on the Diamond Light Source in Oxford, United Kingdom.

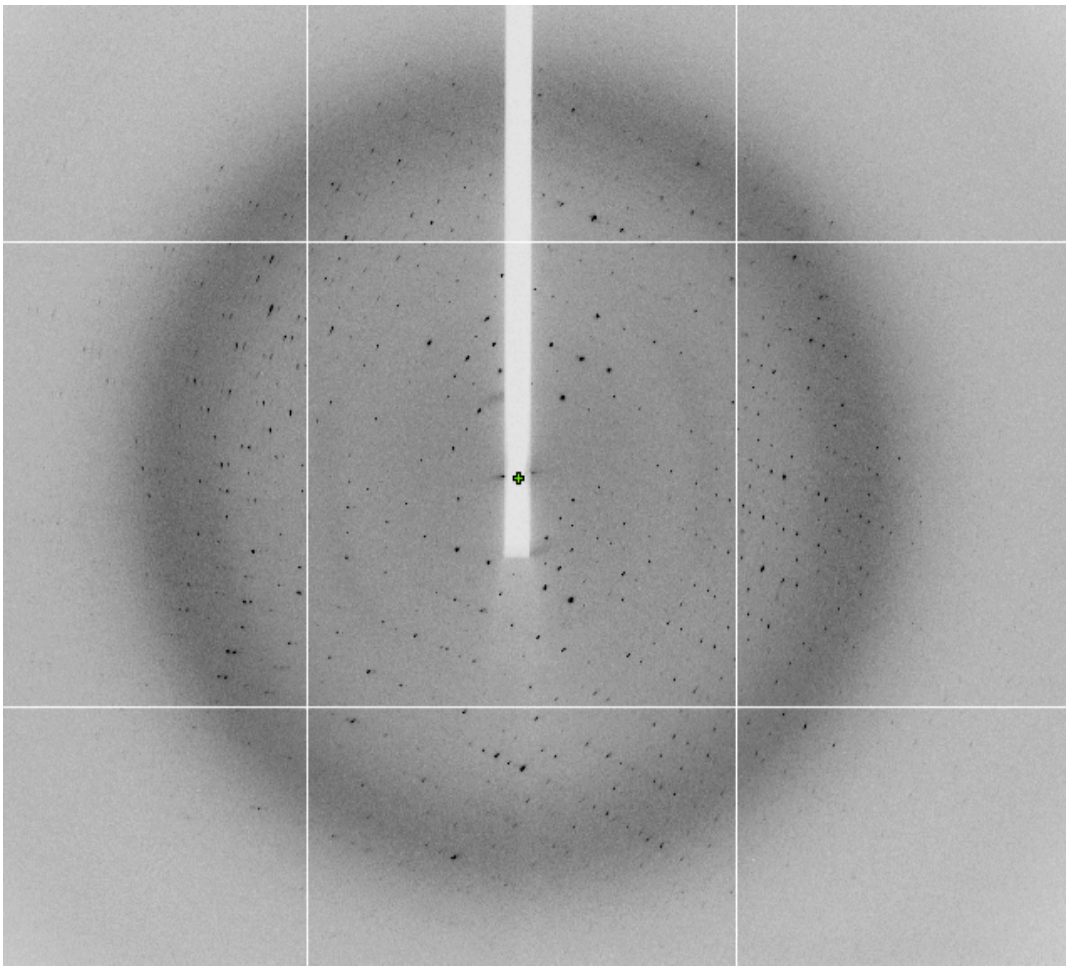


Figure 6.3: X-ray diffraction pattern from SeMet NheA crystal using beamline I02 of Diamond light source.
Diffraction intensities are visible to 2 Å. The exposure time was 0.75 s, the crystal-to-detector distance was 225.0 mm, Peak wavelength 0.9794 Å and the oscillation range per image was 0.5°. The diffraction image was collected on an ADSC Q315 CCD detector.

Table 6.1: The Final data collection and processing statistics of the NheA structure in the *C2* space groups generated.

	Native NheA	**SeMet NheA**		
Data collection statistics				
Beamline At the Diamond Light Source, Oxford, England	I03	I02		
Detectors At the Diamond synchrotron	ADSC Q315R CCD detector			
The crystal-to-detector distance (mm)	259.0	225.0		
Space group	*C2*			
Cell dimensions				
a, b, c (Å)	308.67, 58.17, 172.93	307.91, 58.72, 172.73		
α, β, γ (°)	90, 110.55, 90	90, 110.39, 90		
Crystal system	Monoclinic			
		Peak	Inflection	Remote
Wavelength (Å)	0.96860	0.9794	0.9797	0.9784
Resolution (Å)	50.2-2.05 (2.16-2.05)	29.8-2.05 (2.16-2.05)	29.9-2.09 (2.21-2.09)	29.9-2.19 (2.31-2.19)
No. of observations	504907 (74251)	672961 (96285)	629433 (89066)	547829 (76626)
No. of unique reflections	179016 (26237)	182607 (26191)	170815 (24262)	148979 (21012)
R_{merge}	0.077 (0.358)	0.114 (0.626)	0.10 (0.641)	0.106 (0.656)
I / σI	7.5 (2.5)	7 (2)	7.7 (2)	8.3 (2)
Completeness (%)	98.9 (99.7)	99.6 (98.5)	99.5 (97.6)	99.3 (96.5)
Mosaicity (°)	0.36	0.4		
Multiplicity	2.8 (2.8)	3.7 (3.7)	3.7 (3.7)	3.7 (3.6)
Number of molecules in the asymmetric unit	8			
Solvent content (%)	44.3			
Matthews coefficient ($Å^3$/Da)	2.21			

*Data in brackets are for highest resolution shell.
$R_{merge} = \Sigma hkl \Sigma i |I\ i - \langle I \rangle| / |\langle I \rangle|$, where I *i* is the intensity for the *i* th measurement of an equivalent reflection with indices *h,k,l*

6.3 Matthews number (*Vm*) calculation

The Matthews number calculation allows estimation of the number of molecules residing in the asymmetric unit (Matthews, 1968). It was calculated using the web page http://www.ruppweb.org/Mattprob/ (Kantardjieff & Rupp, 2003). This predicted, based on the unit cell volume and the molecular weight of NheA, that there were 12 possible solutions in the Vm range considered of which the most probable solutions are between 6 and 9 molecules in the asymmetric unit (Table 6.2). In order to examine the possible rotational pseudosymmetric relationships of NheA molecules in the asymmetric unit, a self-rotation function was run from 0° to 180° with a 5° intervals on κ using POLARRFN (CCP4: Supported Program), but there was no non-crystallographic pseudosymmetric rotational symmetry. A self-Patterson function was calculated with 50–6 Å, data which showed that there are two peaks, one with a height of 61% of the origin at (*u, v, w*) = (0.25, 0.25, 0.0), and another one with a height of 41% of the origin at (*u, v, w*) = (0.5, 0.0, 0.0) (Figures 6.4 A and B) in addition to the origin. These indicate that the asymmetric unit of NheA has molecules which are related by two different kinds of translational pseudo-symmetry. As result, it was suggested that the number of the monomers in the asymmetric unit of NheA is a multiple of four, and thus the strongest probability is 8 molecules in asymmetric unit with a solvent content of 44% and a *Vm* of 2.21 $Å^3$/Da (Table 6.2; and Figure 6.5 A and B). This value is in the middle of the range of those predicted for protein crystals (Matthews, 1968).

6.4 Structure determination of NheA

As described below, the NheA structure was solved by SeMet MAD and to solve the structure and to carry out phase improvement we used a combination of the SHELXC, D and E programs to locate heavy atoms (Sheldrick, 2010) and PHENIX to carry out phase improvement and automated model building (Adams et al, 2010).

Table 6.2: Output from the Matthews Probability Calculator at http://www.ruppweb.org/Mattprob/ (Kantardjieff & Rupp, 2003). This suggested that there were 6 to 9 molecules in the asymmetric unit with a V_m of 2.21 A^3/Da giving the crystals a solvent content of 44% if there were 8 molecules in the asymmetric unit.

N(mol)	Prob(N) for resolution	Prob(N) overall	Vm A**3/Da	Vs % solvent	Mw Da
1	0.0005	0.0016	17.68	93.04	41075.00
2	0.0005	0.0016	8.84	86.08	82150.00
3	0.0005	0.0022	5.89	79.12	123225.00
4	0.0024	0.0124	4.42	72.16	164300.00
5	0.0221	0.0622	3.54	65.21	205375.00
6	0.1078	0.1854	2.95	58.25	246450.00
7	0.2914	0.3357	2.53	51.29	287525.00
8	0.3877	0.3016	2.21	44.33	328600.00
9	0.1717	0.0879	1.96	37.37	369675.00
10	0.0143	0.0060	1.77	30.41	410750.00
11	0.0006	0.0017	1.61	23.45	451825.00
12	0.0005	0.0016	1.47	16.49	492900.00

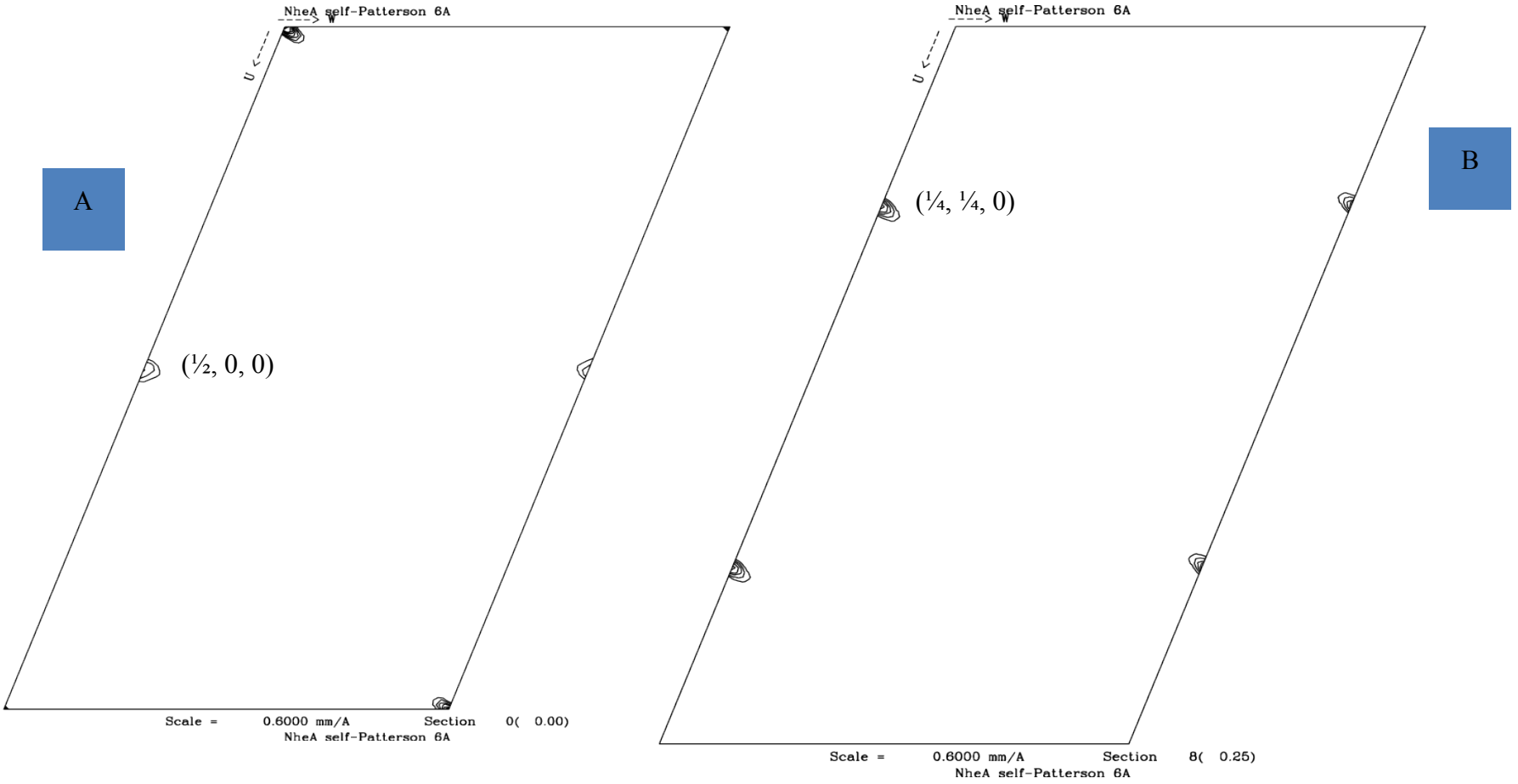


Figure 6.4: A self- Patterson function shows of NheA crystal.
There are two peaks, one at (*u, v, w*) = (0.5, 0.0, 0.0) (A), and another one at (*u, v, w*) = (0.25, 0.25, 0.0) (B) in addition to origin (0, 0, 0). These show that the asymmetric unit of NheA crystal contains molecules, which are related by two different kinds of translational pseudo-symmetry.

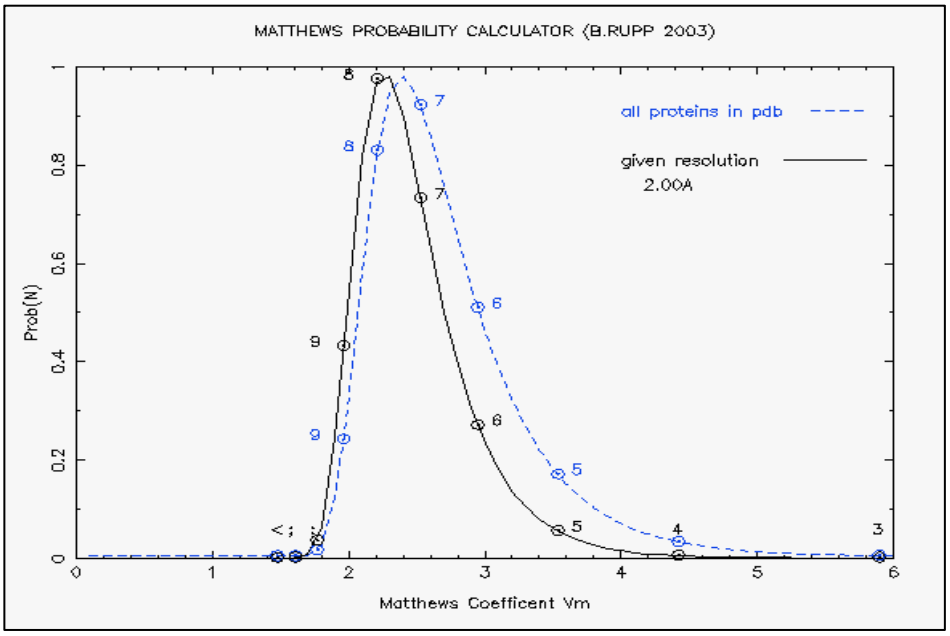


Figure 6.5 A: Output from the Matthews Probability Calculator at http://www.ruppweb.org/Mattprob/ (Kantardjieff & Rupp, 2003).
This suggested that there were 6 to 9 molecules in the asymmetric unit, with a V_m of 2.21 A^3/Da if there were 8 molecules in the asymmetric unit.

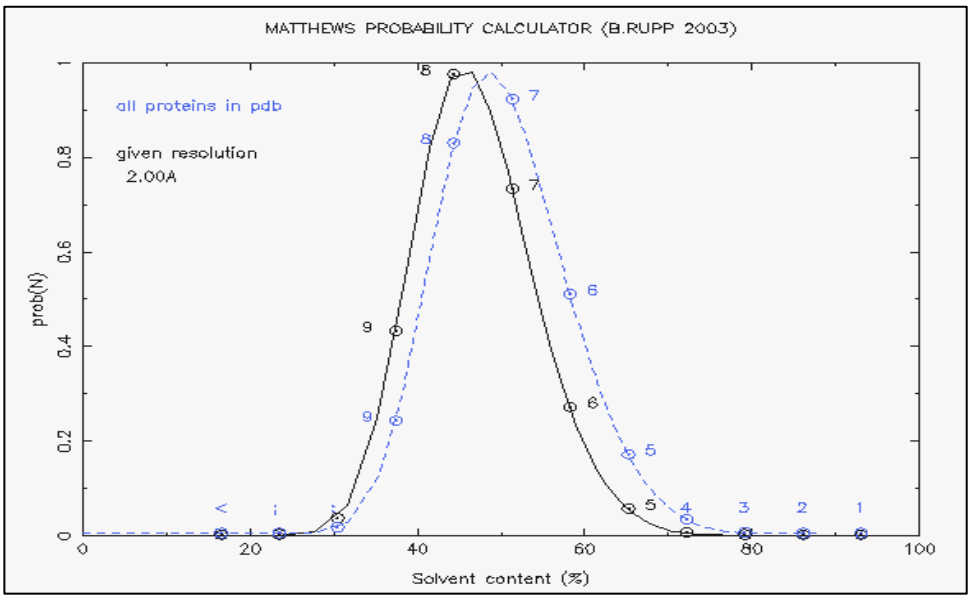


Figure 6.5 B: Output from the Matthews Probability Calculator at http://www.ruppweb.org/Mattprob/ (Kantardjieff & Rupp, 2003).
This suggested the crystals have solvent content of 44% if there are 8 molecules in the asymmetric unit.

Because SHELX cannot read the CCP4 *.mtz reflection file format the *.mtz file, which was produced from SCALA, was converted to *.sca file using SCALEPACK in CCP4 (CCP4, 1994). SHELXC was used to prepare files for SHELXD (Selenium atom location) and SHELXE (phasing and density modification). SHELXC does statistical analysis of the input data, estimates the heavy atom structure factors and prepares files for the other two programs. The SHELXD program solved the heavy atom substructure and located 58 selenium atoms using 1000 attempts and then generated a *.pdb file with the positions of the heavy atoms. SHELXE calculates phases and carried out density improvement over 100 cycles. However, in this case SHELXE could not produce an interpretable map. These were indication of solvent boundaries between elongated molecules and of presence of helical structure but detailed interpretation was impossible (Figure 6.6 A and Figure B). Therefore, the *.pdb file from SHELX with the *.sca files of anomalous data and native data, were used in PHENIX to carry out phase improvement and automated building of the NheA structure. As a result of this autobuild eight chains were built and 2754 residues were found.

Un-modelled regions of electron density were rebuilt using the program COOT (Emsley & Cowtan, 2004). Refinement of the model against the observed structure factors was performed using the program REFMAC (Murshudov et al, 1997). Between rounds of refinement the model was rebuilt using the program COOT (Emsley & Cowtan, 2004). The process of rebuilding and refinement was repeated 19 times until no structure improvement was possible.

The result of refinement was a model with a R work of 20% and a R free of 28% for 176339 unique reflections between 49.85 Å and 2.05 Å resolution. The root mean square deviation from the ideal for bond lengths is 0.02 Å and for bond angles is 1.7°. There are 1507 water molecules included in the final structure. The average B-factor overall is 35.4 $Å^2$ for the protein and 34.72 $Å^2$ for water. These refinement statistics indicated an acceptable and reasonable structure (Branden & Jones, 1990). The final refinement statistics for this structure are shown in Table 6.3.

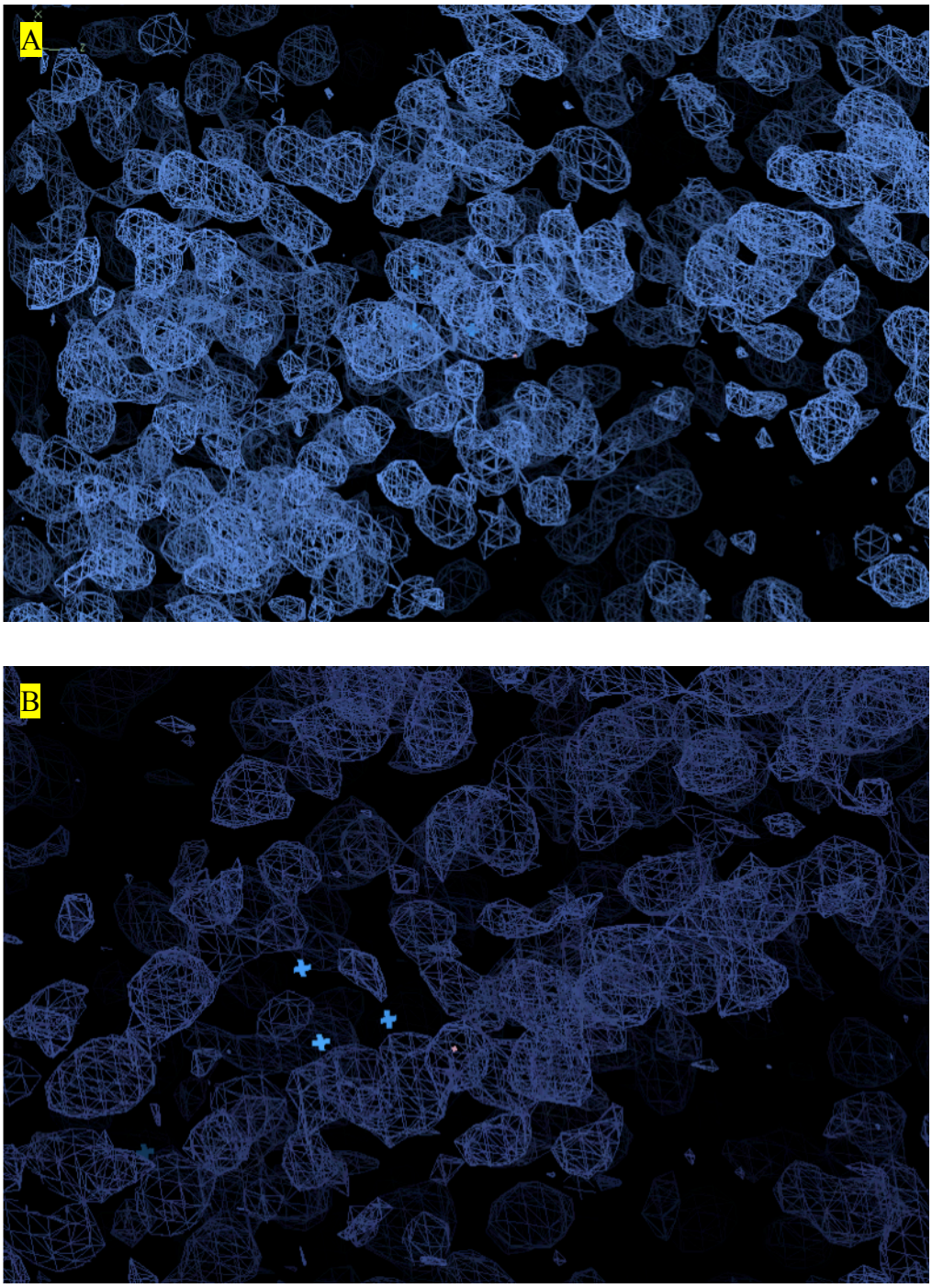


Figure 6.6: The map calculated with the initial phases obtained from SHELX for the NheA structure at 2.05 Å. The $(2F_{obs}-F_{calc})\exp(i\alpha_{calc})$ map is contoured at 1σ.
(A) Non- inverted solution.
(B) The map from the inverted hand.

Table 6.3: Final statistics for the refined structures of the NheA structure in the *C2* space groups generated by REFMAC (Murshudov et al, 1997)

Refinement statistics	
Resolution (Å)	2.05
No. Reflections	176339
R_{work} / R_{free}	0.20/0.28
No. Atoms	22683
No. Residues	2674
No. Waters	1507
Average B values (Å^2)	
Protein	35.4
Water	34.7
R.m.s deviations	
Bond lengths (Å)	0.02
Bond angles (°)	1.7
Ramachandran plot	
Most favored regions (%)	96.1
Additionally allowed regions (%)	3.4
Generously allowed regions (%)	0.4
Disallowed regions (%)	0

$R_{work} = \Sigma|F_{obs} - F_{calc}|/\Sigma|F_{obs}|$ where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively. R_{free} = as for R_{work}, but for 5% of the total reflections chosen at random and omitted from refinement.

6.5 3D crystal structure of NheA

The crystal structure of NheA confirmed that the protein has eight monomers of NheA in the asymmetric unit as expected from the Matthews number (Vm) calculation (Figure 6.7 and 6.8). These molecules were named chain A, chain B, chain C, chain D, chain E, chain F, chain G, and chain H. Electron density for all the eight molecules was well defined except for the residues in the N-terminal regions of all eight molecules in the asymmetric unite and in some turns and loop regions.

PDBePISA http://pdbe.org/PISA (Krissinel & Henrick, 2007) analysis was run in order to investigate if there was any evidence for oligomeric assemblies in NheA (Figure 6.8). The eight molecules all appear to be monomeric (this discussed in more details on section 6.7). The overall fold of each monomer within NheA can be viewed in Figure 6.9 A and 6.9 B. NheA is a long, elongated, rod shaped structure, with dimensions 95 Å x 40 Å x 20 Å. Because all eight molecules are essentially identical (see section 6.6) the description below considers only molecule A, unless otherwise stated. Monomer A has been chosen as it is better ordered than others.

To identify and analyze secondary structures in the NheA protein, the PROMOTIF program was used, which is available on http://www.ebi.ac.uk/pdbsum/ (Hutchinson & Thornton, 1996; Laskowski, 2009). The major secondary structure of NheA is α helix. Figure 6.9 A and Figure 6.9 B shows that the main body of NheA consists of a bundle of five helices: αA consists of 21 residues (32 Å long; SER 24- GLN 44); three long helices: αB1/ αB2 consists of 48 residues (62 Å long; LEU 56 – GLY 104) with a distinct kink at residues TYR 95 and TYR 96; αC consists of 57 residues (82 Å long; GLU 110 - LEU 166) with a distinct kink at residue LEU 141; αF consists of 49 residues (73 Å long; SER 326 –ASN 356) and αG at the C terminal which consists of 31 residues (44 Å long; SER 326-ASN 356). Helix αA, αB, αC and αF or αB, αC, αF, and αG are arranged in an anti-parallel bundle in a simple square, left-handed. Up–down topology with connecting loops or domains.

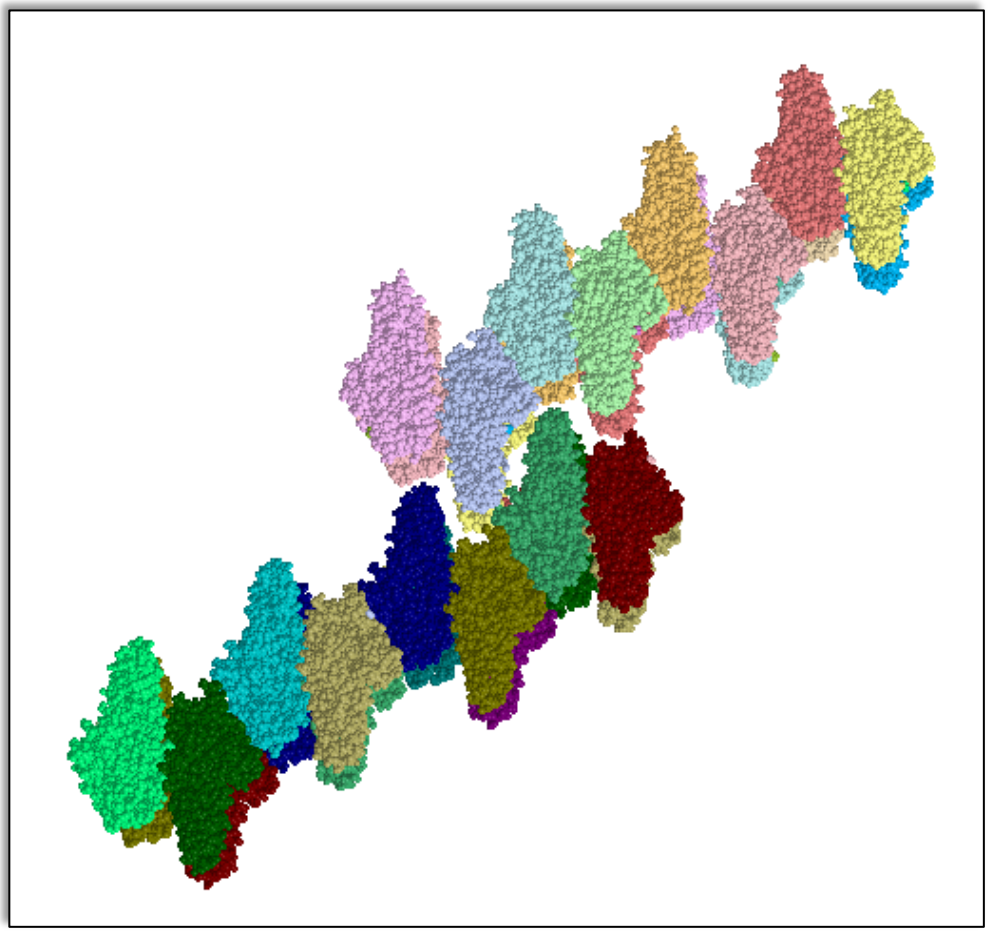


Figure 6.7: 32 NheA molecules in a unit cell (space group C2) of the crystal.
This figure was generated by PDBePISA http://pdbe.org/PISA (Krissinel & Henrick, 2007).

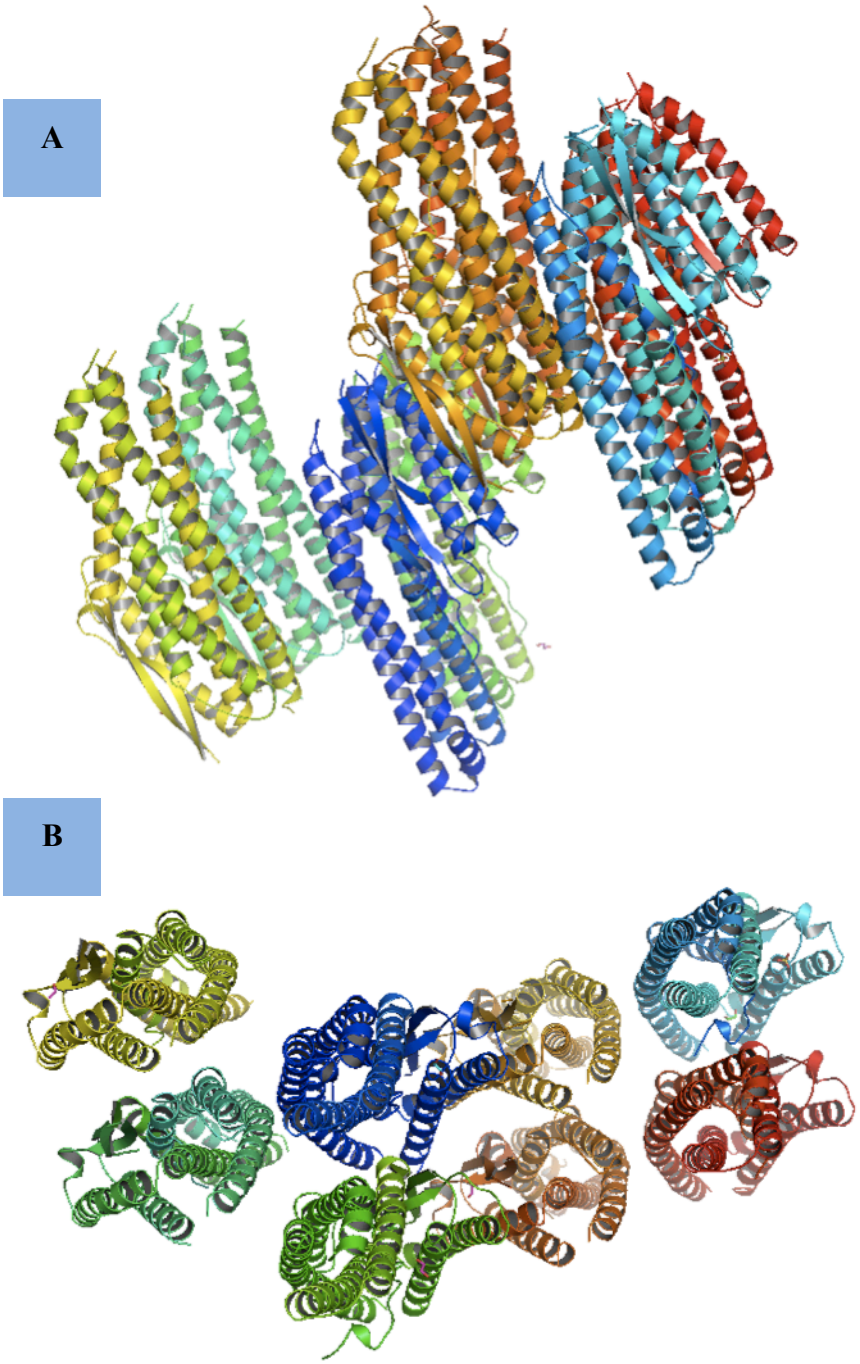


Figure 6.8: 8 NheA molecules in the asymmetric unit cells of the NheA crystal.
The figure (A) shows monomer A (deep blue), monomer B (cyan) monomer C (limegreen), monomer D (green), monomer E (yellow), monomer F (light orange), monomer G (orange), and monomer H (red). The figure (B) shows the top view of the eight monomers of NheA.
Produced using PyMol (DeLano and Lam, 2005).

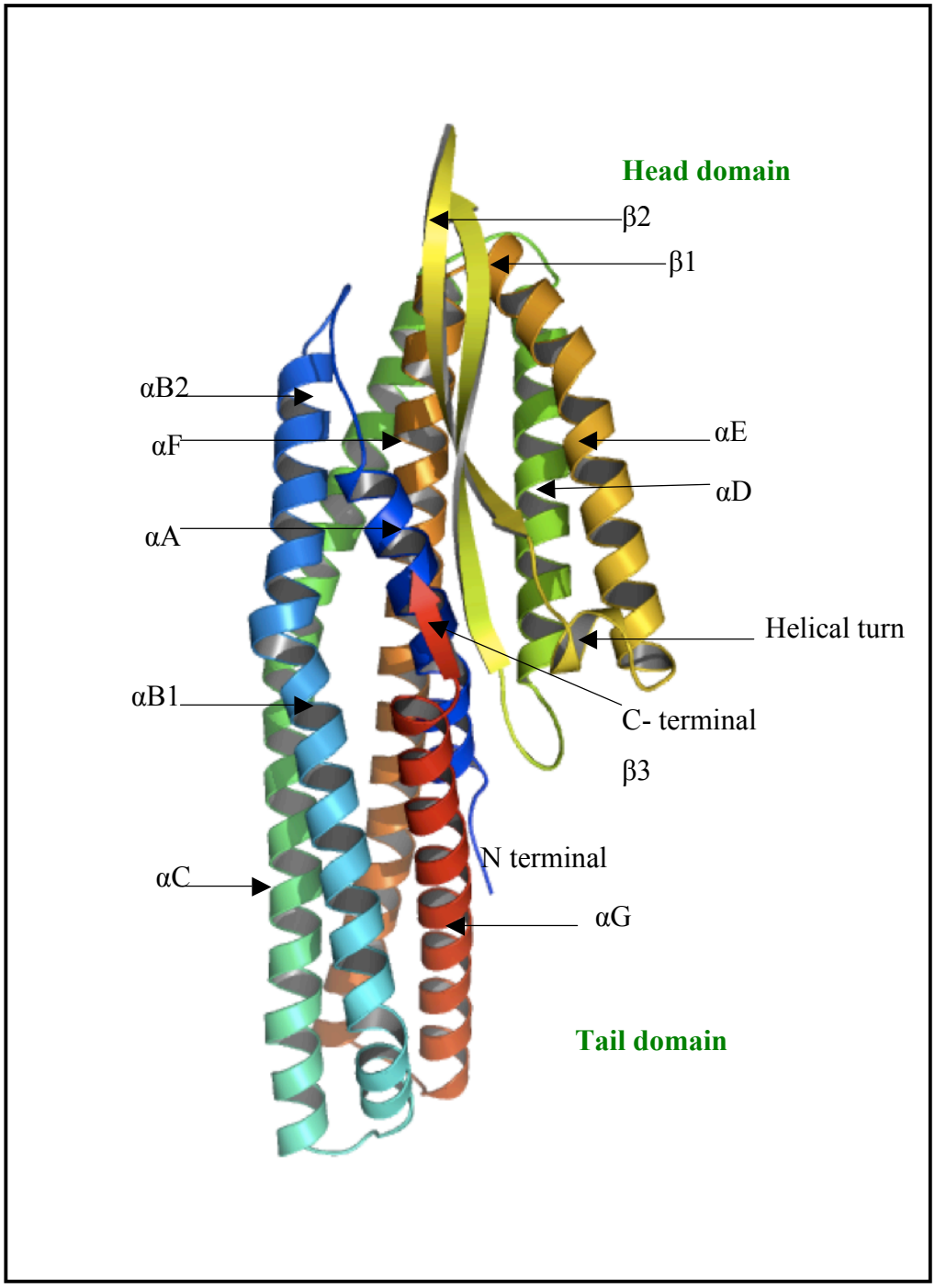


Figure 6.9 A: Ribbon representation of NheA monomer.
The NheA ribbon is coloured from blue to red, from the N to C terminus. The α-helical bundle has dimensions 95 Å x 40 Å x 20 Å. NheA has seven α-helices, labeled αA, αB, αC, αF, αG, αD, αE. The structure has two sub-domains the ‘‘Tail’’ domain, and the “Head” domain, which includes the “ß tongue” which consist of two strands ß1 and ß 2. N- and C-terminal are labelled.
This figure was prepared with PYMOL (DeLano and Lam, 2005).

Figure 6.9 B: Ribbon representation of NheA monomer showing turns in NheA.

NheA coloured according to secondary structure: α helix is red, and β sheet is yellow. The turn regions are coloured in green. The two ß strands coloured in yellow.

This figure was prepared with PYMOL (DeLano and Lam, 2005).

The molecule consists of two unusual topology sub-domains which had been first recognized in the 3D crystal structure of *E. coli* K12 ClyA (Wallace et al, 2000) and then in HblB (Madegowda et al, 2008) but do not occur elsewhere in the Protein Data Bank:

1) **The tail domain,** at the C terminal is made up of the fifth helical bundle, αG. Helix αG packs between helices αA and αB, and thus forms a five -helix bundle for a short region of the structure. Here, it should be pointed out that α A in NheA structure is short, therefore it only a minor part of tail domain. However, in the ClyA molecule (Wallace et al, 2000) αA is much longer and the 5-helical bundle (αA, αB, αC, αF, and αG) occupies the lower half of the structure (Figure 1. 6). The C terminal ends with a short ß strand, ß3, which is formed from 4 residues (VAL 357, GLU 358, VAL 359, HIS 360).

2) **The head domain,** which includes the "ß tongue", is located between the 3rd and 4th (αC and αF) helices of the main body and includes two long α-helices: αD consists of 24 residues (35 Å long; ASP 172-LEU 195), αE consists of 27 residues (38 Å long; SER 243-THR 269), and a very short helical turn (ASN 235-VAL 238). This part also contains two ß strands: ß strand 1 that consists of 13 residues (40 Å long; GLY 204- THR 216) and ß strand 2 that consists of 12 residues (26 Å long; THR 222- LEU 233). These form a 3-stranded antiparallel ß-sheet with ß3 from the tail domain (see above).

The major secondary structures of NheA are connected by regions of random coil and by turn regions that sometimes contain additional short helical segments, giving a structure that over all is 75.8% helix, and 8.4% beta strand as classified by the PROMOTIF program (Hutchinson & Thornton, 1996). The loop structures contain well-characterized structural motifs such as beta turns and gamma turns and there are some residues adopting random coli conformations (Table 6.4). There are 11 β turns, and one inverse γ turn (Table 6.4).

The PROMOTIF program indicates two beta bulges in NheA structure as follows:

1) There is an antiparallel classic β bulge (GLY 231, THR 232) in β2 strand opposite ASN 207 in β1 strand. The classic β bulge involves extra residues on one of β strand, which cause disruption of the regular hydrogen bonding pattern.

2) There is an antiparallel β bulge type G1 (ALA 220, GLN 221) before β2 strand. G1 type β bulges involve the presence of one residue in the left-handed α conformation and are mostly found at the ends of β strands.

Hairpin structures have been also identified and consist of two antiparallel β strands, β1 (GLY 204-THR 216) and β2 (THR 222- LEU 233) and they hydrogen bonded together. The hairpin in the NheA structure is classified as hairpin class 3:5, this class is characterized by presences of type a l β turn followed by a G1 β bulge.

Table 6.4: The secondary structure elements in *B. cereus* NheA at 2.05 Å resolution. Output from the PDBsum http://www.ebi.ac.uk/pdbsum/ (Hutchinson & Thornton, 1996; Krissinel & Henrick, 2007; Laskowski, 2009) (Hutchinson & Thornton, 1996; Krissinel & Henrick, 2007; Laskowski, 2009).

Start - End	Secondary structure	Number of Residue
ASN 19 – SER 22	Type l β turn	4
LEU 21- SER 24	Type ll β turn	4
SER 24- GLN 44	αA helix	21
VAL 50- MET 53	Type l β turn	4
ASN 51 –SER 54	Type VIIl β turn	4
MET 53- LEU 56	Type l β turn	4
LEU 56 – GLY 104	αB1/ αB2 helix	48
ASN 105- GLU108	Type l β turn	4
ILE 106- ASP 109	Type lV β turn	4
GLU 110-LEU 166	αC helix	57
GLN 167- SER 170	Type lV β turn	4
ASP 172-LEU 195	αD helix	24
GLN 199 ILE 202	Type lV β turn	4
GLY 204- THR 216	β1 strand	13
ASN 217-ALA 220	Type l β turn	4
THR 222- LEU233	β2 strand	12
ASN 235- VAL 238	Short helical	4
ALA 240- SER 243	Type lV β turn	4
SER 243- THR 269	αE helix	27
ALA 271- GLU 319	αF helix	49
THR 322- THR 324	Inverse γ turn	3
SER 326 –ASN 356	αG helix	31
VAL 357 – HIS 360	β3 strand	4

6.6 Structure comparison of molecules of NheA

Most of the residues in the molecule are well ordered and visible in the electron density (Figure 6.10, 6.11 and 6.12), with the exception of some the N terminal and turn regions in all eight monomers of NheA, where the experimental electron density was poorly defined (Figure 6.13 and 6.14). This poor density is interpreted as resulting from the inherent flexibility of these turns or from disorder in the crystal. Several N-terminal residues are missing from electron density maps due to their flexible nature. Thus, these poorly defined residues have been truncated in the model for improved accuracy as they are not well represented in the electron density as follows:

- In monomer A: the 15 N-terminal residues (LYS 1- VAL 15) and the residues in turn (ASN217-GLN221), and SER 170

- Monomer B: the 6 N-terminal residues (LYS 1- GLU 6) and the residues in turns (GLY 168 –SER 170), (ASN 217-THR 222);

- Monomer C: the 18 N-terminal residues (LYS 1- PRO 18) and the residues in turns (GLU108- SER 112), (GLY 168-SER 170), (GLN 218-GLN 221);

- Monomer D: the 5 N-terminal residues (LYS 1- THR 5) and the residues in turns ASP 109, (168 GLY-GLY 171), (GLN 218- GLN 221);

- Monomer E: the 18 N-terminal residues (LYS 1- PRO 18) and the residues in turns (GLU 108-GLU 110), (SER 169 -GLY 171), (ASN 217-THR 222), (GLU 320-THR 324);

- Monomer F: the 19 N-terminal residues (LYS 1- ASN 19) and the residues in turns (GLU 108-SER 112), (GLY 168-SER 170), (GLN 218-GLN 221), (VAL 318-GLY 321);

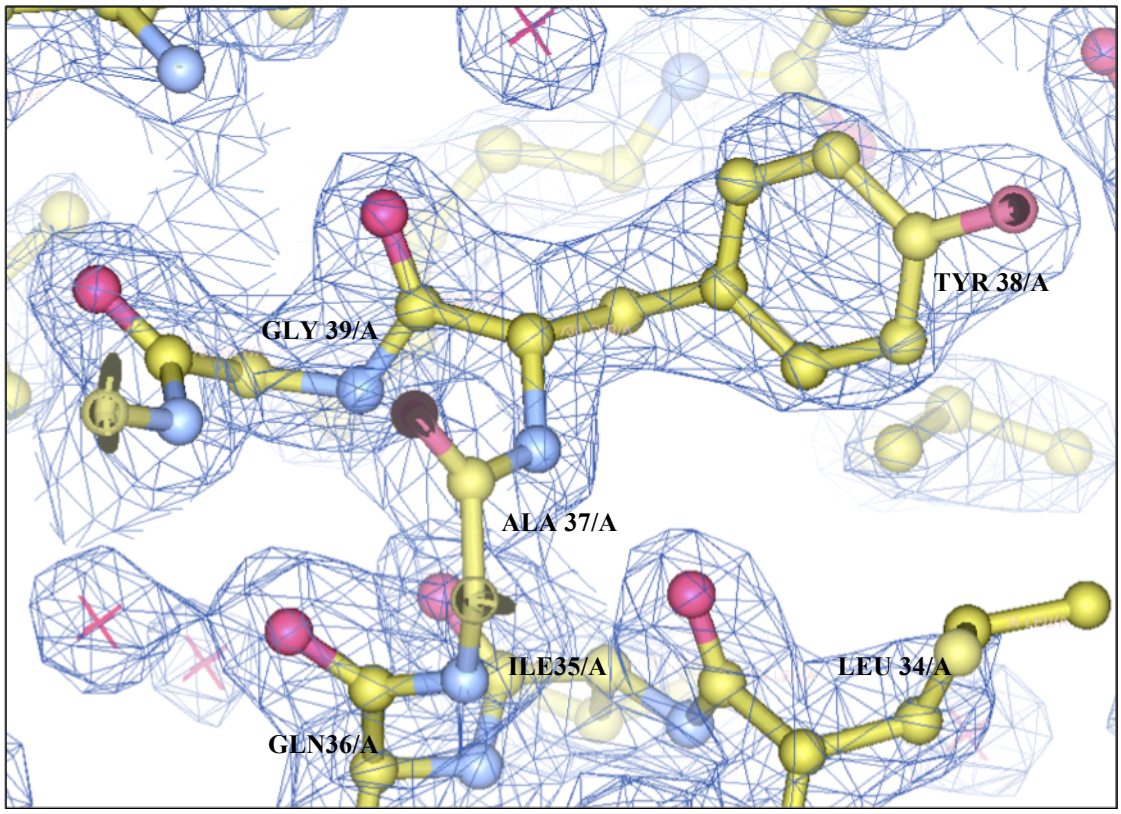


Figure 6.10: Example of the final electron density observed for the NheA structure (monomer A) at 2.05 Å. The $(2F_{obs}-F_{calc})\exp(i\alpha_{calc})$ map is contoured at 1σ and shown in blue. The protein is shown in ball and stick representation with yellow Carbon, blue Nitrogen, and red Oxygen atoms. X symbol represents water molecules.

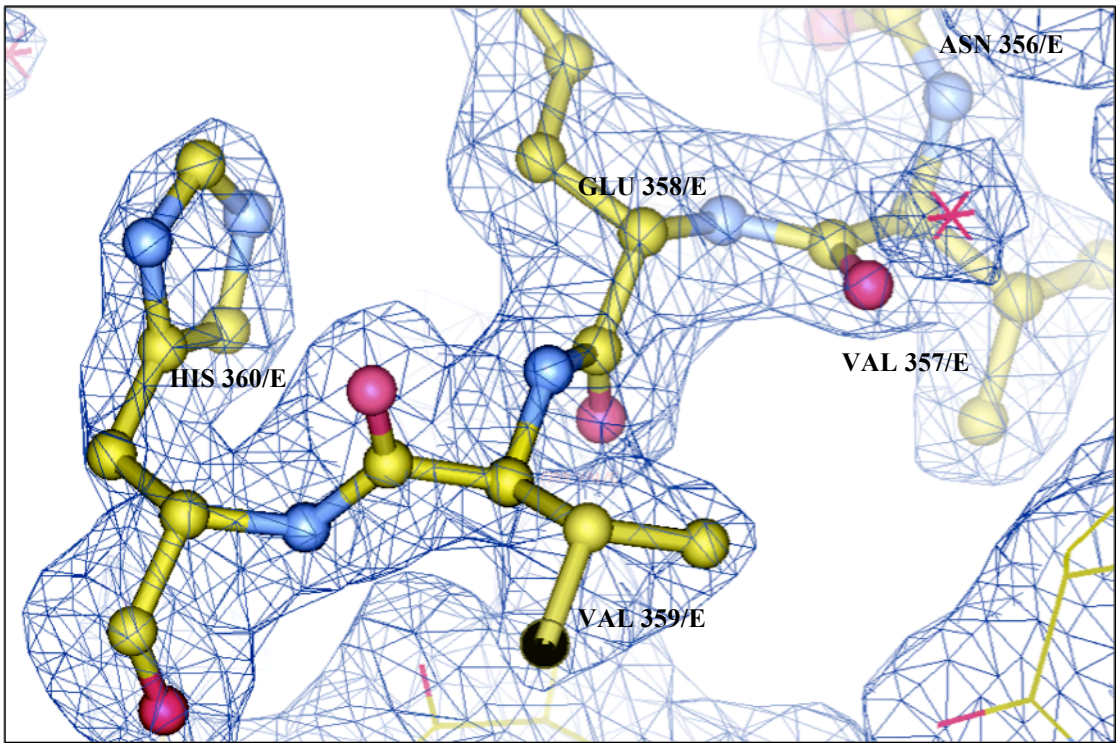


Figure 6.11: The final electron density observed at the C terminal for the NheA structure (monomer E) at 2.05 Å. The $(2F_{obs}-F_{calc})\exp(i\alpha_{calc})$ map is contoured at 1σ and shown in blue. The protein is shown in ball and stick representation with yellow Carbon, blue Nitrogen, and red Oxygen atoms. X symbol represents water molecules.

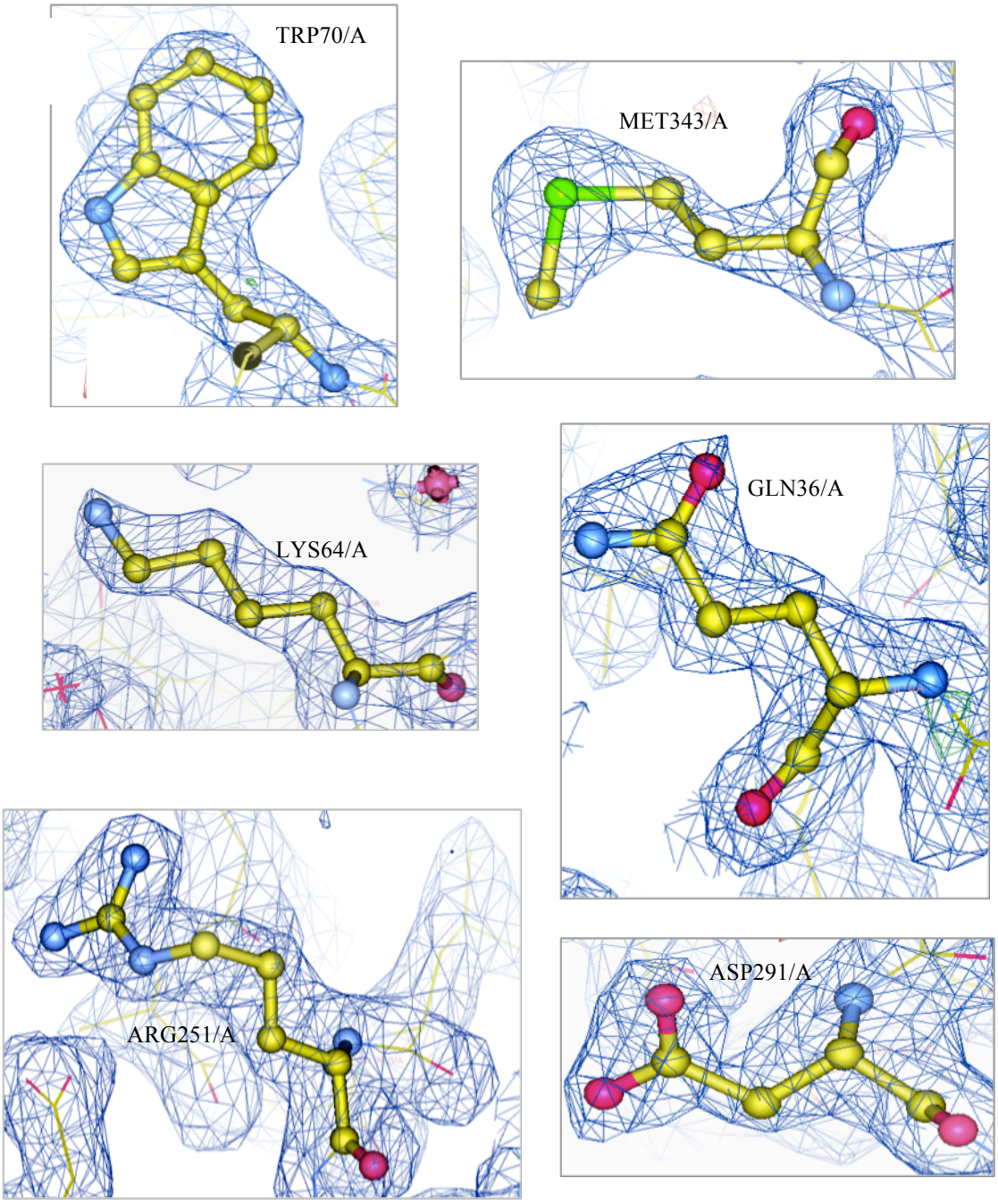


Figure 6.12: Examples of the final electron density for some side chains in the NheA structure at 2.05 Å.
The $(2F_{obs}-F_{calc})\exp(i\alpha_{calc})$ map is contoured at 1σ and shown in blue. The protein is shown in ball and stick representation with yellow Carbon, blue Nitrogen, red Oxygen atoms, and green Sulfur atom. X symbol represents water molecules.

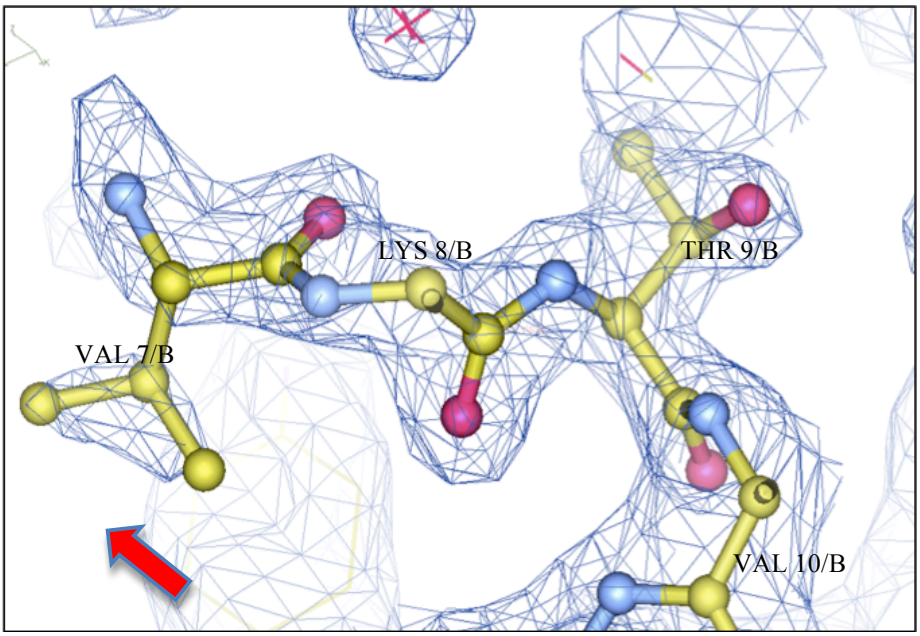


Figure 6.13: The final electron density observed at N terminal for the NheA structure (monomer B) at 2.05 Å.
The $(2F_{obs}-F_{calc})\exp(i\alpha_{calc})$ map is contoured at 1σ and shown in blue. The protein is shown in ball and stick representation with yellow Carbon, blue Nitrogen, and red Oxygen atoms. X symbol represents water molecules. The arrow shows the electron density was lacking for the N-terminal therefore amino acids 1-6 were un-modeled.

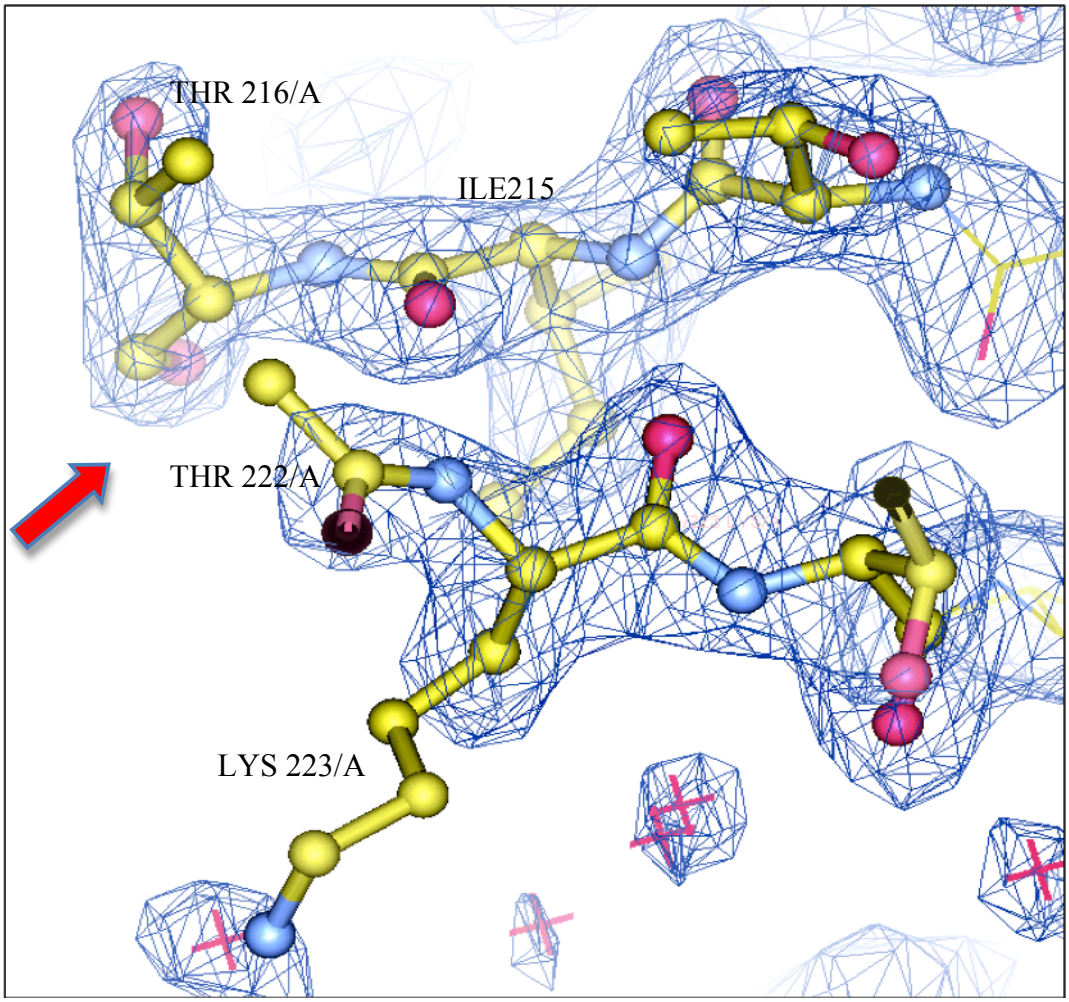


Figure 6.14: The final electron density observed at type l β turn regions for the NheA (monomer A) structure for loop regions at 2.05 Å.
The $(2F_{obs}-F_{calc})\exp(i\alpha_{calc})$ map is contoured at 1σ and shown in blue. The protein is shown in ball and stick representation with yellow Carbon, blue Nitrogen, and red Oxygen atoms. X symbol represents the water molecules. The arrow shows the electron density in this region was poor thus amino acids from 217 to 221 were un-modeled.

- Monomer G: the 18 N-terminal residues (LYS 1- PRO 18) and the residues in turns (ASP 109 - GLU 110), (GLY 168 - GLY 171), (ASN 217 - THR 222), (GLY 321- ASP 325);

- Monomer H: the 14 N-terminal residues (LYS 1- ASN 14) and the residues in turns (168 GLY- GLY 171), (ASN 217- THR 222).

To investigate whether there are significant differences between the eight subunits in the NheA crystal structure, monomer A was superposed onto monomers B, C, D, E, F, G, and H. Monomer A was chosen for comparison because overall it is the best-ordered monomer and it is the most complete one in the model of the structure.

Each Chain of monomer A was compared structurally to each chain of monomers B, C, D, E, F, and G using LSQKAB (CCP4, 1994) for comparing protein subunits in 3D. This was done over the amino acid range 1-360, and resulted in the overall root mean-square deviations (RMSDs), listed in Table 6.5. The superpositions are shown in Figure 6.15 A and B. Each chain aligns very well onto chain A with an RMSD of 0.3 -0.4 Å and no indication of any significant differences between monomers, except for individual residues in loop region (Figure 6.16). The highest deviation value of RMSD was between THR 222 in chain A and chain B which was about 6.6 Å, but it is important to consider that this residue is located in the psi loop regions which is characterized by poor density, and therefore there may be differing interpretations during model building.

Table 6.5: Output from LSQKAB (CCP4: Supported Program (CCP4, 1994) showing the average overall RMSD for molecules superposed on chain A.

	RMSD	Maximum RMSD
Chain A on chain B	0.3Å between 346 α carbon atom pairs	6.563 Å (THR 222)
Chain A on chain C	0.4 Å between 323 α carbon atom pairs	1.364 Å (THR 222)
Chain A on chain D	0.3 Å between 345 α carbon atom pairs	1.355 Å (THR 222)
Chain A on chain E	0.4 Å between 322 α carbon atom pairs	1.731 Å (ASP 325)
Chain A on chain F	0.3 Å between 323 α carbon atom pairs	1.340 Å (THR 222)
Chain A on chain G	0.4 Å between 323 α carbon atom pairs	1.935 Å (GLU 108)
Chain A on chain H	0.3 Å between 331 α carbon atom pairs	1.117 Å (TYR 323)

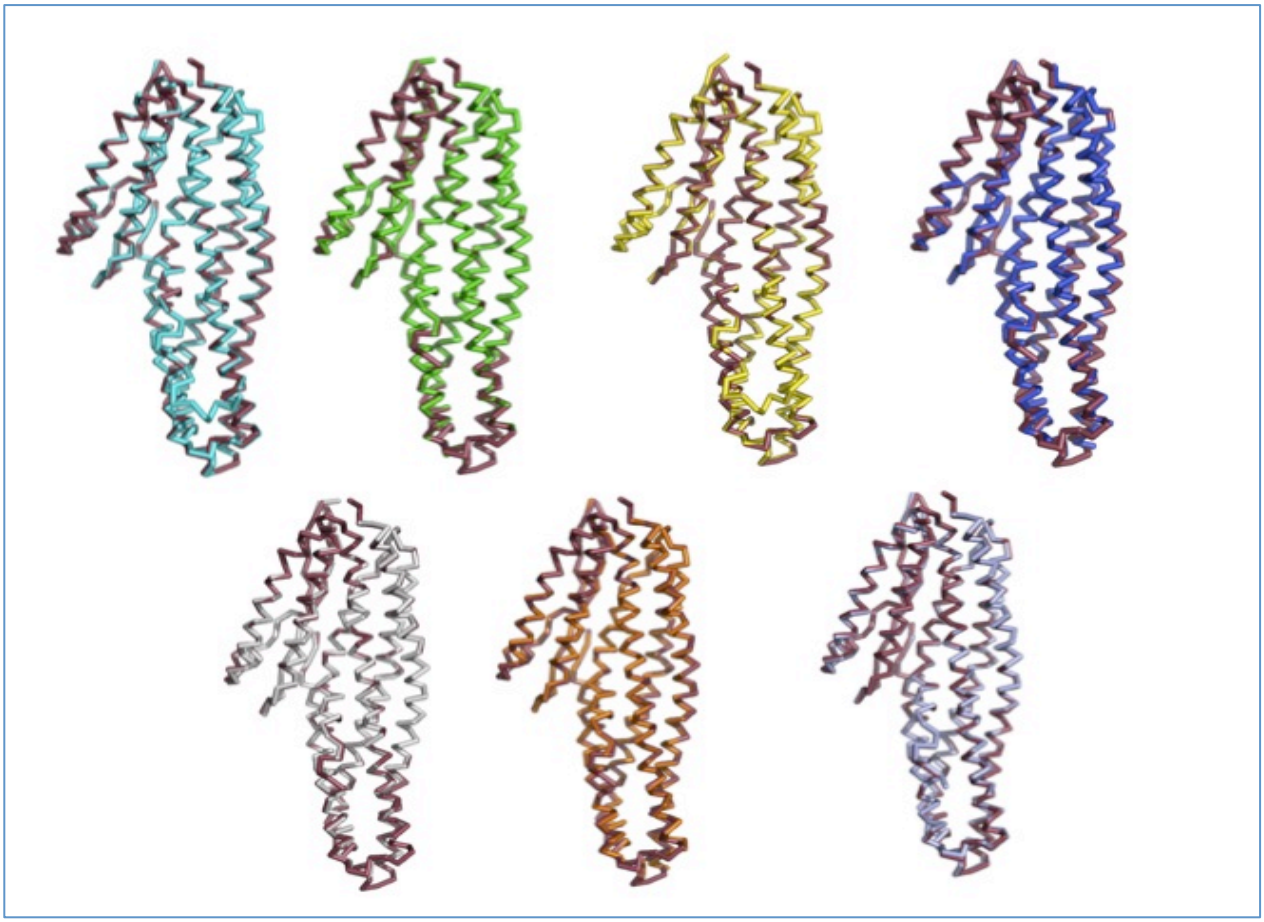


Figure 6.15 A: Superposition of the Cαs of monomer A (red) onto monomers B (cyan), C (green), D (yellow), E (blue), F (grey), G (orange), and H (light blue) represented as protein ribbons.
It is clear that the overall folds of all monomers are very similar with low deviation. Produced using PyMol (DeLano and Lam, 2005).

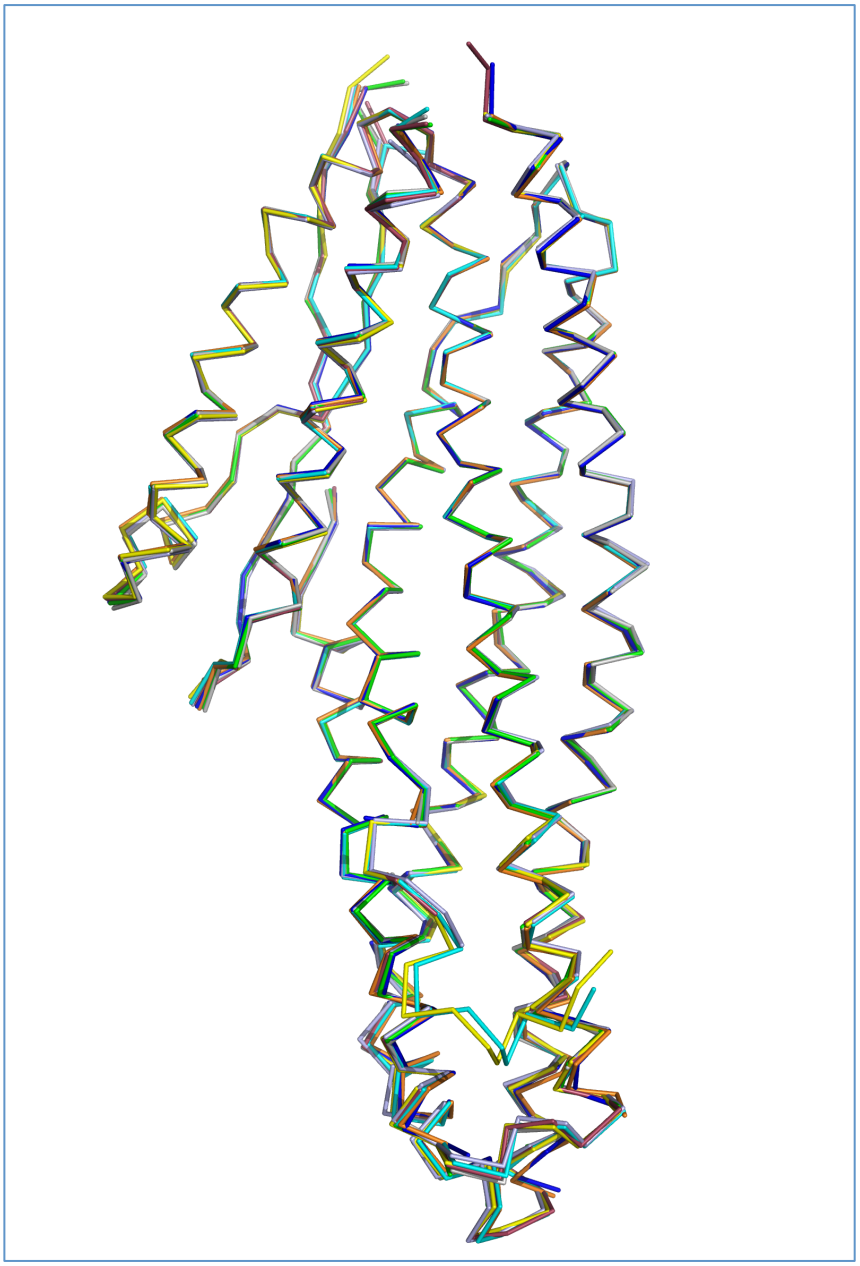


Figure 6.15 B: Superposition of the Cαs of monomer A (red) onto all eight molecules, **monomers B (cyan), C (green), D (yellow), E (blue), F (grey), G (orange), and H (light blue) represented as protein ribbons**.
It is clear that the overall folds of all monomers are almost identical with low deviations. Produced using PyMol (DeLano and Lam, 2005).

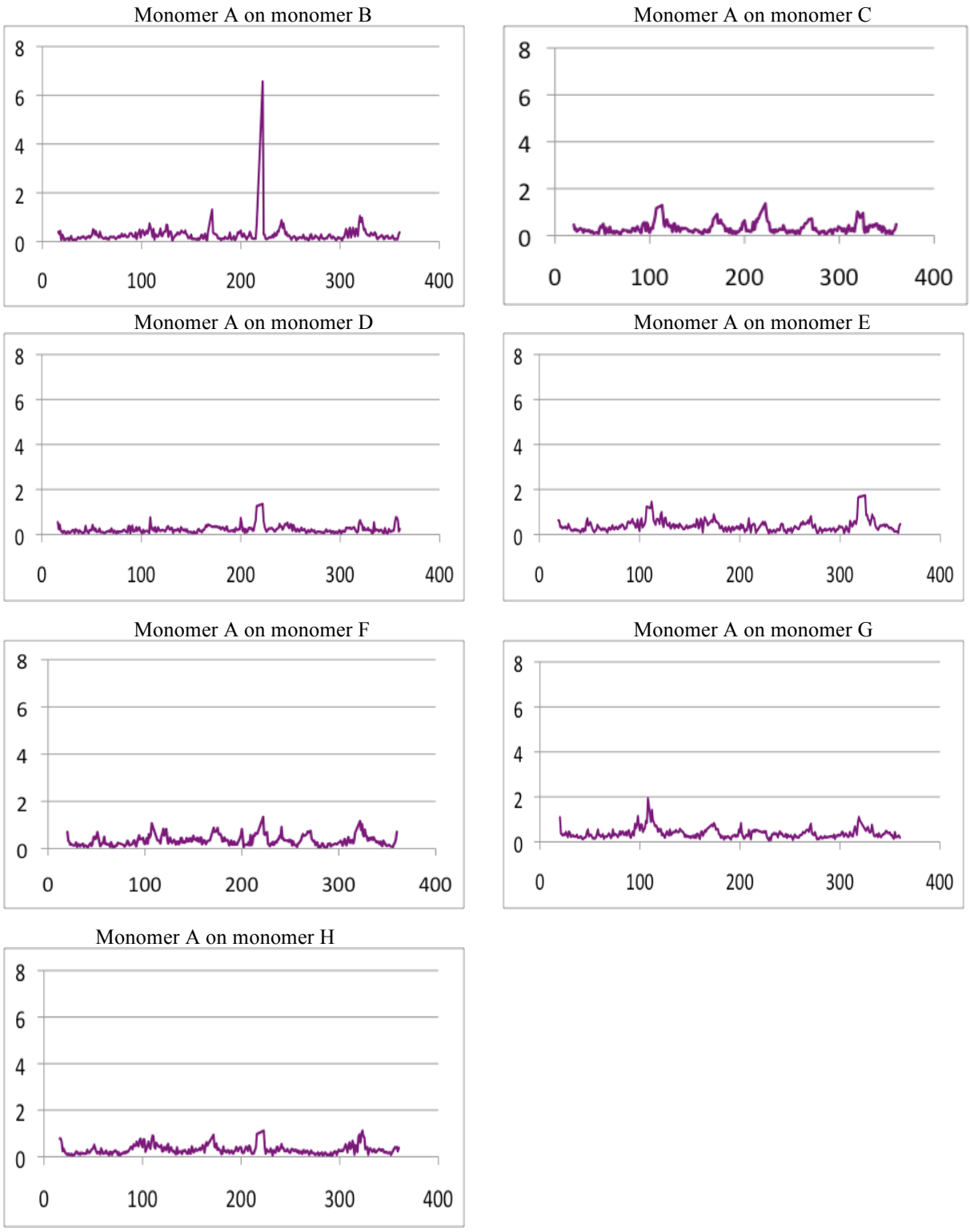


Figure 6.16: The RMSD in Å (Y- axis) between the Cαs of residues of monomer A compared to monomers B, C, D, E, F, G, and H plotted against residue number (X-axis).
It clear that there are no really significant differences in RMSD. The maximum RMSD value is seen between monomer A and monomer B (6.563, THR 222), which is located in a disordered part of the structure.

References

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66:** 213-221

Branden CI, Jones TA (1990) Between objectivity and subjectivity. *Nature* **343:** 687-689

CCP4 (1994) The CCP4 suite- programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50:** 760–763

DeLano WL, Lam JW (2005) PyMOL: A communications tool for computational models. *Abstr Pap Am Chem Soc* **230:** 254-COMP

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta crystallographica* **60:** 2126-2132

Evans G, Pettifer RF (2001) CHOOCH a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *J Appl Crystallogr* **34:** 82-86

Hutchinson EG, Thornton JM (1996) PROMOTIF - A program to identify and analyze structural motifs in proteins. *Protein Sci* **5:** 212-220

Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66:** 125-132

Kantardjieff KA, Rupp B (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12:** 1865-1871

Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* **372:** 774-797

Laskowski RA (2009) PDBsum new things. *Nucleic Acids Res* **37:** 355-359

Leslie AGW (1994) MOSFLM. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26**

Madegowda M, Eswaramoorthy S, Burley SK, Swaminathan S (2008) X-ray crystal structure of the B component of hemolysin bl from *Bacillus cereus*. *Proteins-Structure Function and Bioinformatics* **71:** 534-540

Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* **33:** 491–497

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53:** 240-255

Sheldrick GM (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr* **66:** 479-485

Wallace AJ, Stillman TJ, Atkins A, Jamieson SJ, Bullough PA, Green J, Artymiuk PJ (2000) *E. coli* hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy. *Cell* **100:** 265-276

Wlodawer A, Minor W, Dauter Z, Jaskolski M (2008) Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J* **275:** 1-21

Chapter 7: NheA Structure Analysis

This chapter will describe the analysis of the structure of the NheA protein.

7.1 NheA interfaces, surfaces and assemblies

In crystals, protein molecules interact together and the molecules arrange in orderly lattices to form crystals. It is possible to examine the interactions between molecules in the crystals and estimate their binding energy, solvation energy, interface area, hydrophobic interactions, hydrogen bonds and salt bridges across the interface. If the crystal contacts are large they may indicate possible biological assemblies between NheA monomers.

To investigate the crystal contacts and interfaces in NheA, as well to investigate the size of the crystal interfaces, and the energetic factors that keep the NheA molecules together and to see is there any evidence for oligomeric states in NheA, PDBePISA was used. This is a web tool for the analysis of macromolecular interactions, interfaces, surfaces and assemblies http://pdbe.org/PISA (Krissinel & Henrick, 2007). The calculation of protein assemblies from protein crystallography data on PDBePISA to identify significant interfaces is based on the free energy of protein binding which includes the free energy of binding and the entropy change. Although obviously the free energy consideration is paramount, the difficulty comes in accessing the factors that contribute to this -for example if the area of displaced solvent is small then the interface is likely to be unstable. Krissinel and Henrick (2005) point out that when the complex dissociates, any hydrogen bonding sites are likely to be satisfied by new (different) hydrogen bonds with the solvent (Krissinel & Henrick, 2005). Analysis of the NheA interfaces by PDBePISA did not reveal any specific interactions that could result in the formation of stable oligomers of NheA, and the eight molecules all seem to be in the monomer form. The results of PISA are shown in Table 7.1, which gives the details of each interface found between NheA

molecules in the crystal. Because there are 8 NheA molecules in the asymmetric unit related by pseudosymmetry many pairwise interactions are essentially the same. There are four types of interaction, which can be seen between NheA subunits:

1) The two-fold symmetric back-to-back interface: this is the strongest type of interface, e.g. between chain H and chain G, where 9 hydrogen bonds (Figure 7.1) and 8 salt bridges are formed. (Table 7.2 and 7.3).

2) The two-fold symmetric head to head interface, such as that between chain C and chain B, in which 4 hydrogen bonds (Table 7.2 and Figure 7.2). However, no salt bridges are formed.

3) The two-fold symmetric head to tail interface, such as that between chain A and chain D, in which 10 hydrogen bonds (Table 7.2 and Figure 7.3) and no salt bridges are formed.

4) The two-fold symmetric side by side interface, such as that between chain D and chain A, in which 4 hydrogen bonds and 3 salt bridges are formed (Table 7.2 and Figure 7.4).

PDBePISA indicates that none of the interactions are sufficiently extensive to suggest NheA is a dimer or oligomer. This is confirmed by gel filtration (see figure 5.5), which shows that NheA runs as a monomer. In contrast, the soluble form *o*f *E. coli* ClyA exists in the crystal as a dimer where two monomers of ClyA pack together to form a 2-fold symmetric dimer. The two subunits interact together to form 10 hydrogen bonds and two salt bridges (Figure 7.5). Also these two ClyA monomers orientate in a "head to tail" fashion that includes the two major hydrophobic surfaces of each monomer, and buries a significant area of the molecule away from the solvent (Wallace et al, 2000). Our PISA analysis of *B. cereus* HlbB (Madegowda et al, 2008) indicated that the HlbB interfaces do not reveal any specific interactions that could result in the formation of stable oligomers

Table 7.1: presents a list of all unique interfaces found in the NheA crystal with the number of hydrogen bonds and salt bridges. $^{i}N_{at}$ is the number of interfacing atoms, $^{i}N_{res}$ is the number of interfacing residues, $\Delta^{i}G$ is the solvation free energy gain on formation of the interface, in kcal/M, N_{HB} is the number of potential hydrogen bonds across the interface, N_{SB} is the number of potential salt bridges across the interface, N_{DS} is the number of potential disulfide bonds across the interface, **CSS** Sands Significance Score, which shows how important for assembly formation the interface is (CSS ranges from 0 to 1 as interface relevance to complex formation increases). This table is Output from the PDBePISA http://pdbe.org/PISA (Krissinel & Henrick, 2007)

	Structure 1			×	Structure 2			interface area, Å²	ΔiG kcal/mol	ΔiG P-value	NHB	NSB	NDS	CSS
	Range	iNat	iNres		Range	iNat	iNres							
Two fold head-to head interface	H	78	29	◊	B	76	22	723.4	-4.4	0.363	3	1	0	0.000
	C	79	19	◊	B	69	20	711.0	-8.8	0.113	6	0	0	0.000
	F	70	18	◊	A	73	20	679.1	-8.3	0.130	4	0	0	0.000
	G	73	19	◊	D	73	19	667.1	-7.4	0.185	4	0	0	0.000
	E	73	18	◊	H	72	18	648.8	-7.8	0.152	6	0	0	0.000
							Average:	676.5	-8.1	0.145	5	0	0	0.000
Two fold back to back interface	H	68	20	◊	G	64	20	691.2	-5.7	0.237	11	8	0	0.100
	F	64	20	◊	B	74	22	686.5	-6.6	0.201	8	7	0	0.100
	E	69	20	◊	D	66	22	670.2	-8.0	0.144	9	7	0	0.100
	C	65	19	◊	A	64	20	662.7	-7.0	0.164	10	6	0	0.100
							Average:	677.6	-6.8	0.186	10	7	0	0.100
Two fold side-by side interface	D	69	26	◊	A	61	24	584.2	-0.0	0.688	7	3	0	0.000
	G	61	23	◊	G	61	23	554.9	-1.5	0.574	4	2	0	0.000
	E	55	19	◊	C	55	20	470.5	-1.6	0.501	1	3	0	0.000
	F	45	17	◊	F	44	17	344.2	-2.3	0.470	0	2	0	0.000
							Average:	456.5	-1.8	0.515	2	2	0	0.000
	A	64	22	◊	D	55	20	508.2	-4.0	0.315	4	3	0	0.000
	B	34	15	◊	H	25	13	257.2	-1.1	0.507	1	2	0	0.000
							Average:	382.7	-2.5	0.411	3	3	0	0.000
Two fold head-to tail	A	50	18	◊	D	47	18	463.0	-0.6	0.611	10	0	0	0.000
	B	55	18	◊	B	55	18	457.9	-1.9	0.514	8	0	0	0.000
	H	43	16	◊	H	43	16	377.3	0.3	0.675	4	0	0	0.000
							Average:	432.7	-0.7	0.600	7	0	0	0.000
Two fold side-by side interface	G	47	17	◊	F	43	15	382.6	-1.3	0.553	1	0	0	0.000
	E	43	17	◊	C	43	15	371.4	-1.4	0.556	1	0	0	0.000
							Average:	377.0	-1.4	0.555	1	0	0	0.000
	F	29	9	◊	G	22	9	189.7	1.7	0.786	1	2	0	0.000
	E	22	9	◊	C	29	10	175.3	0.1	0.637	0	3	0	0.000
							Average:	182.5	0.9	0.711	1	3	0	0.000

Table 7.2: Output from PDBePISA http://pdbe.org/PISA (Krissinel & Henrick, 2007) shows the hydrogen bonds formed between two monomers within the main interface types between NheA molecules in the crystal. In some cases there are multiple bonds between the same side chains, which indicated by the square parentheses.

##	Monomer H	Distance [Å]	Monomer G	Type of Interfaces
1	H:ARG 143[NE]	3.28	G:ASP 151[OD1]	Two-fold back-to-back interface
2	H:ARG 143[NE]	3.14	G:ASP 151[OD2]	
3	H:ARG 143[NH2]	3.32	G:ASN 66[OD1]	
4	H:ARG 143[NH2]	3.15	G:ASP 151[OD2]	
5	H:ASN 58[OD1]	3.21	G:GLN 136[NE2]	
6	H:ASN 66[OD1]	3.39	G:ARG 143[NH2]	
7	H:GLN 136[OE1]	2.57	G:ASN 58[ND2]	
8	H:ASP 151[OD2]	3.04	G:ARG 143[NE]	
9	H:ASP 151[OD2]	3.16	G:ARG 143[NH2]	
##	**Monomer C**	**Distance [Å]**	**Monomer B**	
1	C:GLN 253[NE2]	3.56	B:THR 214[O]	Two-fold head to head interface
2	C:GLN 254[NE2]	2.77	B:PRO 261[O]	
3	C:THR 214[O]	3.20	B:GLN 253[NE2]	
4	C:PRO 261[O]	2.97	B:GLN 254[NE2]	
##	**Monomer A**	**Distance [Å]**	**Monomer D**	
1	A:TYR 100[OH]	2.92	D:ASP 242[OD1]	Two-fold head to tail interface
2	A:SER 243[N]	2.96	D:GLN 330[OE1]	
3	A:ARG 246[NH2]	3.24	D:ASN 105[OD1]	
4	A:ARG 246[NH2]	3.15	D:GLY 104[O]	
5	A:GLN 330[NE2]	2.51	D:ALA 241[O]	
6	A:ASP 242[OD1]	3.08	D:TYR 100[OH]	
7	A:GLN 330[OE1]	3.08	D:SER 243[N]	
8	A:GLN 330[OE1]	3.22	D:SER 243[OG]	
9	A:GLY 104[O]	3.31	D:ARG 246[NH2]	
10	A:ALA 241[O]	2.85	D:GLN 330[NE2]	

Continue table 7.2

##	Monomer D	Distance [Å]	Monomer A	
1	D:THR 224[OG1]	2.71	A:GLN 167[OE1]	Two-fold side by side interface
2	D:HIS 360[ND1]	3.12	A:GLU 288[OE2]	
3	D:THR 214[OG1]	3.57	A:GLY 171[N]	
4	D:GLU 358[OE2]	3.10	A:ARG 292[NH1]	

Table 7.3: **Output from PDBePISA http://pdbe.org/PISA (Krissinel & Henrick, 2007)** shows example of the 8 salt bridges that were formed between monomers H and G that form the two-fold back-to-back interface. In some cases there are multiple bonds between the same side chains, which indicated by the square parentheses.

##	Monomer H	Distance [Å]	Monomer G
1	H:ARG 143[NE]	3.28	G:ASP 151[OD1]
2	H:ARG 143[NE]	3.14	G:ASP 151[OD2]
3	H:ARG 143[NH2]	3.15	G:ASP 151[OD2]
4	H:GLU 140[OE1]	3.71	G:LYS 61[NZ]
5	H:GLU 140[OE2]	3.71	G:LYS 61[NZ]
6	H:ASP 151[OD1]	3.35	G:ARG 143[NE]
7	H:ASP 151[OD2]	3.04	G:ARG 143[NE]
8	H:ASP 151[OD2]	3.16	G:ARG 143[NH2]

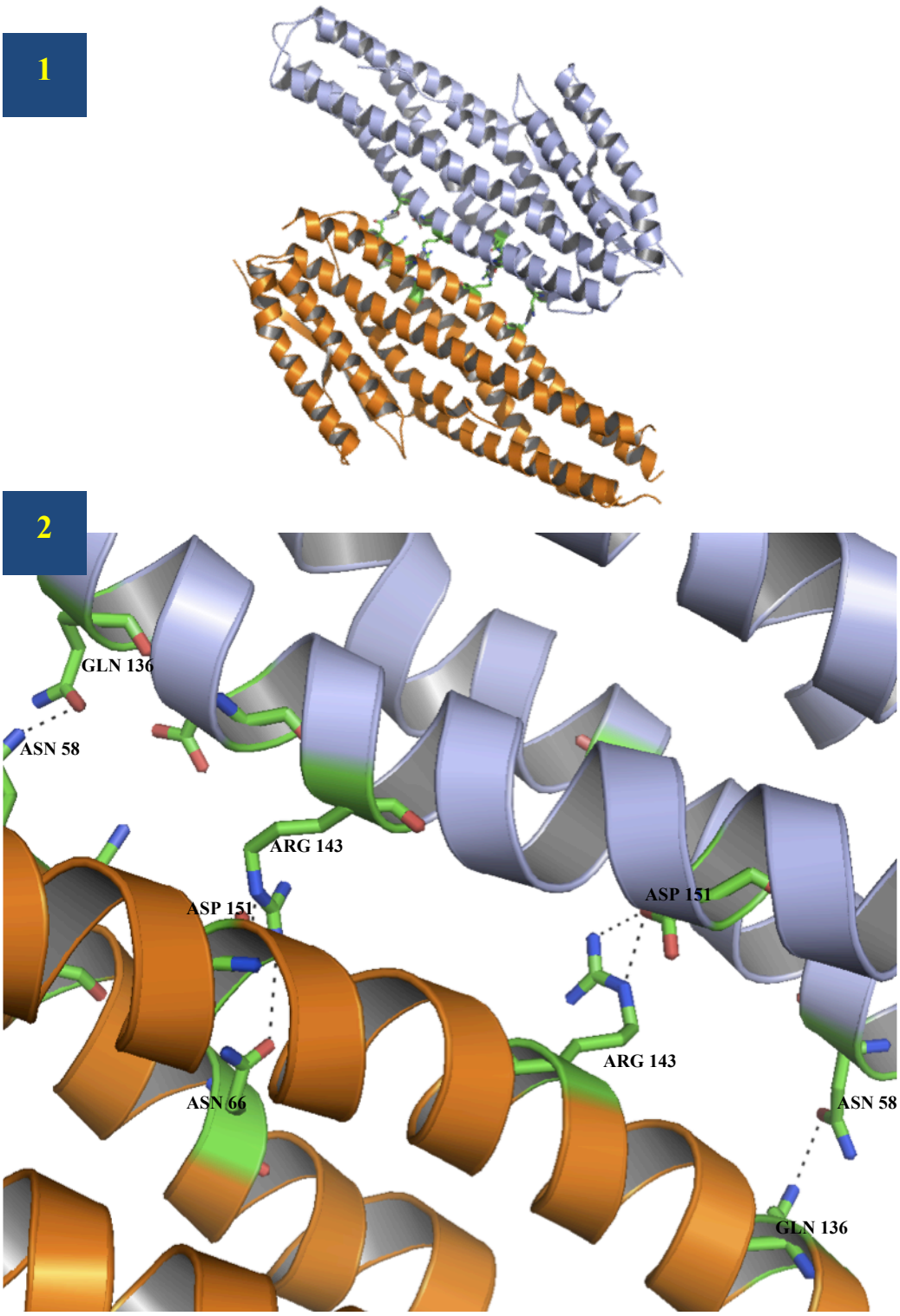


Figure 7.1: (1) Example of a two-fold back-to-back interaction between monomers H (light blue) and G (orange) in the crystal contacts of NheA. (2) The residues involved in this interaction are labelled. The hydrogen bonds between H and G are shown by grey dotted lines. Hydrogen bonds that are more than 3.4 Å long have not been shown. This interface is generated from PDBePISA http://pdbe.org/PISA (Krissinel & Henrick, 2007) and the pictures were generated using PyMol (DeLano and Lam, 2005).

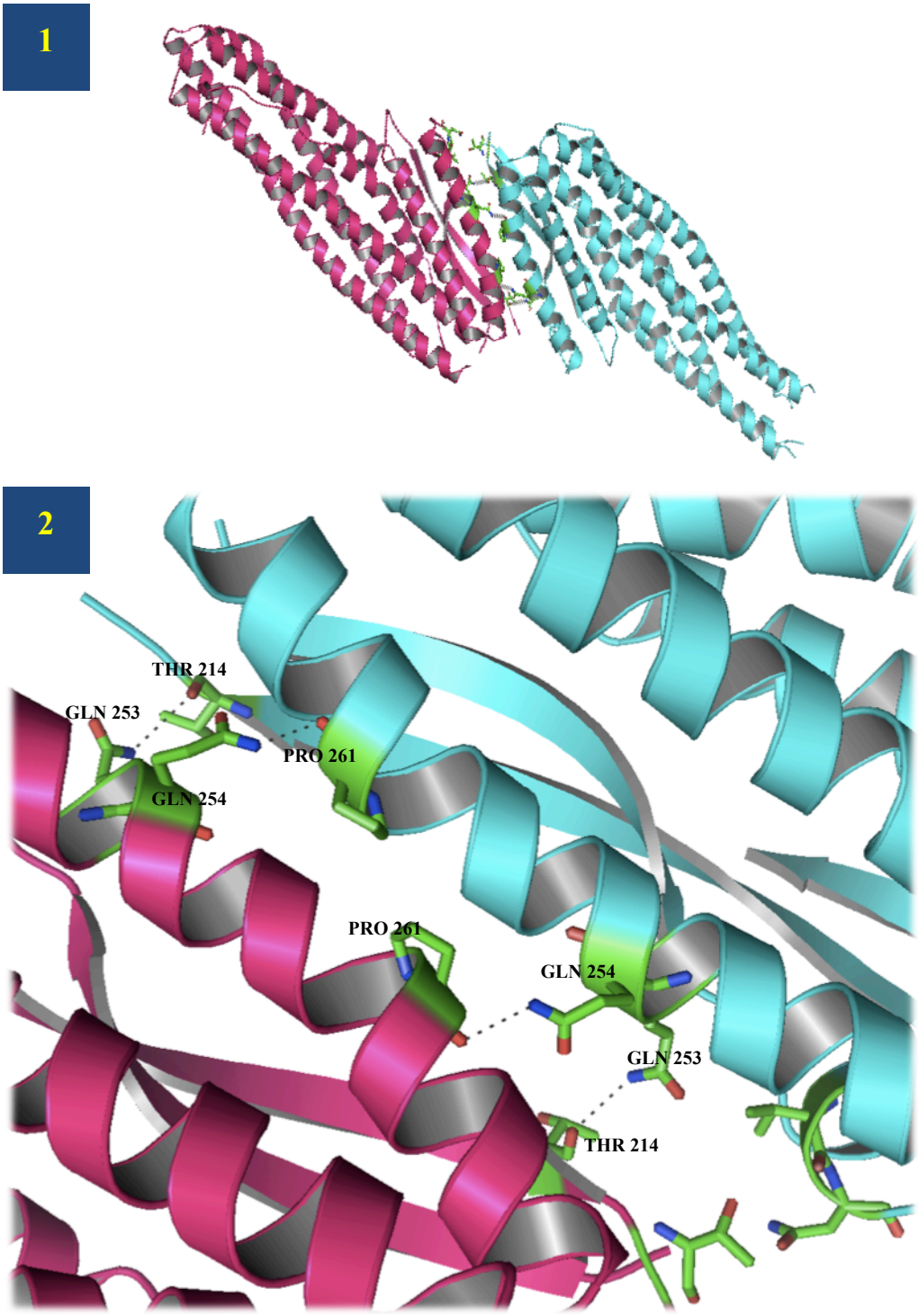


Figure 7.2: (1) Example of a two- fold head-to-head interaction between monomers C (hot pink) and B (cyan) in the crystal contacts of NheA. (2) The residues involved in this interaction are labelled. The hydrogen bonds between C and B are shown by grey dotted lines. Hydrogen bonds that are more than 3.4 Å long have not been shown. This interface is generated from PDBePISA http://pdbe.org/PISA (Krissinel & Henrick, 2007) and the pictures were generated using PyMol (DeLano and Lam, 2005).

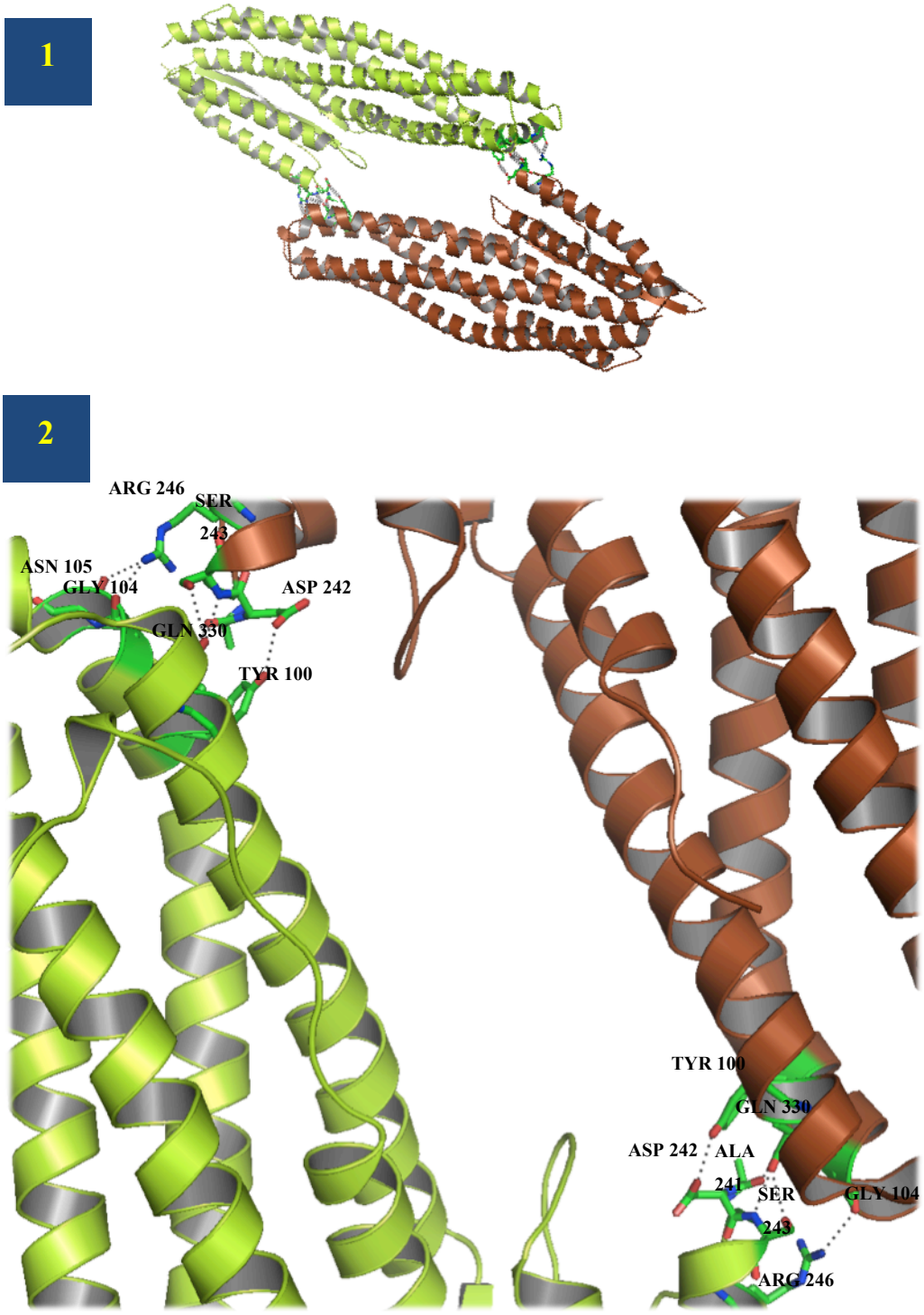


Figure 7.3: (1) Example of a two-fold head-to-tail interaction between monomers A (brown) and D (green) in the crystal contacts of NheA.
(2) The residues involved in this interaction are labelled. The hydrogen bonds between A and D are shown by grey dotted lines. Hydrogen bonds that are more than 3.4 Å long have not been shown. This interface is generated from PDBePISA http://pdbe.org/PISA (Krissinel & Henrick, 2007) and the pictures were generated using PyMol (DeLano and Lam, 2005).

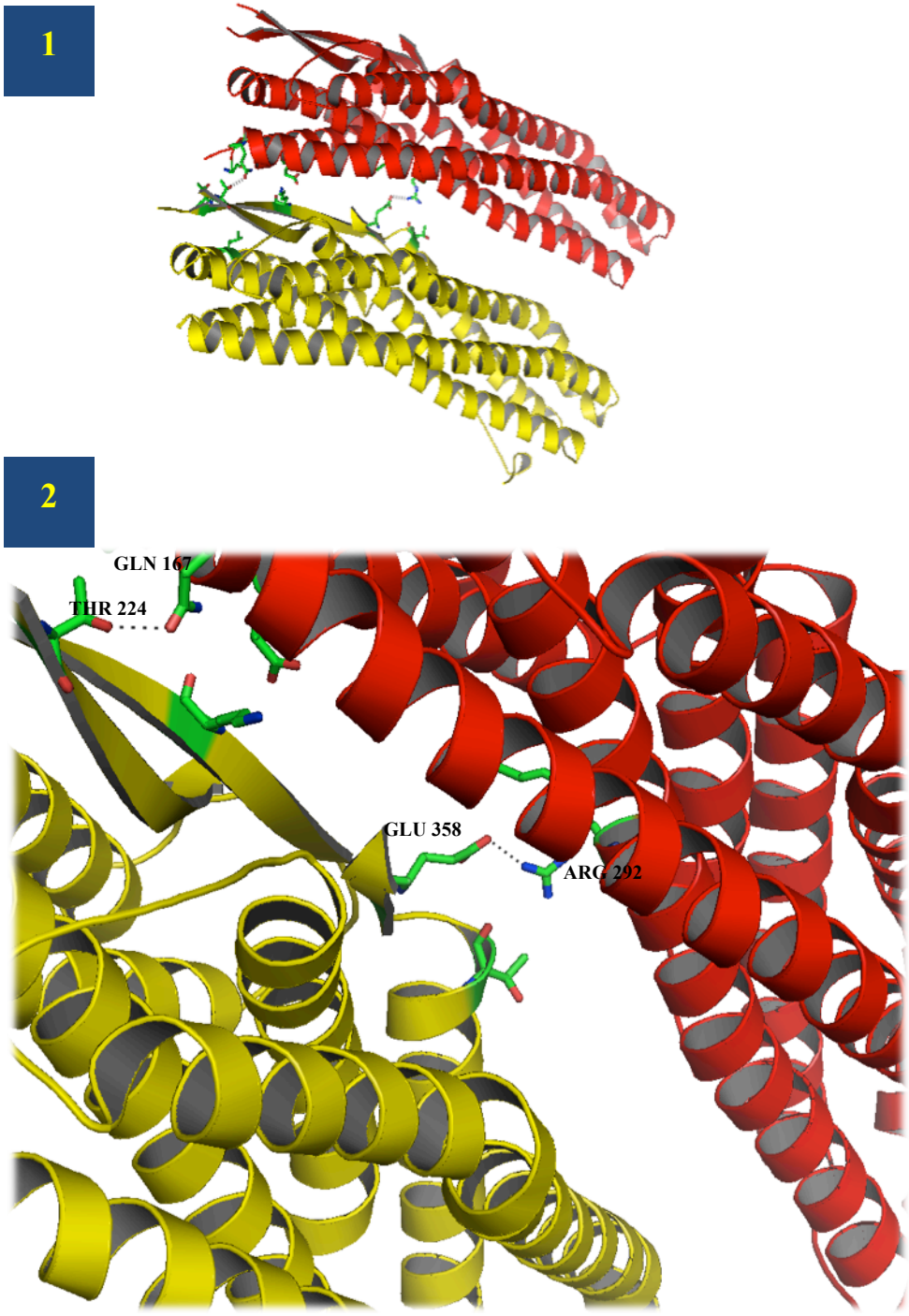


Figure 7.4: (1) Example of a two-fold side-to-side interaction between monomers A (red) and D (yellow) in the crystal contacts of NheA.
(2) The residues involved in this interaction are labelled. The hydrogen bonds between A and D are shown by grey dotted lines. Hydrogen bonds that are more than 3.4 Å long have not been shown. This interface is generated from PDBePISA http://pdbe.org/PISA (Krissinel & Henrick, 2007) and the pictures were generated using PyMol (DeLano and Lam, 2005).

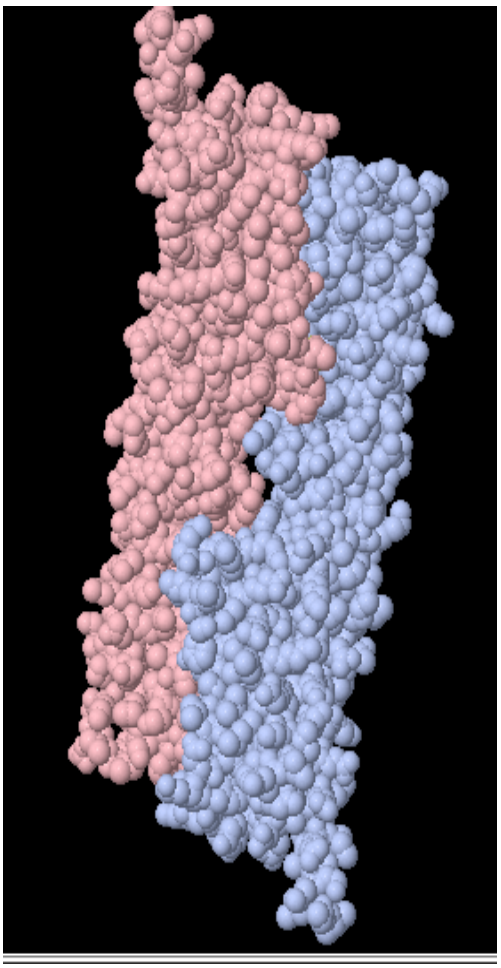


Figure 7.5: Homodimers of the water-soluble form of ClyA.
ClyA is a dimer in the crystal and in solution. This dimeric state of ClyA is very significant in stabilizing the crystal of ClyA in the soluble form. This figure is generated from the PDBePISA http://pdbe.org/PISA (Krissinel & Henrick, 2007).

7.2 Related toxin structures

Although the underlying four helix bundle structure is very common and is present in a range of proteins, the length of the helix bundle coupled with the presence of two subdomains, known as the head subdomain and the tail subdomain, led to the conclusion that ClyA family possesses a novel fold (Wallace et al, 2000).

NheA also has the long helical bundle, elongated shape and extra domains. Other toxins also have extensive α helix bundles such as the water soluble form of the antibacterial toxin colicin Ia (Wiener et al, 1997), which contains two alpha helices, approximately 210 Å long and arranged in a elongated ‘‘Y’’ shape (Figure 7.6) with three functional domains: a receptor domain, a translocation domain and a pore formation domain. The enzymatic membrane damaging, botulinum neurotoxin type A, employs a pair of 105 Å long α-helices in the proposed translocation domain (Lacy et al, 1998). Also the elongated shape of NheA is reminiscent of some other classes of β pore forming toxins such as aerolysin (Parker et al, 1996), perfingolysin (Rossjohn et al, 1997) and staphylococcal LukF (Olson et al, 1999). As well some membrane-associated proteins involved in membrane fusion contain long helical bundles, such as protein HIV-1 gp41 (Weissenhorn et al, 1997) and SIV gp41 (Caffrey et al, 1998), influenza virus haemagglutinin (Bullough et al, 1994), and mammalian SNARE protein (Sutton et al, 1998). However, these proteins do not have significant structure similarity to NheA as estimated by the DALI server (Holm & Rosenström, 2010), which is used for comparing protein structures in 3D.

The Dali search revealed only two significant structure similarities, which were both expected. The search gave a superposition of NheA with *B. cereus* HblB (2NRJ) with a Z-score of 24.3, sequence identity 15%, and a root mean-square deviation (RMSD) of 3.0 Å between 303 α-carbon atom pairs. The superposition between NheA with *E. coli* ClyA monomeric form (1QOY) gave a Z-score of 10.4, sequence identity 8%, and a RMSD of 5.3Å between 221 α-carbon atom pairs.

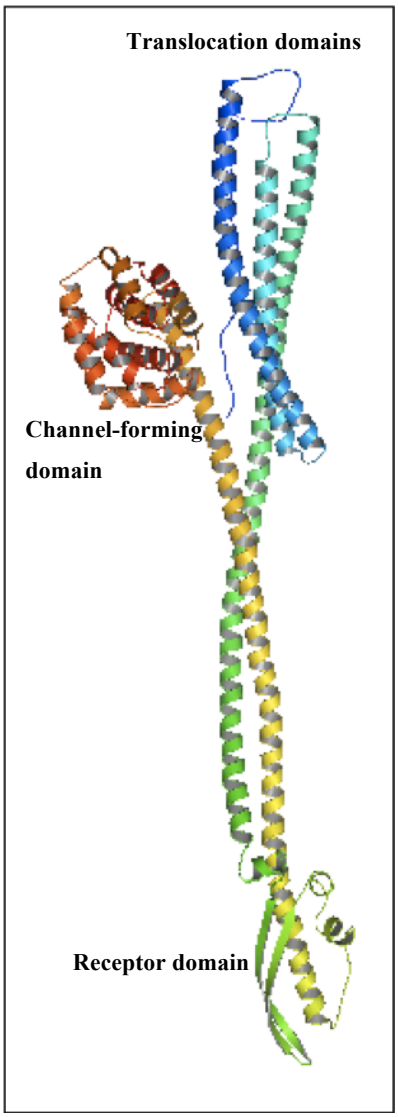


Figure 7. 6: 3D Crystal structure of the colicin Ia monomer.
The structure is shown in ribbon representation with the three domains labelled (Wiener et al, 1997). Produced using PyMol (DeLano and Lam, 2005).

The superposition of NheA with the *E. coli* ClyA pore form (2WCD, chain C) gave a Z-score of 10.3, sequence identity 11%, and a RMSD of 3.4 Å between 198 α-carbon atom pairs All other proteins, gave Z scores less than 10, which indicates that no very significant fold similarity had been found.

This result is as expected, because there is a high level of structural similarity among the three proteins, *B. cereus* NheA, *B. cereus* HblB and *E. coli* ClyA (Figure 7.7), in spite of the lack of sequence homology between them (see Table 1.1 and Figure 1.14).

The main body of all of three proteins consists of four helical chains αA, αB, αC, and αF. In addition, all of them have significant novel structure elaborations at both ends: the tail domains and the head domains (the "ß tongue"). While the overall topography of the three proteins similar in the structure there are some interesting observation to be made:

Firstly, αA in NheA is shorter than in the other toxins (21 residues; SER 24- GLN 44; 32 Å long). In ClyA αA is much longer as it consists of 43 residues (ILE3- ARG 46; 60 Å long) and in HblB consists of 27 residues (GLU 19- LYS 45; 41 Å long). Therefore αA in ClyA and HblB is considered as part of tail domain and as a result the tail domain is a five -helix bundle (Figure. 1.6 and 1.12). However, in NheA the αA becomes shorter (Figure 6.9 A) and αA does not occupy the lower half of the NheA structure.

Furthermore, In ClyA a distinct kink is present at residue PRO 36 (within αA), but this becomes part of the turn between αA1 and the extended αB of the protomer of ClyA pore form (Figure 1.7 and Figure 1.8). HblB has a few distinct kinks, which are located at LYS 83 -VAL 84 (within αB), LEU189 -GLY 190 (αD), LEU 228 -GLY 229 (αE), and LYS 310-PRO 311 (αG). However, in NheA there are the distinct kinks within αB at residues TYR 95 and TYR 96 and within αC with distinct kink at residue LEU 141.

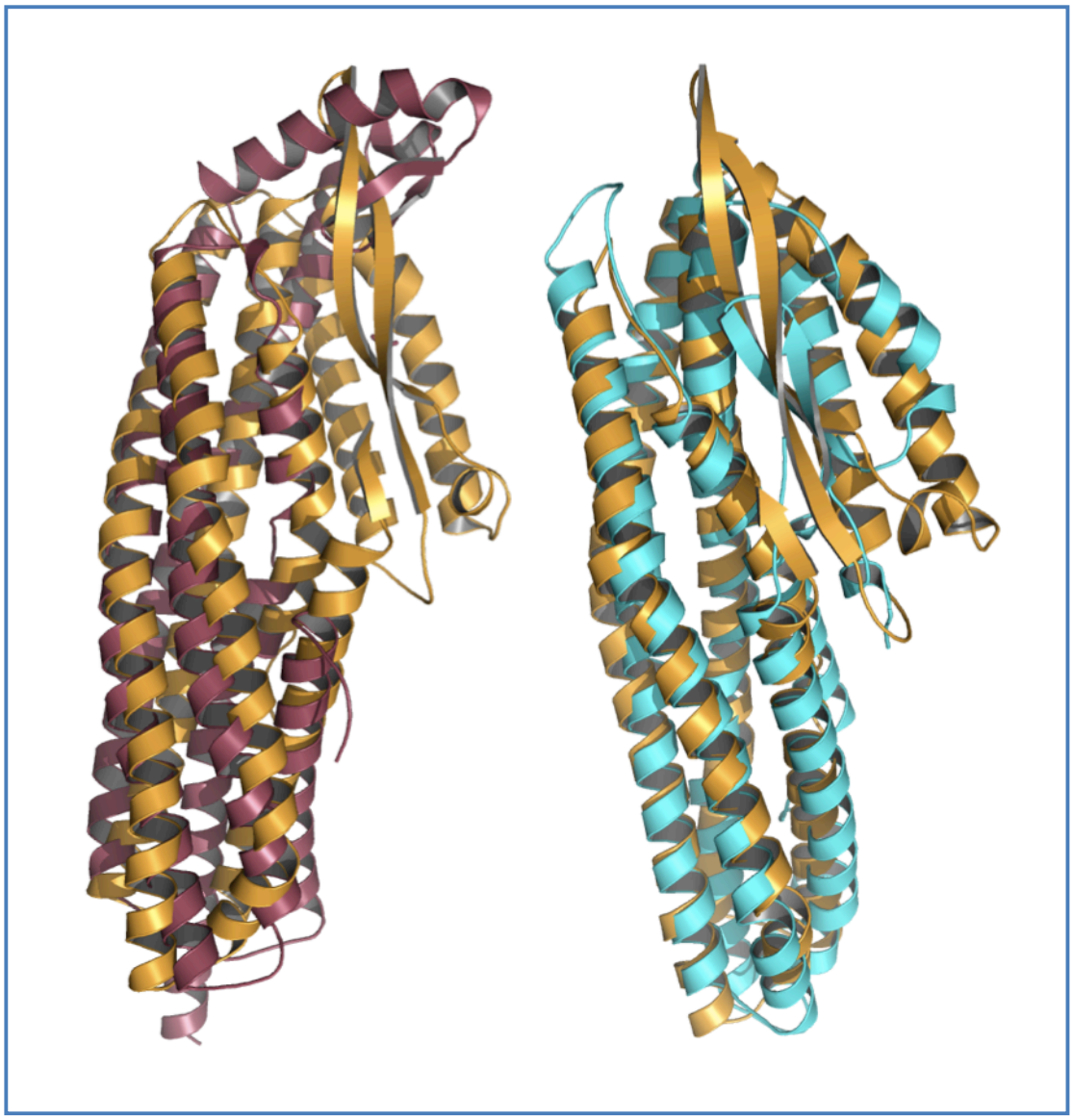


Figure 7.7: The superimposition of NheA (orange) with HblB (cyan) and ClyA (pink).
Dali superimposed the NheA and HblB over 238 α-carbon atom pairs with an RMSD of 3 Å; NheA and ClyA over 303 α-carbon atom pairs with an RMSD of 5.3 Å.
This figures were generated using DALI web server (Holm & Rosenström, 2010) (http://ekhinda.biocenter.halsinki.fi/dali_server/) and PyMol (DeLano and Lam, 2005).

- The second and the main structural difference is the orientation of the head domain with respect to the tail domain. This differs substantially between *E. coli*. ClyA and both *B. cereus* NheA and *B. cereus* HblB. The head domain is turned downward and interacts with the tail domain in *B. cereus* NheA and HblB, whereas it is turned upward in ClyA and forms no interactions with its tail domain (Figure 7.8). The interdomain angle for NheA is ≈20° and HblB is ≈30°, whereas it is ≈120° in ClyA.
  The different orientation of the head domain in *E. coli* ClyA and *B. cereus* NheA, and HblB may indicate that the head domain can reorient itself in these toxins. However, it is notable that the eight independent copies of NheA in asymmetric unit all have the same orientation.

In addition the NheA beta tongue is much longer than in the other two toxins (ClyA or HlbB). It is approximately 53 Å long in NheA, but it is 20 Å long in ClyA and 30 Å long in HblB; (Figure 7.8). These measurements were done in the program COOT (Emsley & Cowtan, 2004).

- The third area of dissimilarity is in the type of amino acids in the beta tongue. In the beta tongue of HblB the 17 residue hydrophobic segment (GAILGLPIIGGIIVGVA) and in the beta tongue of ClyA the 27 residue segment (AYAGAAAGVVAGPFGLIISYSIAAGVV) are mainly hydrophobic in sequence. They have been identified as transmembrane regions (Madegowda et al, 2008; Mueller et al, 2009; Wallace et al, 2000). In the pore form of ClyA the beta tongue becomes a hydrophobic transmembrane α helix (see Figure 1.8) (Mueller et al, 2009). However, in NheA the beta tongue of 25 residues is amphipathic (GSINIGKQ NFTITIKT IDFVSIGTL). In fact NheA has no predicted transmembrane region (Granum et al, 1999).

Although NheA, HblB and ClyA do not have significant sequence similarity (identities: NheA /ClyA =18%, NheA /HblB =20%, HblB /ClyA =20%), they are nevertheless similar in structure, suggesting that once again these toxins represent a novel super-family of pore forming toxins and may

be have a common mode of pore formation.

While NheA has no predicted transmembrane region (Granum et al, 1999), its elongated amphipathic beta sheet (Figure 7.9 A) is reminiscent of those in some other toxins for which structures have been reported, such as *Staphylococcus aureus* α-hemolysin (Song et al, 1996) (see section 1.3 for more details), *Staphylococcus aureus* Leukocidin F (LukF) (Olson et al, 1999), aerolysin (Parker et al, 1996), and perfringolysin (Rossjohn et al, 1997). These form pores based on β sheet barrels rather than α helical architectures (see Figure 1.2). These pore forming toxins contains long β strands ranging from 93 Å length in aerolysin to 70 Å in perfringolysin, and staphylococcal LukF. The hydrophobic residues on one side of the β strands face out toward the membrane. However, the hydrophilic ones on the other side of the β strands form the pore.

In *S. aureus* α-hemolysin it is thought that a short N-terminal β strand acts as "latch" to release the β strands that form the β sheet barrels in the oligomeric pore (Figure 7.9 B and C) (Song et al, 1996). It is therefore interesting that in NheA the short beta sheet at the C-terminal, which is known as β3 in the NheA structure (Figure 7.9 A), could possibly similarly act as a "latch" to release the beta tongue. Consequently, the questions arise: Could the NheA protein be a β pore-forming toxin? Could it be considered as a link between α and β pore forming toxins? And could this member of an α-PFT family form a transmembrane β-barrel like *S. aureus* α-toxin? The NheA structure presented here offers a starting point for further biophysical investigations of the pore form of the complex to answer these questions. Cleary knowledge of the structures of NheB and NheC will be important in this future work.

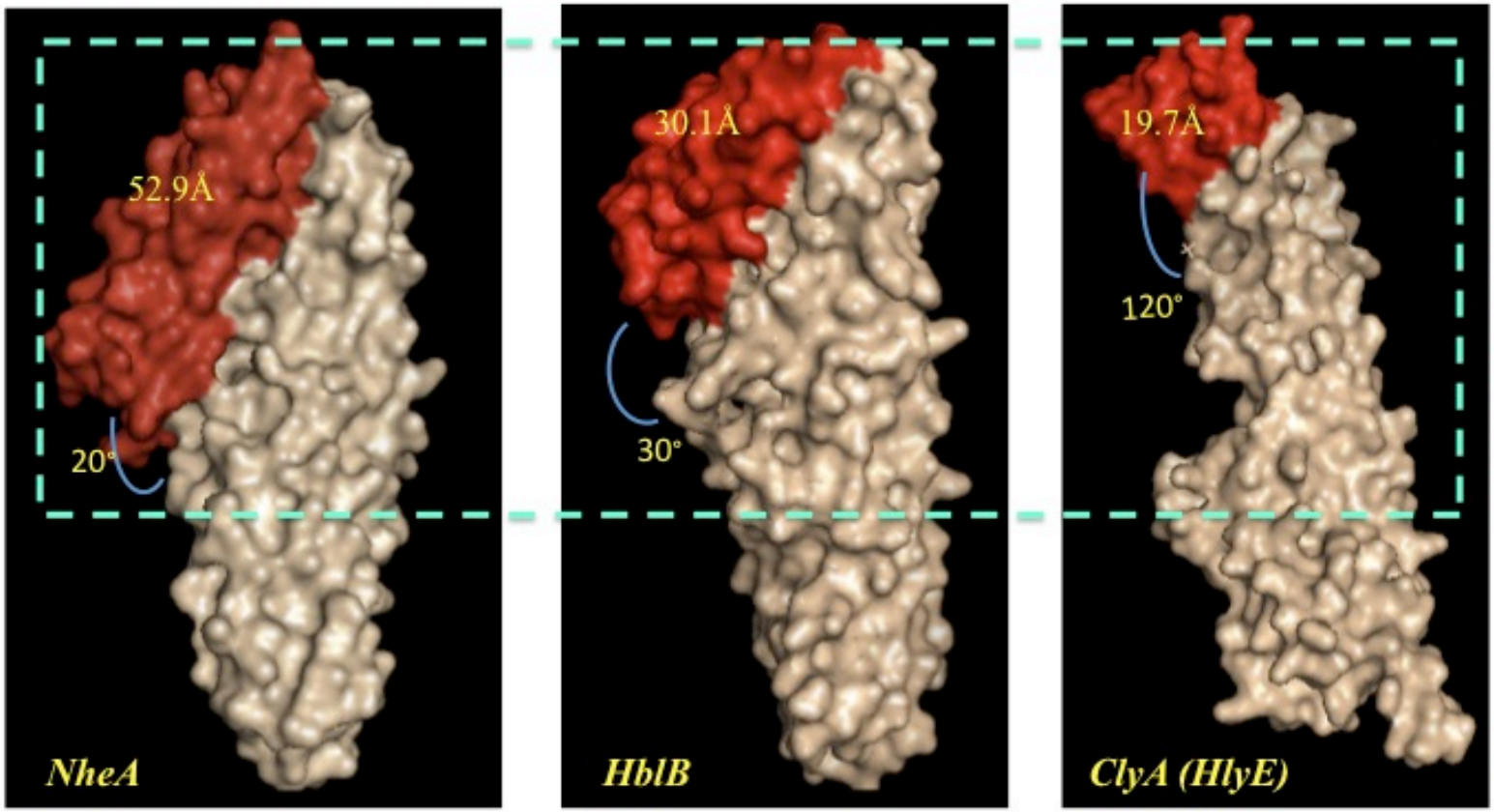


Figure 7.8: Molecular surface view of NheA, HblB and ClyA.
Red colour indicates the head domains. Calculated inter-domain angles and long distances of the toxins are also depicted. The different orientations of the head domain are shown. Downward and upward β-tongue orientations of NheA, HblB and ClyA are labeled. Produced using PyMol (DeLano and Lam, 2005).

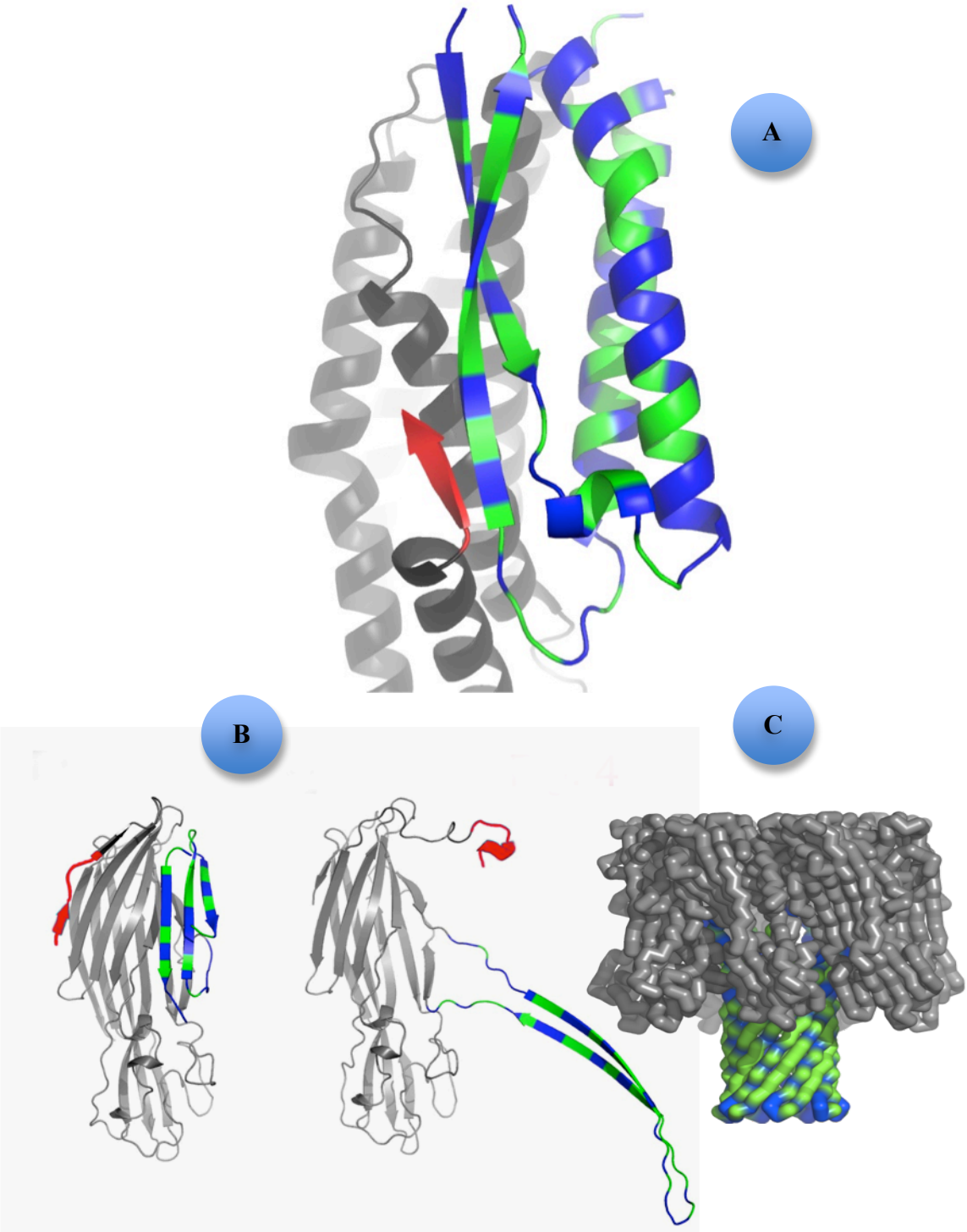


Figure 7.9: Comparison between the amphipathic β sheet of NheA and the transmembrane region in the classic example of a β-PFT.
(A) The 25 residue long β sheet of NheA is amphipathic (hydrophobic =green & hydrophilic=blue in diagram). Red region represents the short ß3 at the C-terminus, which could act as a "latch" to release the beta tongue. An enlarged image of the β sheet of NheA is shown on the top.
(B) The water -soluble form of staphylococcal LukF and (C) The pore form of *Staphylococcus* aureus α-hemolysin, which has long amphipathic transmembrane regions. Produced using PyMol (DeLano and Lam, 2005).

7.3 Electrostatic surface potential of NheA

The electrostatic surface potential of NheA was calculated using GRASP (Nicholls et al, 1991) (Figure 7.10). Examination of the charge distribution on the surface of NheA shows that NheA has fairly negative potential in the middle of the main body and the head domain. However, positive charge is distributed randomly in the body of NheA. The electronegative potential surface of NheA may be involved in specific binding to NheB or NheC. In the water-soluble forms of ClyA and HblB the negative charge is present mostly in tail domain (Figure 7.10 C and B).

However, it is clear from Figure 7.10 that positively charged LYS and ARG amino acids are present randomly in the bodies of all the three toxins, NheA, HblB and ClyA.

The protomer surface of the ClyA pore form is largely lighter color shades, which means the molecule is mostly neutral (Figure 7.10 D). On the other hand, the protomer surface has slightly negative potential if rotated 180°. (Mueller et al, 2009) indicating that the surface of the transmembrane region is neutral with a width of approximately 30Å and the inner surface of the pore has a mostly negative potential, which reflects the cation selective nature of the ClyA pore (Ludwig et al, 1999). ClyA is a homo-oligomeric pore forming toxin and needs only one protein to achieve maximum cytotoxic activity where both Hbl and Nhe are hetero-oligomeric pore forming toxins and require three proteins to obtain maximum cytotoxic activity so the structures of Hbl and Nhe may be close to each other than to ClyA .

Although the binding of PFTs to lipid depends mostly on their hydrophobicity, it was found that most pore forming toxins are cationic as the positive charge residues (LYS and ARG) interact with the negative phosphate groups of phospholipids, the major component of cell membrane. This electrostatic interaction is increased the binding of hydrophobic core into lipid bilayer of the cell membrane (Fuertes et al, 2010).

For some pore forming toxins the electrostatic surface potential of the lipid bilayer is very important for insertion of the toxin into the membrane of target cells. The ion–channel forming toxins such as Colicin la form a voltage gated conducting channel through the cell membrane of the target bacteria. This toxin requires negatively charged lipids in the target membrane at acidic pH (as the pH is reduced below 5.2) to form channels in the membrane (Mel & Stroud, 1993).

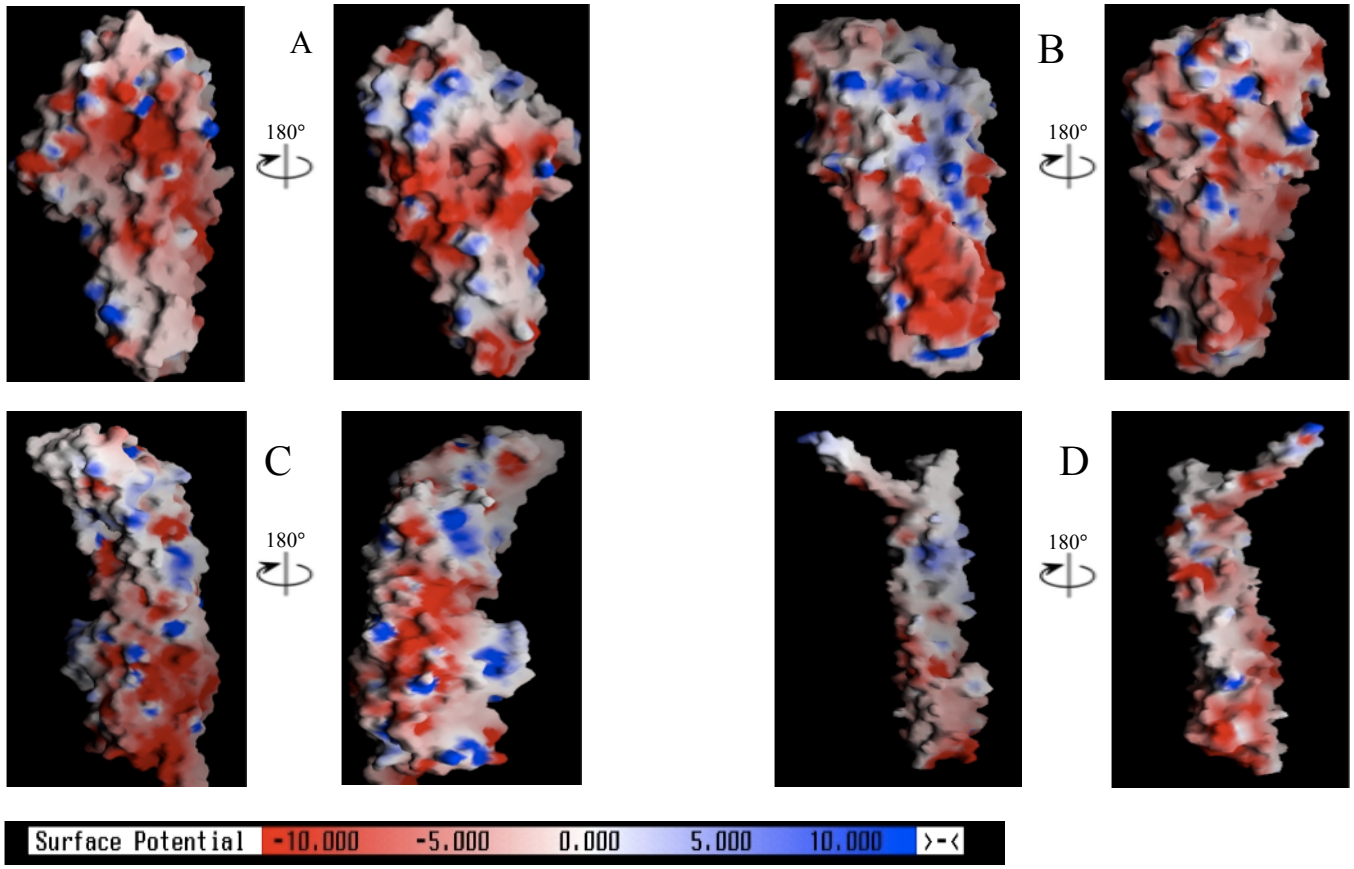


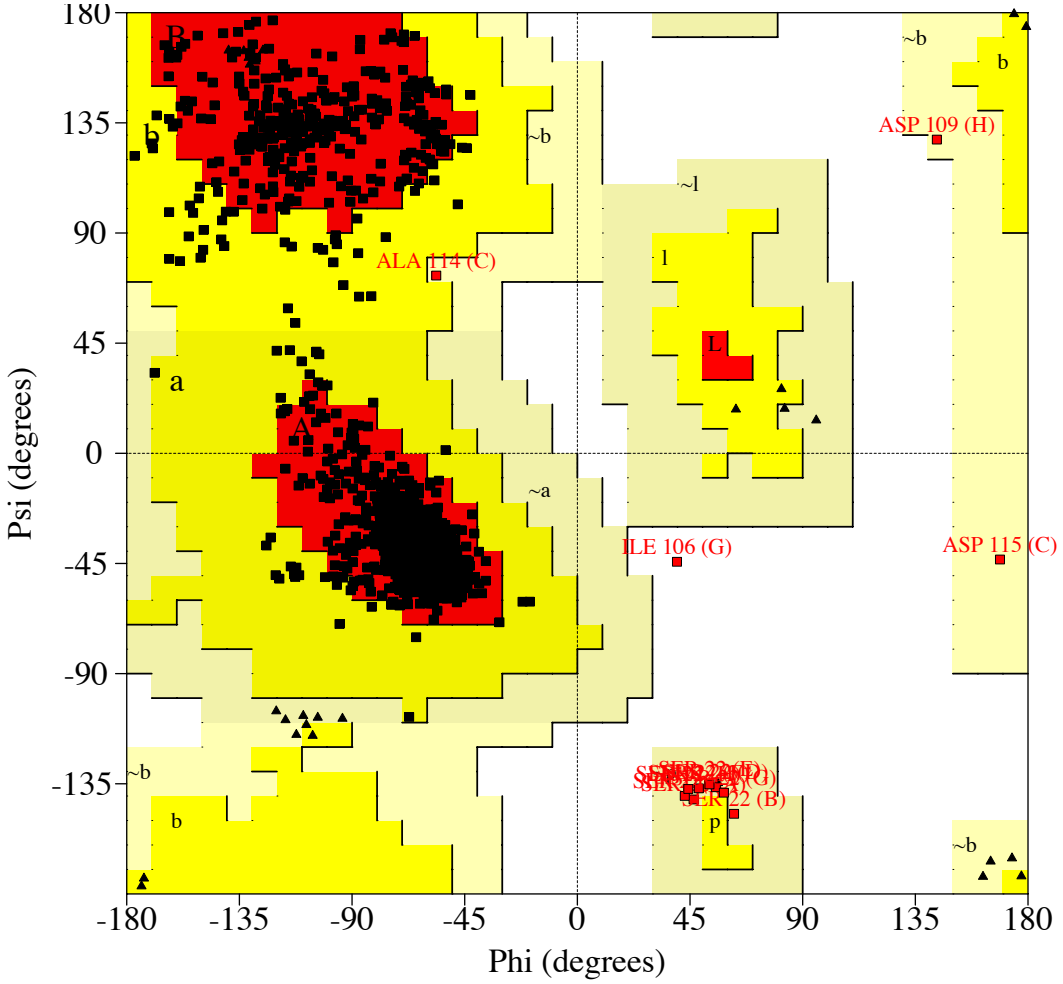
Figure 7.10: Electrostatic surface charge potential diagrams for (A) NheA, (B) HblB, (C) water soluble form of ClyA and (D) the protomer of ClyA pore form. Positively and negatively charged surfaces are shown by blue and red colouring respectively according to the scale at the bottom of the diagram. Electrostatic surface charge potential diagrams for each toxin was drawn after a 180° rotation. This figure was generated using GRASP (Nicholls et al, 1991).

7.4 NheA model stereochemistry

The final model of NheA from the crystal that diffracted to 2.05 Å resolution comprised of 2672 residues in the eight molecules in the asymmetric unit. The lower numbers of observed amino acids were due to absences of these amino acids at N- terminal or loop and turn regions.

The NheA coordinates were input into PROCHECK (http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/) (Laskowski et al, 1993) for structure validation and analysis. Ramachandran plots (Ramachandran et al, 1963) are shown in Figure 7.11. It is clear that there is one residue in the disallowed regions, which is ILE 106 of monomer G. We have no explanation of why this residue has ϕ, and ψ values in the disallowed regions as it has reasonably good electron density. There are 11 residues in the generously allowed region, and 96.1% in the most favoured regions. The main chain and side chain parameters with Chi1-Chi2 plots are shown in Figures 7.12, 7.13 and 7.14 and also indicate acceptable geometry. These results all indicate that the *B. cereus* NheA structure is of very good quality.

The NheA structure was also input into the MOLPROBITY program (http://molprobity.biochem.duke.edu/) (Chen et al, 2010) to do extra model validation by identifying possible steric clashes within the model and adjustment of side chain amide rotamers to satisfy optional hydrogen bonding interactions. Overall the MOLPROBITY score is in the 97th percentile (100% being the best amongst structures of comparable resolution). MOLPROBITY reported a low clash score of 7.4 corresponding to the 93th percentile. 98% of Ramachandran favored, and 0.35% of Ramachandran outliers were reported. 1.45% of side chain conformations marked as rotamer outliers with no residues with bad lengths and 0.15% of residues with bad angles.



Plot statistics

Residues in most favoured regions [A,B,L]	2392	96.1%
Residues in additional allowed regions [a,b,l,p]	85	3.4%
Residues in generously allowed regions [~a,~b,~l,~p]	11	0.4%
Residues in disallowed regions	1	0.0%
	----	------
Number of non-glycine and non-proline residues	2489	100.0%
Number of end-residues (excl. Gly and Pro)	1577	
Number of glycine residues (shown as triangles)	75	
Number of proline residues	44	

Figure 7.11: Ramachandran plot for NheA.
Produced by the programme PROCHECK (Laskowski et al, 1993).

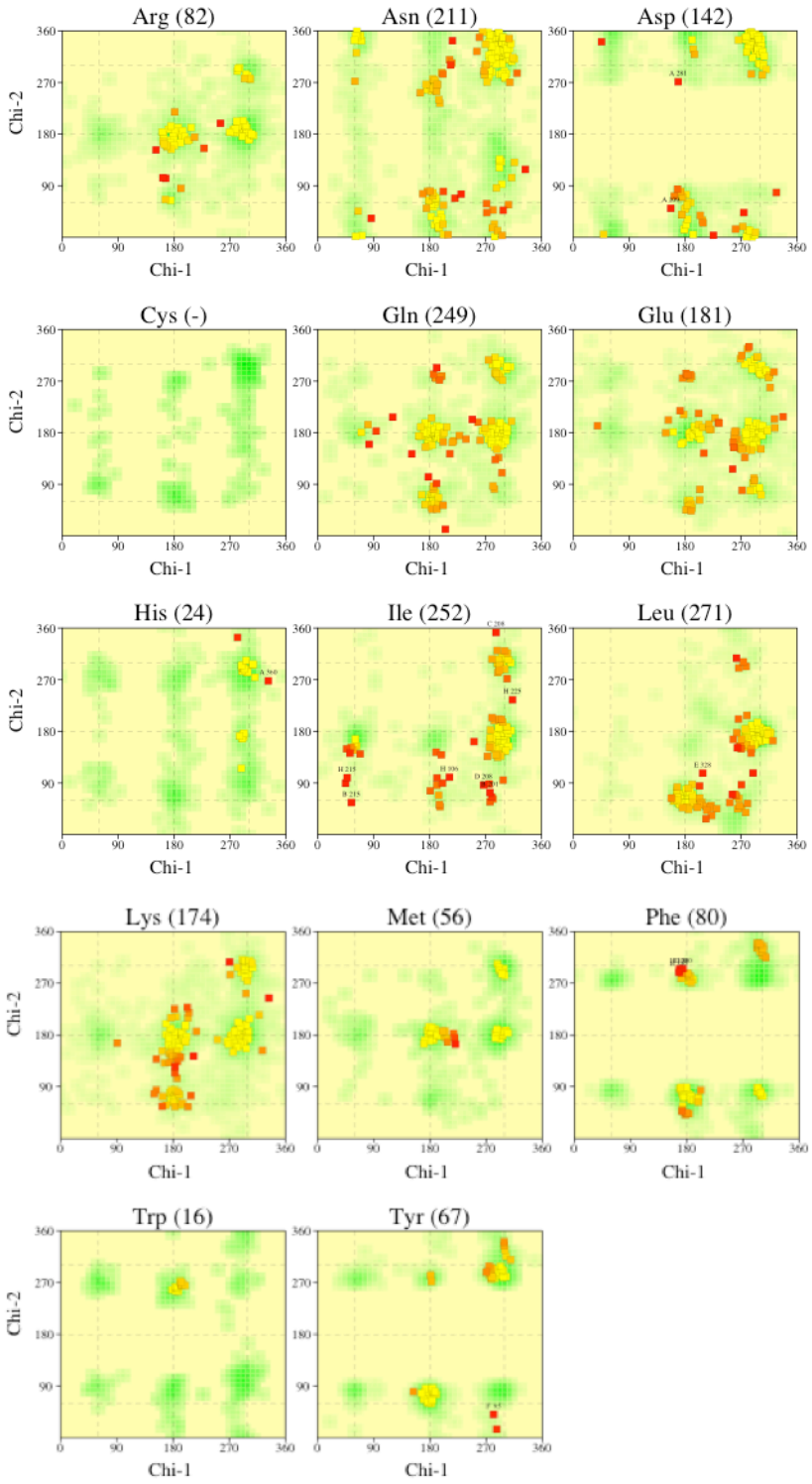
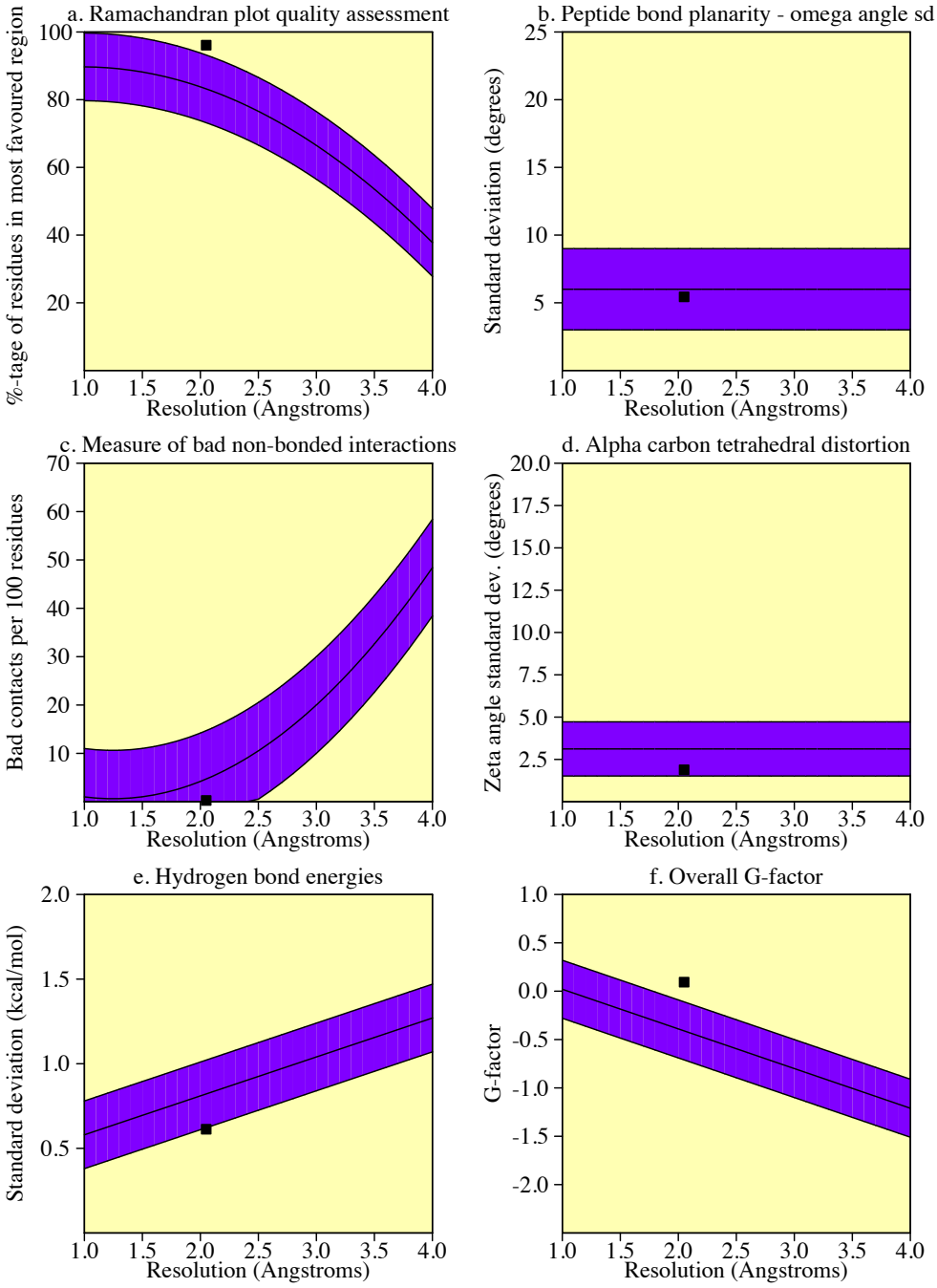


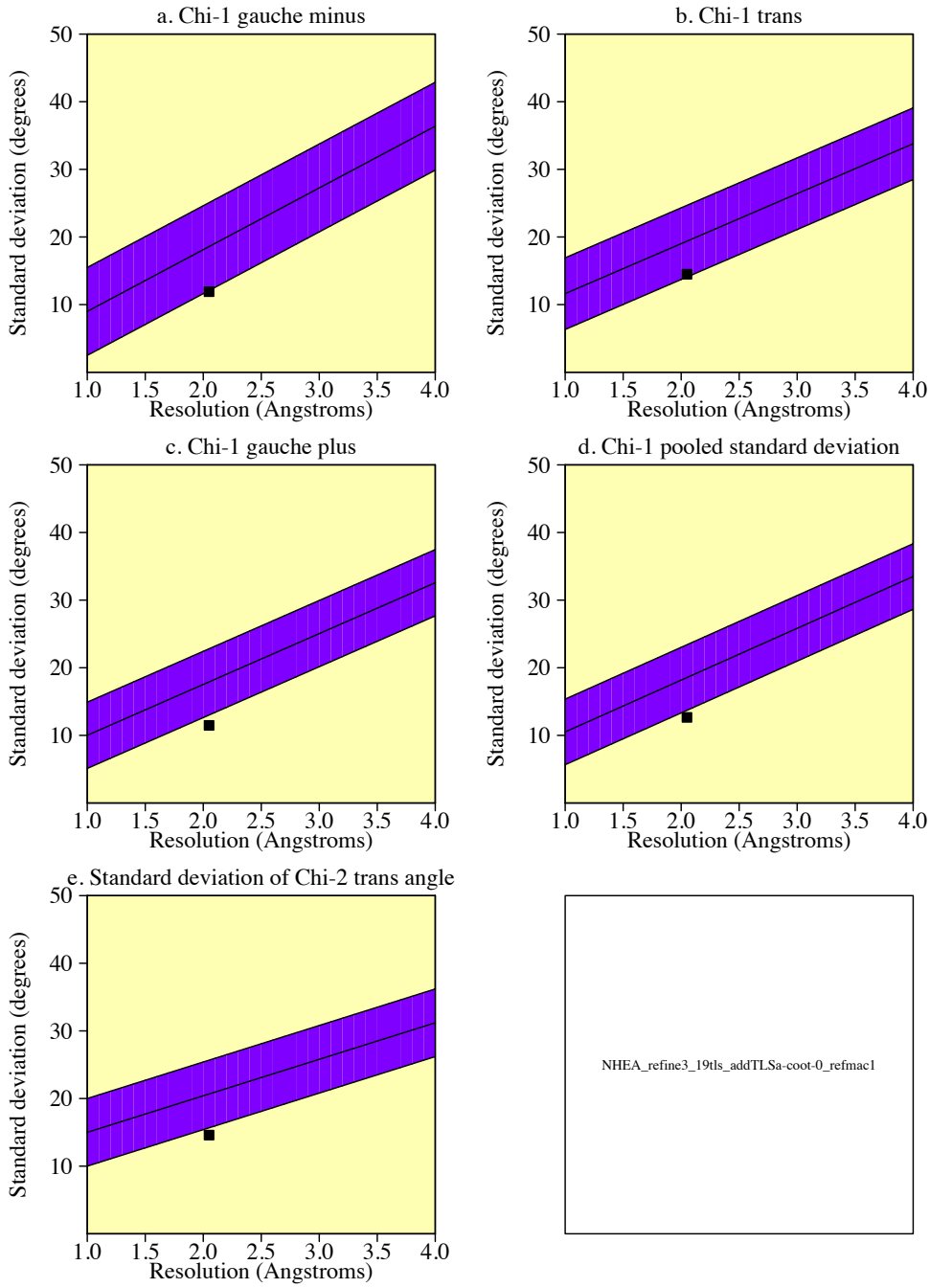
Figure 7.12: Chi 1-Chi 2 plots for various residue types in NheA. Number of residues are shown in brackets favourable conformations lie in shaded regions as obtained from an analysis of 163 structures at resolution 2.0 Å or better, produced by the programme PROCHECK (Laskowski et al, 1993).



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	2489	96.1	83.2	10.0	1.3	BETTER
b. Omega angle st dev	2632	5.4	6.0	3.0	-0.2	Inside
c. Bad contacts / 100 residues	11	0.3	4.7	10.0	-0.4	Inside
d. Zeta angle st dev	2576	1.9	3.1	1.6	-0.8	Inside
e. H-bond energy st dev	2168	0.6	0.8	0.2	-1.0	BETTER
f. Overall G-factor	4185	0.1	-0.4	0.3	1.7	BETTER

Figure 7.13: Main chain parameters for NheA. Showing all values are better than the average for a structure at this resolution.
Produced by the programme PROCHECK (Laskowski et al, 1993).



Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	155	11.9	18.6	6.5	-1.0	BETTER
b. Chi-1 trans st dev	821	14.5	19.4	5.3	-0.9	Inside
c. Chi-1 gauche plus st dev	1346	11.5	17.9	4.9	-1.3	BETTER
d. Chi-1 pooled st dev	2322	12.7	18.6	4.8	-1.2	BETTER
e. Chi-2 trans st dev	891	14.6	20.7	5.0	-1.2	BETTER

Figure 7.14: Side chain parameters for NheA. Showing all values are better than the average for a structure at this resolution.
Produced by the programme PROCHECK (Laskowski et al, 1993).

7.5 B value distribution

Thermal motion means oscillations of an atom around its mean position due to temperature. Crystal disorder means the atoms do not occupy the same position in every unit cell or in every asymmetric unit (Rhodes, 2006). The atomic scattering factor is commonly affected by the variation in atomic position caused by the real motion of temperature changes and/or crystal disorder. A higher value of B implies a greater uncertainty in the positions of atoms. Temperature can cause atomic mobility/fluctuations, that affect the B-values (B factors), which also known as "temperature-factor " or "Debye-Waller factors".

The average temperature factor for the NheA structure is 35.4 $Å^2$ and for the solvent atoms are 34.7 $Å^2$, which means in general the distribution of B-values is reasonable. The average B factor distribution for the each monomer in NheA was shown in Table 7.4. Generally the highest average B-value for a whole chain was in monomer H, which is about 42.3 $Å^2$, where monomer A has the lowest average B-value for the whole chain, which is about 31 $Å^2$.

From Figure 7.15 and Figure 7.16, it is clear that the main chain has a lower B factor distribution than the side chains because the atoms in side-chains are expected to have more freedom of movement than the atoms of the main chain. Several residues have side-chains with high B factors such as ASP 109, GLU 110, SER 170 in monomer A; ASP 109, GLU 110, GLN 111 in monomer B; ASP 325, GLU 320, GLU 319 in monomer C; TYR 95, GLU 110, GLN 111 in monomer D; ASN 107, ASP 325, 112 SER in monomer E; ASP 115, ASP 325, ASN 107 in monomer F; ASN 105, ASN 105, GLU 108 in monomer G; ASP 325, ASP 109, GLU 110 in monomer H. However, it is believed that the major contributors to the B-values comes from the crystal disorder than the thermal motion of the atoms (Rhodes, 2006). This is particularly true when crystals are cooled to liquid nitrogen temperature.

Most of the disordered amino acids, which have high B values, are part of coils, turns, bulges or hairpins. As can seen from Figure 7.15 and Figure

7.16; residues in turns of each monomer have generally higher B-values indicating that turns have a higher degree of mobility than the α helix or β sheets which confirm that regular secondary structure have lower conformational mobility than turns (Krishnaswamy & Rossmann, 1990). This is expected because the hydrogen-bonded α-helices and β-sheets tend to have lower atomic fluctuations than the random turns of a protein (Yuan et al, 2003).

The N-terminal comprising amino acids LYS 1– ASN 23 is not as well defined as the C-terminal, especially at the beginning. The electron density in this region is weak, and more a flexible and mobile region. As result, even when these residues could be built they have also high B-values.

The similarity of the temperature factor distribution of the 64 α helices and 24 β strands of the NheA structure suggest a similarity in the level of mobility of the eight monomers of NheA.

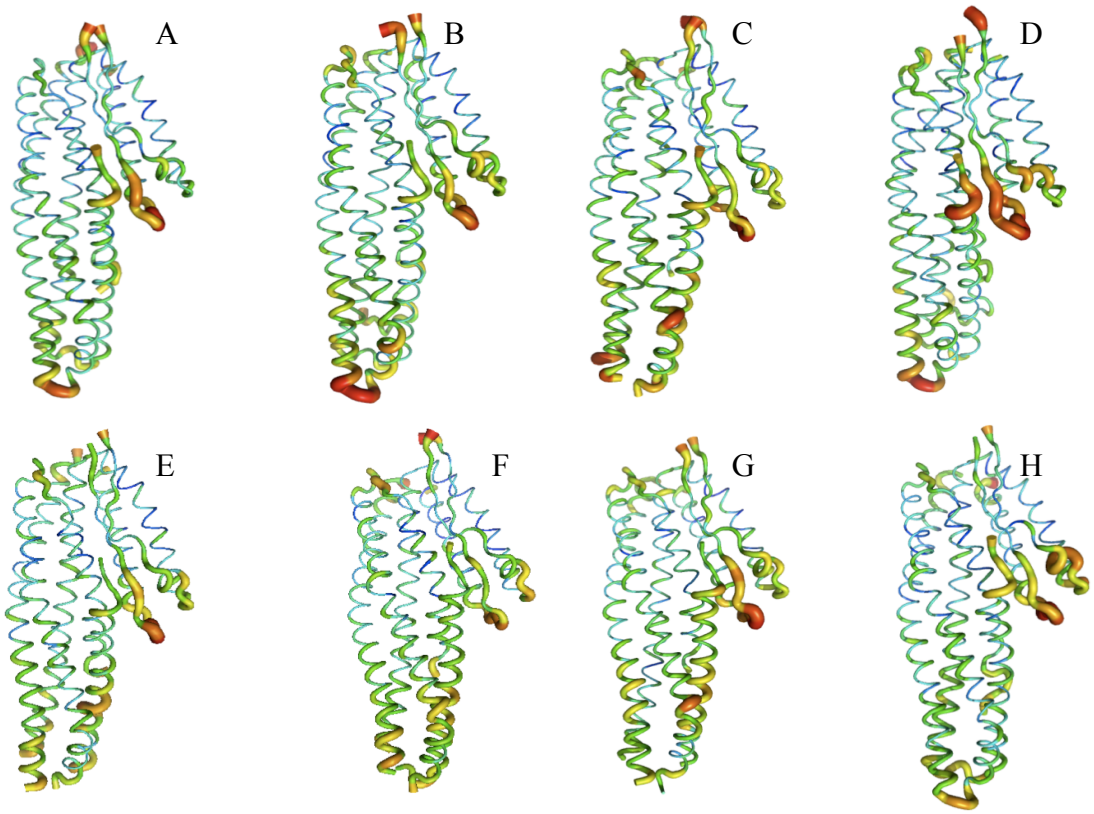


Figure 7.15: B-factors of NheA molecules.
This figure shows a cartoon reproduction of the monomers of NheA coloured by B-factor. The thin blue lines indicate the lowest B-factors changing to thick red lines to indicate the highest B-factors. As expected, turns have B-factors higher than other parts of the structure.
Produced using PyMol (DeLano and Lam, 2005).

Table 7.4: The average B factors for main chain, side chain and overall value for each monomer of NheA.

	Average B value for main chain ($Å^2$)	**Average B value for side chain ($Å^2$)**	**Average B value for whole chain ($Å^2$)**
Chain A	28.1	34	31
Chain B	32.6	38.3	34.4
Chain C	30.1	35.3	32.7
Chain D	29.5	35.3	32.4
Chain E	33.9	39.8	36.8
Chain F	32.5	37.6	35
Chain G	36.1	41.1	38.6
Chain H	39.9	44.8	42.3

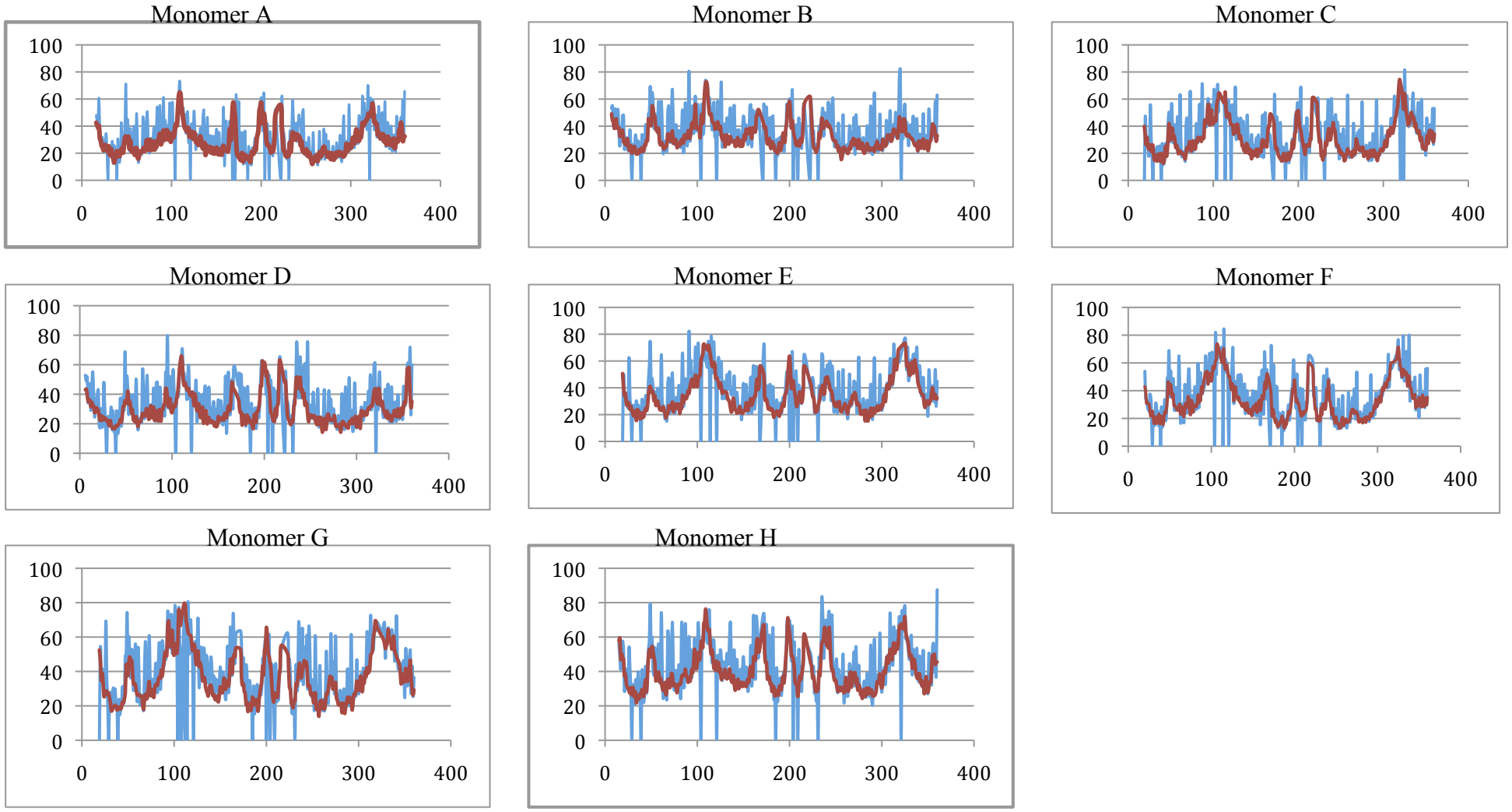


Figure 7.16: B-factors (Å^2) of main and side chain residues of NheA molecules plotted against residue numbers (X-axis). Red line represents B-factors of main chain and blue line represents B-factors of side chains. It is clear that the B-factors of side chains are higher than main chain. Zero side chain values correspond to Glycine or unbuilt side chain.

References

Bullough PA, Hughson FM, Skehel JJ, Wiley DC (1994) Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* **371:** 37-43

Caffrey M, Cai M, Kaufman J, Stahl JS, Wingfield TP, Covell DG, Gronenborn MA, Clore GM (1998) Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41. *The EMBO J* **17:** 4572–4584

Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66:** 12-21

DeLano WL, Lam JW (2005) PyMOL: A communications tool for computational models. *Abstr Pap Am Chem Soc* **230:** 254-COMP

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta crystallographica* **60:** 2126-2132

Fuertes G, Giménez D, Esteban-Martín S, García-Sáez AJ, Sánchez O, Salgado J (2010) Role of membrane lipids for the activity of pore forming peptides and protein. In *Proteins: membrane binding and pore formation* Anderluh G, Lakey J (eds), Vol. 677, pp 56-66. Springer Science and Business Media, LLC

Granum PE, O'Sullivan K, Lund T (1999) The sequence of the non-haemolytic enterotoxin operon from *Bacillus cereus*. *FEMS Microbiol Lett* **177:** 225-229

Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. *Nucl Acids Res* **38:** 545-549

Krishnaswamy S, Rossmann MG (1990) Structural refinement and analysis of mengo-virus. *J Mol Biol* **211:** 803-844

Krissinel E, Henrick K (2005) Detection of protein assemblies in crystals. In *Computational Life Sciences, Proceedings*, Berthold MRGRDKKOFI (ed), Vol. 3695, pp 163-174.

Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* **372:** 774-797

Lacy DB, Tepp W, Cohen AC, DasGupta BR, Stevens RC (1998) Crystal structure of botulinum neurotoxin type A and implications for toxicity. *Nat Struct Biol* **5:** 898-902

Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) Prochecka program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26:** 283-291

Ludwig A, Bauer S, Benz R, Bergmann B, Goebel W (1999) Analysis of the SlyA-controlled expression, subcellular localization and pore-forming activity of a 34 kDa haemolysin (ClyA) from *Escherichia coli* K-12. *Mol Microbiol* **31:** 557-567

Madegowda M, Eswaramoorthy S, Burley SK, Swaminathan S (2008) X-ray crystal structure of the B component of hemolysin bl from *Bacillus cereus*. *Proteins-Structure Function and Bioinformatics* **71:** 534-540

Mel SF, Stroud RM (1993) Colicin Ia inserts into negatively charged membranes at low pH with a tertiary but little secondary structural change. *Biochemistry* **32:** 2082-2089

Mueller M, Grauschopf U, Maier T, Glockshuber R, Ban N (2009) The structure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism. *Nature* **459:** 726-730

Nicholls A, Sharp KA, Honig B (1991) Protein folding and association insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct Funct Genet* **11:** 281-296

Olson R, Nariya H, Yokota K, Kamio Y, Gouaux E (1999) Crystal structure of Staphylococcal LukF delineates conformational changes accompanying formation of a transmembrane channel. *Nat Struct Bio* **6:** 134-140

Parker MW, van der Goot FG, Buckley JT (1996) Aerolysin -The ins and outs of a model channel-forming toxin. *Mol Microbiol* **19:** 205-212

Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol* **7:** 95-99

Rhodes G (2006) *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models* London: Academic Press.

Rossjohn J, Feil S, McKinstry W, Tweten R, Parker M (1997) Structure of a cholesterol-binding, thiol-activated cytolysin and a model of its membrane form. *Cell* **89:** 685–692

Song LZ, Hobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux JE (1996) Structure of Staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **274:** 1859-1866

Sutton RB, Fasshauer D, Jahn R, Brünger AT (1998) Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature* **6700:** 347–353

Wallace AJ, Stillman TJ, Atkins A, Jamieson SJ, Bullough PA, Green J, Artymiuk PJ (2000) *E. coli* hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy. *Cell* **100:** 265-276

Weissenhorn W, Dessen A, Harrison SC, Skehel JJ, Wiley DC (1997) Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **387:** 426-430

Wiener M, Freymann D, Ghosh P, Stroud RM (1997) Crystal structure of colicin 1a. *Nature***:** 461–464

Yuan Z, Zhao J, Wang ZX (2003) Flexibility analysis of enzyme active sites by crystallographic temperature factors. *Protein Eng* **16:** 109-114

Chapter 8: NheB Preparation and Crystallization

This chapter will describe the work on the preparation, crystallization and attempts to solve the structure of the NheB component of the *B. cereus* Nhe toxin.

8.1 Expression of NheB from *Bacillus subtilis* strain JH642

Due to difficulties in the Oslo group in expressing the NheB in *E. coli* as a result of producing inactive protein (the reason for which was not clear, P. E. Granum and D. Phung' personal communication) the protein was instead expressed in *B. subtilis*.

B. subtilis strain JH642 was used as the host cell of expression vector pDG148-Stu (Vector pDG148-*nheB*), which was supplied by Prof. Per Einar Granum, Norwegian School of Veterinary Science in Oslo.

The primary culture was grown at 37 ˚C in LB media under the selection of kanamycin (5 μg/ml) with shaking (150 rpm) overnight.

20 litres of LB liquid media containing kanamycin (5 μg/ml) in 20 flasks were prepared. The primary culture was added to the secondary culture medium in a ratio of 3:100 (primary: secondary). Therefore 30 ml of primary culture was added to each 1000 ml of secondary culture medium. The cells were then incubated at 37 °C whilst being shaken at 150 rpm for 5-6 hours. When the cultures reached on optical density OD_{600} of 0.7 to 0.9, NheB expression was induced by the addition of IPTG to a final concentration of 200 μg/ml. The cells were left for a further 24 hours after induction. The culture was then spun in a cold centrifuge (4 °C) for 20 minutes at 15,344 g.

It was assumed that the protein would be released from the cell into the media due to the presence of signal peptide that acts as a signal for transportation of the protein outside *B. subtilis* (Figure 8.1).

The NheB was precipitated from the media using 65% $(NH_4)_2SO_4$ overnight at 4 ˚C (Figure 8.2 A). After centrifugation the pellet was collected at 4 ˚C (70000 g ; 15 minutes) and dissolved in 50 mM Tris HCl at pH 8. Finally, the sample was dialyzed twice against 1 L of the same buffer to remove $(NH_4)_2SO_4$.

Western immunoblotting was carried out to check for the presence of NheB in the media (as explained in chapter 3 in section 3.4.4), and it is clear from (Figure 8.2 B) that NheB was successfully expressed in LB media but with a low level of protein expression.

8.2 Purification of NheB from *Bacillus subtilis* strain JH642

Purification of NheB is a complicated process and ideal protocol for purification has not yet been found.

The following protocol was used to purify NheB from *B. subtilis* strain JH642, developed by Lindbäck et al, (2004). The dialyzed sample was loaded onto a DEAE sepharose fast flow ion exchange column pre-equilibrated with loading buffer (50 mM Bis Tris at pH 8). Un-bound proteins were washed from the column using the same buffer. The bound proteins were eluted using a linear gradient of increasing 0–1 M NaCl in buffer. Most of the NheB was present in the unbound fractions (Figure 8.3).

The unbound fractions that contained the largest amount of NheB were selected, and cut with 1.5 M $(NH_4)_2SO_4$ and then the supernatant $(NH_4)_2SO_4$ was collected at 4˚C (15,344 g ; 20 minutes) and put through a 5 ml Phenyl Hp column pre-equilibrated with 25 mM Tris-HCl at pH8. The NheB was eluted using a linear gradient of increasing 1.5 M $(NH_4)_2SO_4$ buffer.

```
        10         20         30         40         50         60
MTKKPYKVMA LSALMAVFTA GNIMPAHTYA AESTVKQAPV HAVAKAYNDY EEYSLGPEGL

        70         80         90        100        110        120
KDAMERTGSN ALVMDLYALT IIKQGNVNFG NVSSVDAALK GKVIQHQDTA RGNAKQWLDV

       130        140        150        160        170        180
LKPQLISTNQ NIINYNTKFQ NYYDTLVAAV DAKDKATLTK GLTRLSSSIN ENKAQVDQLV

       190        200        210        220        230        240
EDLKKFRNKM TSDTQNFKGD ANQITSILAS QDAGIPLLQN QITTYNEAIS KYNAIIIGSS

       250        260        270        280        290        300
VATALGPIAI IGGAVVIATG AGTPLGVALI AGGAAAVGGG TAGIVLAKKE LDNAQAEIQK

       310        320        330        340        350        360
ITGQVTTAQL EVAGLTNIKT QTEYLTNTID TAITALQNIS NQWYTMGSKY NSLLQNVDSI

       370        380        390        400
SPNDLVFIKE DLNIAKDSWK NIKDYAEKIY AEDIKVVDTK KA
```

Figure 8. 1: **NheB amino acid sequence**.
The blue sequence colour shows a signal peptide, which functions as a signal for transportation of the protein. The bacterial cells cut off this signal peptide during expression, which means it is not a part of the mature NheB protein after purification.

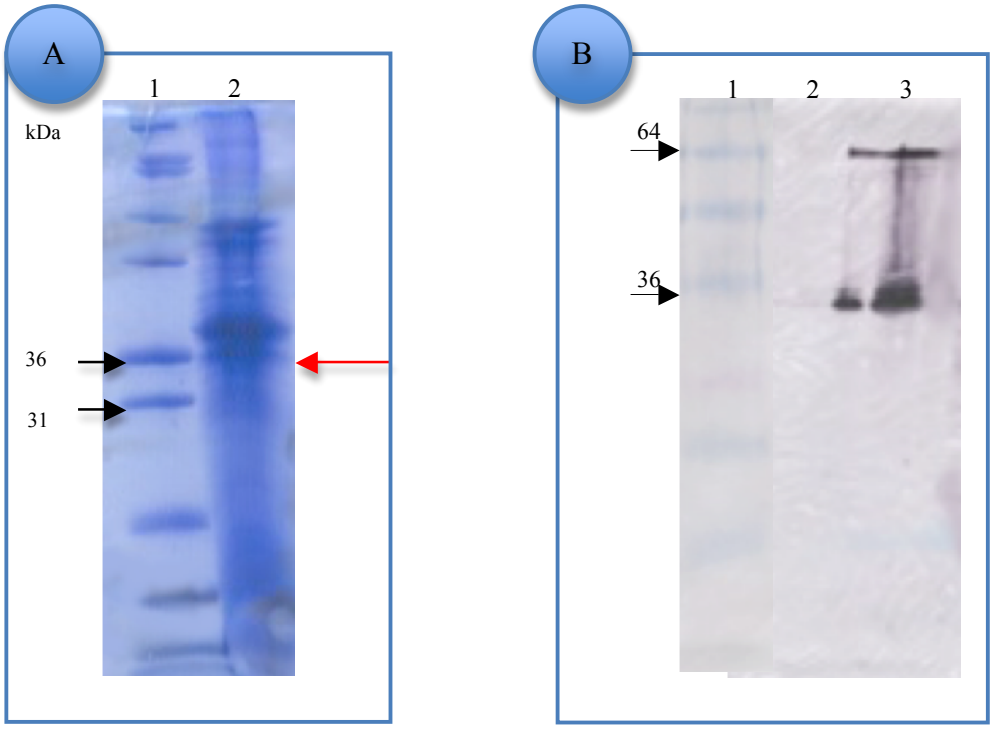


Figure 8.2: The expression of NheB protein in *B. subtilis* strain JH642.
(A) 12% SDS-PAGE analysis of NheB expression visualised by staining with Coomassie blue stain. Lane 1: Marker 12 from top to bottom: 200, 116, 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: induction of NheB by 200 µg/ml of IPTG in LB media for 24 h. The molecular weight of the protein should be around 36 kDa (red pointer).
(B) Western immunobloting of 12% SDS-PAGE analysis of expression for NheB visualised by the colorimetric kit method. The primary antibody MAb 1C2 was used. Lane 1: Marker SeeBlue® Plus2 Pre-Stained Standard from top to bottom: 98, 64, 50, 36, 22, 16, and 6 kDa; Lane 2: the cell extract after induction of NheB by 200 µg/ml of IPTG in LB for 24 h; and Lane 3: the media after induction of NheB by 200 µg/ml of IPTG in LB for 24 h. It is clear that the protein is secreted in the media as it was expected.

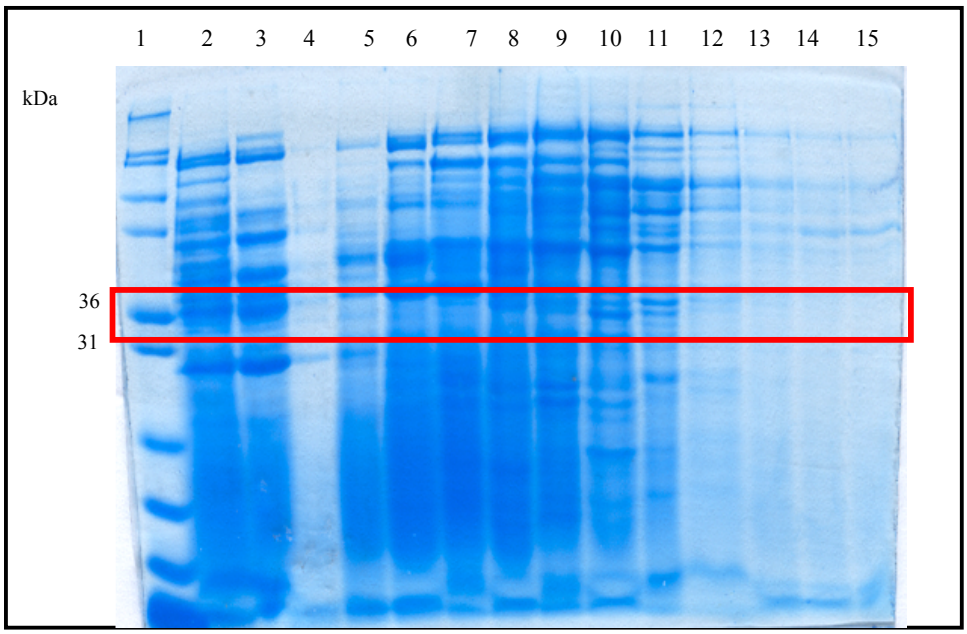


Figure 8.3: **12% SDS-PAGE analysis of DEAE sepharose column purification for NheB from *B. subtilis* strain JH642**.
Visualised by staining with Coomassie blue stain. Lane1: Marker12 from top to bottom: 200, 116, 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: DEAE sepharose column load; Lane3: unbound proteins; Lane 4: to lane 15 the fractions collected from the DEAE sepharose column.

It can be seen from 12% SDS PAGE analysis the NheB is present in fractions 14 to 17 (Figure 8.4). The fractions were analyzed by Western Blot analysis, which confirmed the existence of the protein in those fractions (Figure 8.5).

NheB fractions after the Phenyl HP column were applied to a calibrated Hi load superdex200 gel filtration column pre-equilibrated with 50 mM Tris HCl, at pH 8, and 0.5 M NaCl (Figure 8.6).

8.3 Properties of pure NheB protein from *Bacillus subtilis* strain JH642

Purified NheB had a band molecular mass of 36 kDa and was estimated to be ≥85% pure and heterogeneous by SDS-PAGE. There is a monomer and a dimer form of the protein (Figure 8.6). Western Blot analysis of the sample confirmed it was NheB and it has two forms (Figure 8.7).

Protein concentration was estimated with the Bio-Rad protein assay protocol (Bradford method) and the concentration of the protein was performed using a 10 kDa molecular-weight cutoff Centricon centrifugal ultrafilteration device (Vivascience). NheB was concentrated to 8.5 mg/ml before crystallization trials.

8.4 3D crystallization trials of NheB

3D crystallization trials for NheB were preformed manually via vapour diffusion methods (sitting drop). Different crystallization screens from Hampton Research, Jena Bioscience and QIAGEN screens were utilized at 7 ºC and 17 ºC with protein concentration (8.5 mg/ml).

Small protein needles were observed in condition H10 of the JCSG [25% PEG500, 100 mM Bis-Tris at pH 5.5 and 200 mM CH_3COONH_4] after two days (Figure 8.8).

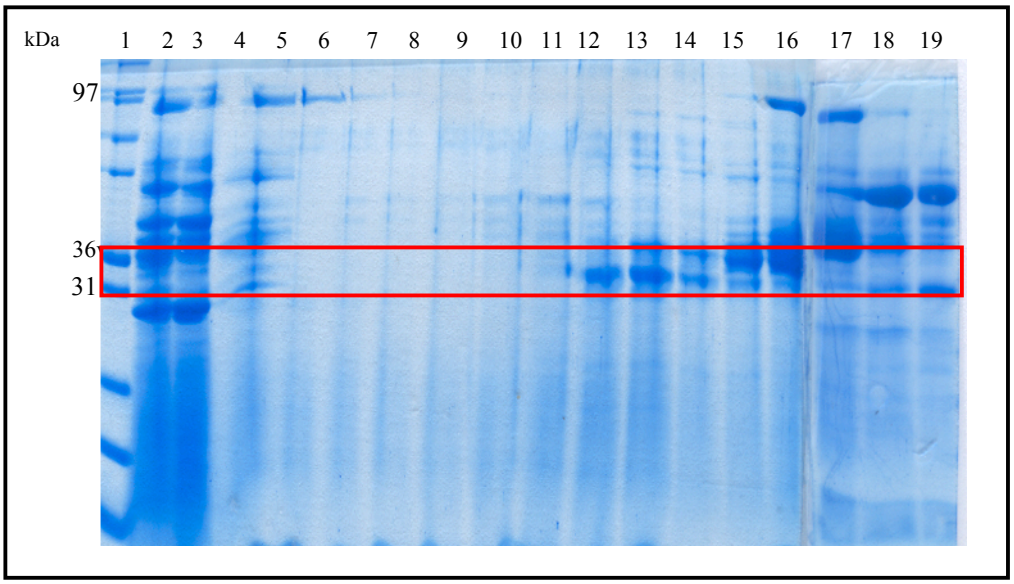


Figure 8.4: **12% SDS-PAGE analysis of a Phenyl HP column purification for NheB from *B. subtilis* strain JH642.**
Visualised by staining with Coomassie blue stain. Lane 1: Marker12 from top to bottom: 200, 116, 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: sample after DEAE sepharose; Lane 3: 1.5 M $(NH_4)_2SO_4$ cut ; Lane 4: Phenyl HP column load; Lane 5: to lane 19 the fractions collected from Phenyl HP column.

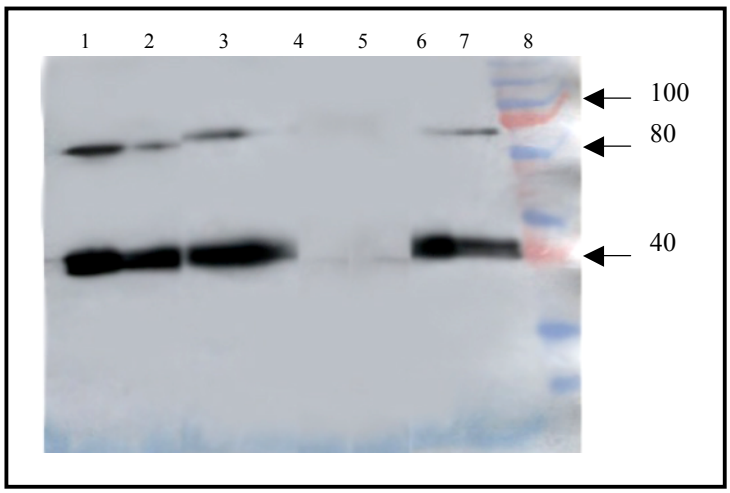


Figure 8.5: Western immunobloting of 12% SDS-PAGE analysis of DEAE sepharose column and a Phenyl HP column purification for NheB from *B. subtilis* strain JH642 visualised by the chemiluminescence method.

The primary antibody MAb 1C2 was used.

Lane 1: fraction 27 of Phenyl HP column; Lane 2: fraction 22 of Phenyl Hp column; Lane 3: Phenyl HP column load; Lane 4: fraction 11 of DEAE sepharose; lane 5: fraction 17 of DEAE sepharose; Lane 6: unbound proteins of DEAE sepharose; Lane 7: DEAE sepharose load; Lane 8: Marker Nexus View 20 kDa Dual Color from top to bottom: 140, 120, 100, 80, 60, 40, 30, and 20 kDa.

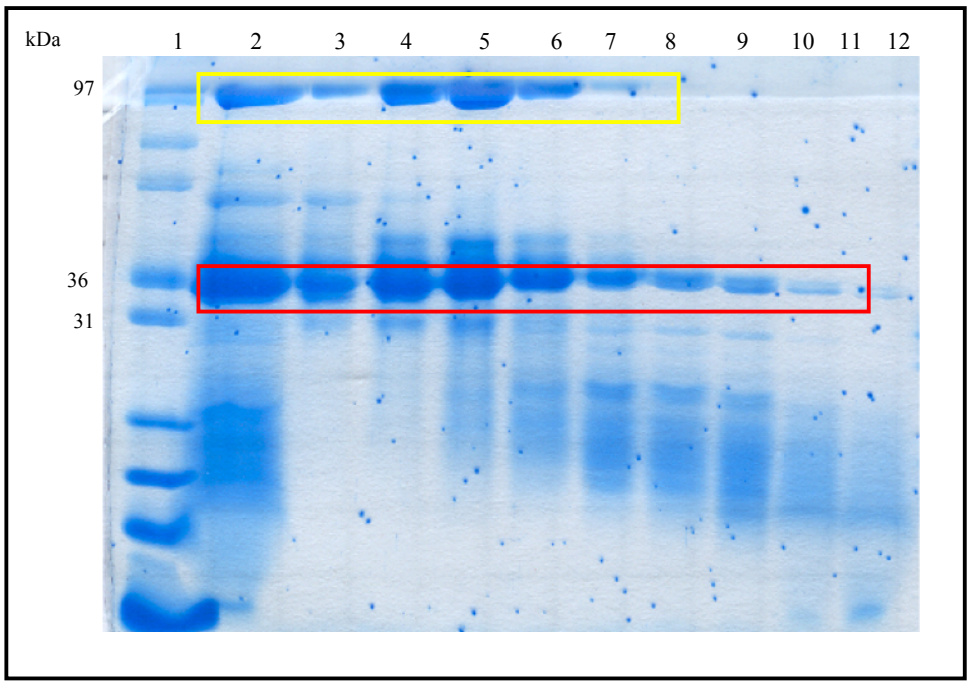


Figure 8.6: **12% SDS-PAGE analysis of gel filtration column purification for NheB from *B. subtilis* strain JH642.**
Visualised by staining with Coomassie blue stain. Lane 1: Marker12 from top to bottom: 116, 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: Gel filtration load; Lane 3: fraction 3 to 12 the fractions collected from gel filtration column. Red rectangle represents monomer form of NheB and yellow rectangle represents dimmer form of NheB.

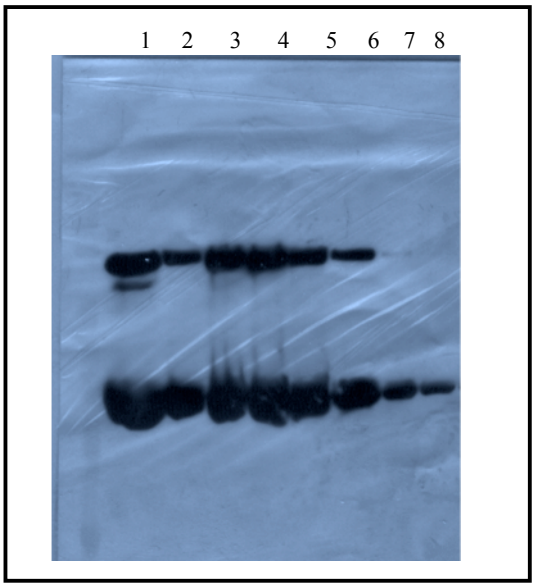


Figure 8.7: Western immunobloting of 12% SDS-PAGE analysis of gel filtration column purification for NheB from *B. subtilis* strain JH642 visualised by chemiluminescence method. The primary antibody MAb 1C2 was used. Lane 1: gel filtration load; Lane 2 to 8 fractions from gel filtration NheB.

This condition was optimized using the sitting-drop vapour-diffusion method in 24-well plates (Hampton Research). Protein drops were made up of 1 µl protein solution (8.5 mg/ml) mixed with 1 µl of various precipitants with varying PEG3350 concentration and pH.

Further extensive trials were undertaken and finally rod crystals were obtained which were of sufficient size (with dimensions of about 0.01 µm x 0.01 µm x 0.3 µm) to attempt to collect data on. These crystals were obtained from optimized conditions 22% PEG500, 100 mM Bis-Tris at pH 7.5 and 200 mM CH_3COONH_4, and took around one week to grow. An image of the crystals is shown in (Figure 8.9). Unfortunately no better crystals than these were grown for NheB.

As can be seen from the 12 % SDS-PAGE analysis of the gel filtration column purification for NheB (Figure 8.6) the protein was not completely pure and to make sure that the crystal was NheB and not a contaminating protein, further experiments were carried out as summarized in Figure 8.10. These involved washing the crystals, dissolving them in water followed by 12% SDS-PAGE and western blotting analysis.

8.5 Data collection

The crystals were soaked in mother liquor supplemented with 25% ethylene glycol, as cryo-testing in house had indicated that 25% ethylene glycol was a potential cryoprotectant for the crystals, and then the crystals were flash-cooled in liquid nitrogen. The crystal was tested for diffraction on our in-house source at 100 K using the Rigaku X-ray generator with a Mar345 detector. Data collection showed no diffraction, but this was potentially promising as it showed these crystals were not salt.

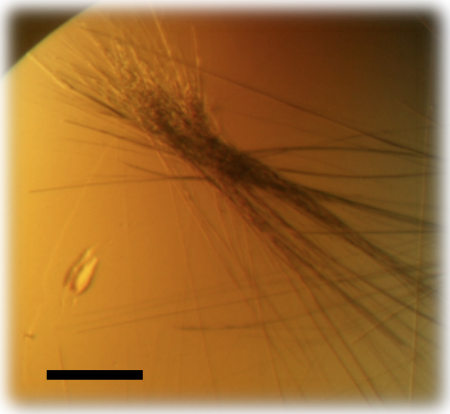


Figure 8.8: First hit for NheB crystals. The needles grow in 25% PEG500, 100 mM Bis-Tris at pH5.5 and 200mM CH_3COONH_4, NheA 8.5 mg/ml, 17 ˚C. Scale bar: 100 µm

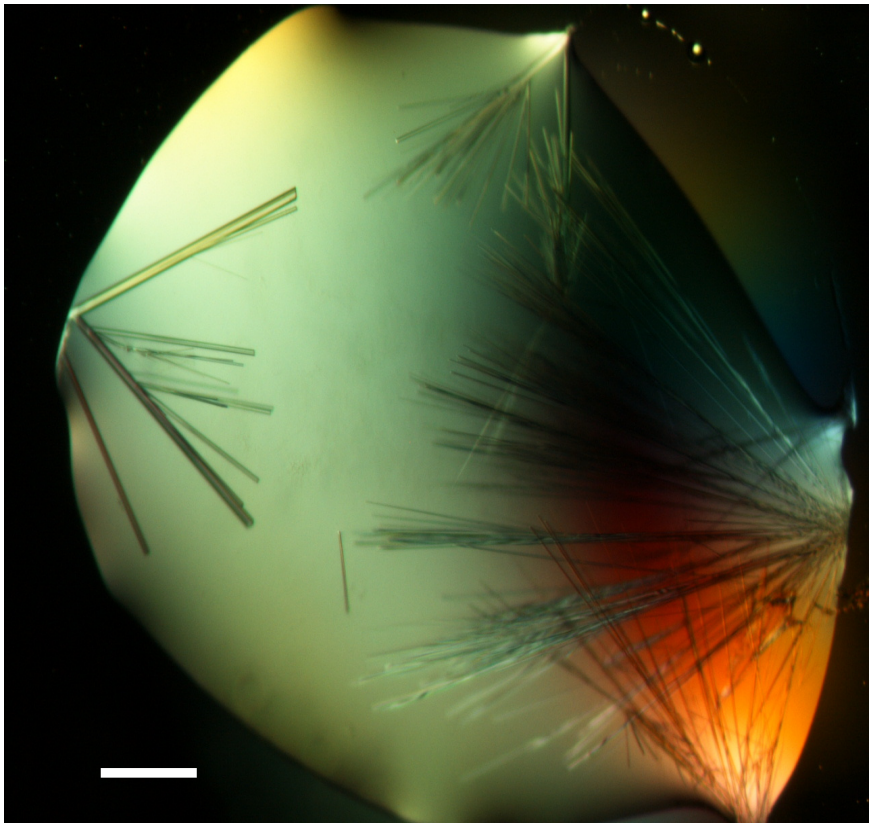


Figure 8.9: A cluster of crystals of NheB grown from 22% PEG500, 100 mM Bis-Trisat pH7.5 and 200 mM CH_3COONH_4, NheB 8.5 mg/ml, 17 ˚C. Scale bar: 100 µm.

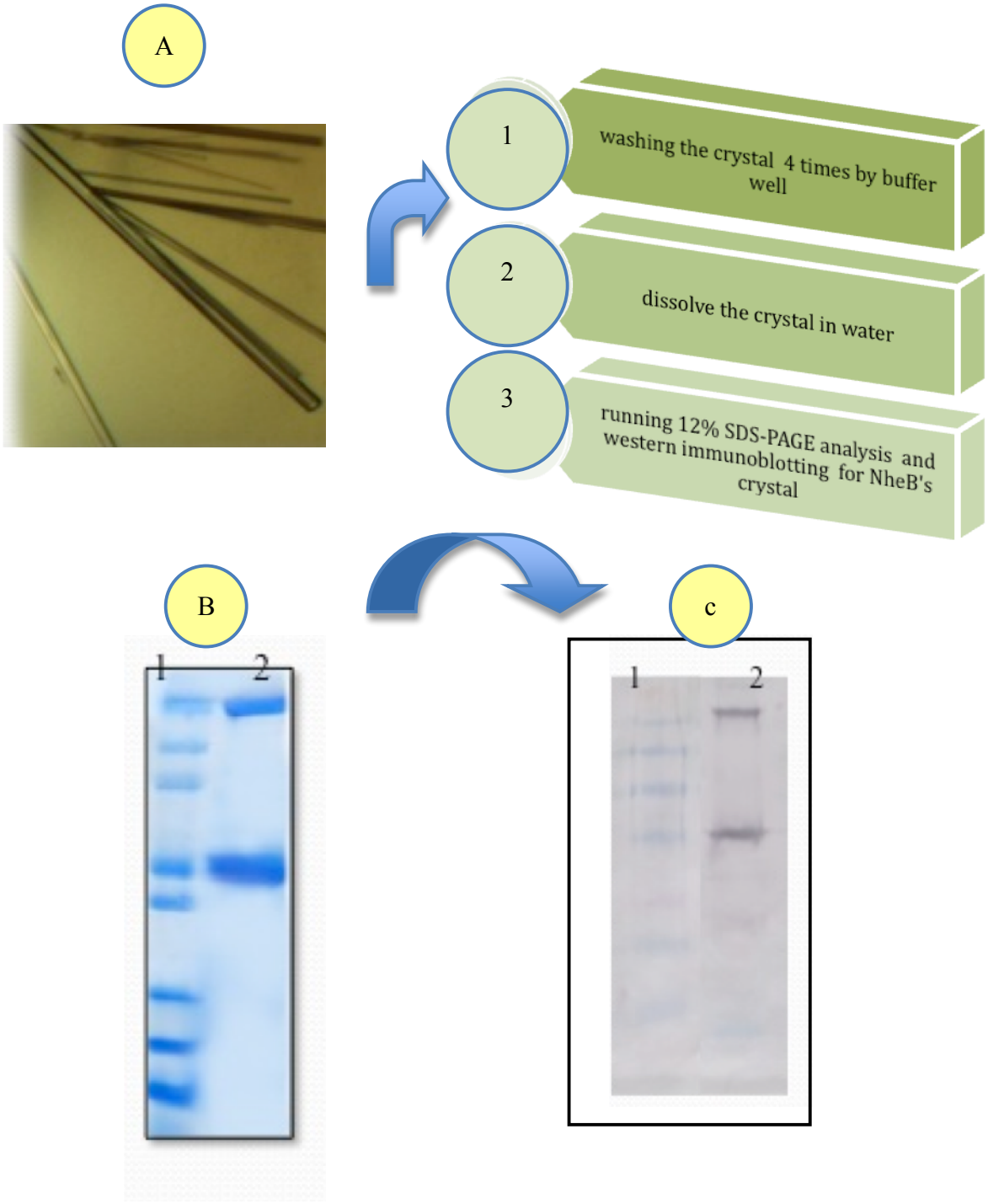


Figure 8.10: Summary of further experiment to make sure that the crystal was NheB.
(**A**) The protein crystal.
(B) 12% SDS-PAGE analysis of crystal visualised by staining with Coomassie blue stain. Lane 1: Marker12 from top to bottom: 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: NheB's crystal.
(C) Western immunobloting of 12% SDS-PAGE analysis of crystal visualised by colorimetric kit method. The primary antibody MAb 1C2 was used. Lane 1: Marker SeeBlue® Plus2 Pre-Stained Standard from top to bottom: 98, 64, 50, 36, 22, 16, and 6 kDa; Lane 2: NheB's crystal. It is clear that the crystal is of NheB.

The NheB crystal was stored in liquid nitrogen, and then tested with synchrotron X-ray beams at the Diamond Light Source where it diffracted to a final resolution of 3.5 Å (Figure 8.11). The diffraction images were collected on an ADSC Q315 CCD detector. A total of 240 images were collected with an oscillation angle of 0.5°. The exposure time was 0.8 second per image with the detector distance set at 457.7 mm.

8.6 Space group determination, indexing and data processing

Auto-indexing of NheB crystal was performed at the Diamond Light Source using the XDS program (Kabsch, 2010) using two images separated by 90°, with each image a 0.5° oscillation. The auto-indexing of NheB indicated that the crystal belongs to the hexagonal system and space group $P6_2$ or $P6_4$. Systematic absences were observed on all 00L reflections where h, k, l = 3n. Pseudo-precession pictures covering h, k, and l axes from NheB crystal are revealed in Figure 7.12 A, B, and C.

The cell dimensions of the NheB crystal are *a*= 130.9 Å, *b*= 130.9 Å, *c*= 61.9 Å, α= 90°, β= 90°, and γ= 120°. The data were also processed at the Diamond Light Source using XDS (Kabsch, 2010). This data had an overall R_{merge} of 0.14 with a value of 0.61 in the outermost resolution shell. The data were 99.7% complete with an I/σ of 10.4, overall. The statistics for X ray data processing are reported in (Table 8.1).

S`8.7 Matthews number (*Vm*) calculation in space group $P6_2$

The Matthews coefficient was calculated using the web page http://www.ruppweb.org/Mattprob/ (Kantardjieff & Rupp, 2003). This indicated the most probable number of NheB molecule in each a asymmetric unit is around two molecules, giving a crystal volume per protein mass (V_M) of 2.07 $Å^3$/Dalton and a solvent content of 44.55% (Matthews, 1968).

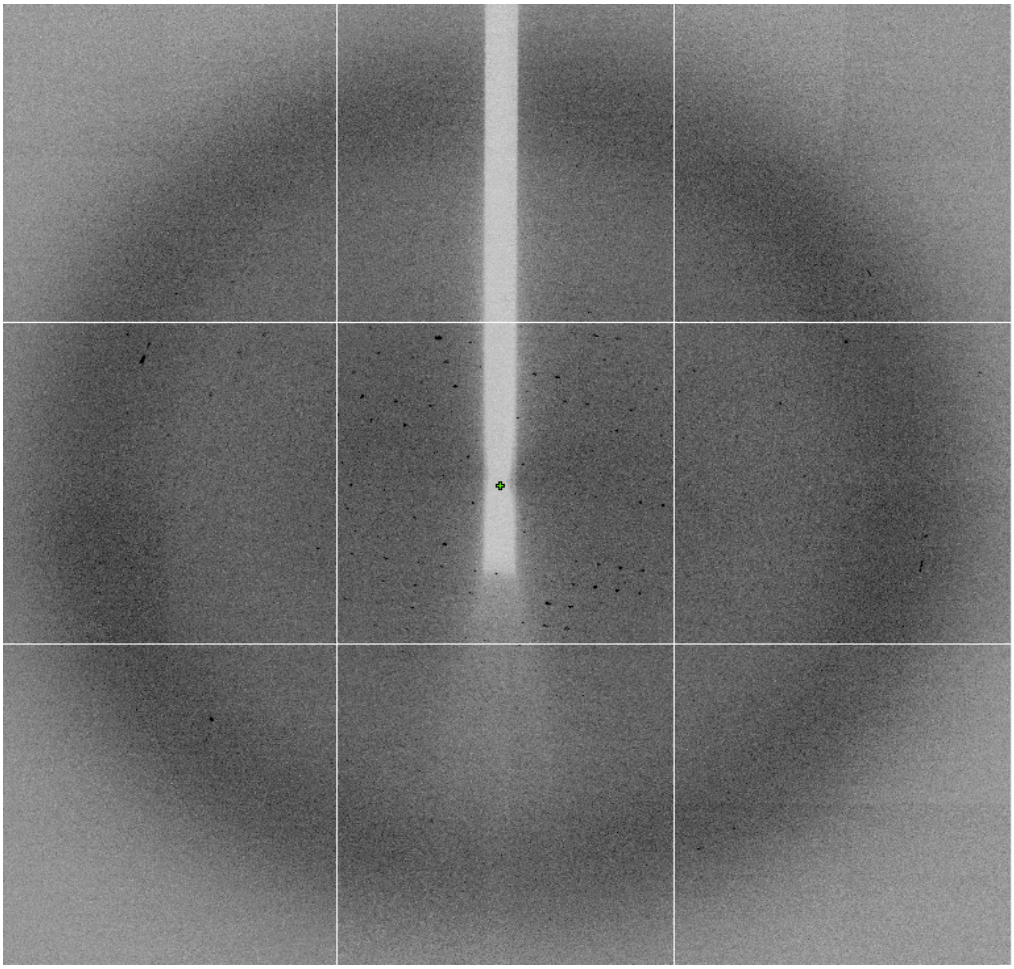


Figure 8.11: X-ray diffraction pattern of NheB crystal using beamline I02 of Diamond light source.
A dataset was collected to 3.5Å resolution on the Diamond synchrotron. The exposure time was 0.8 s, the crystal-to-detector distance was 457.7 mm, wavelength 0.9796 Å and the oscillation range per image was 0.5°. The diffraction image was collected on an ADSC Q315 CCD detector

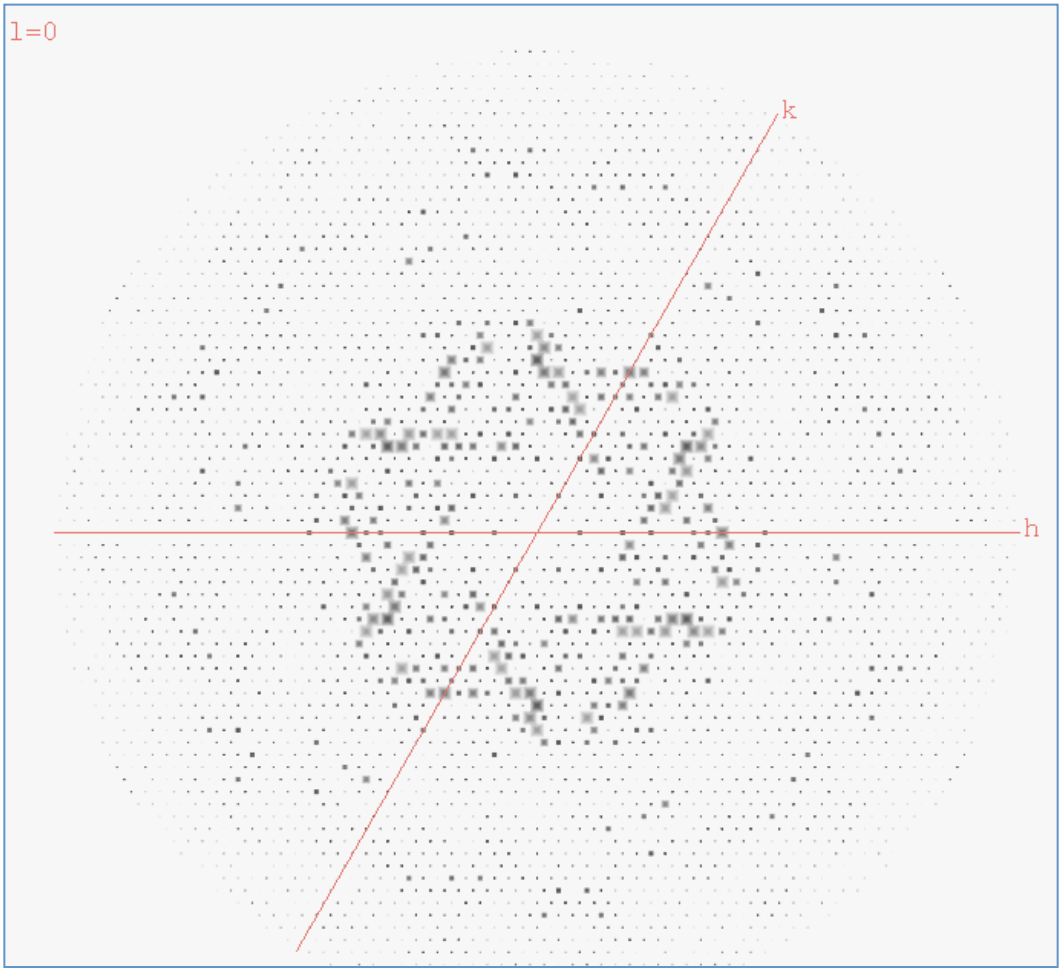


Figure 8.12 A: hkl view pseudo-precession image of zone hk0 from the NheB crystal.

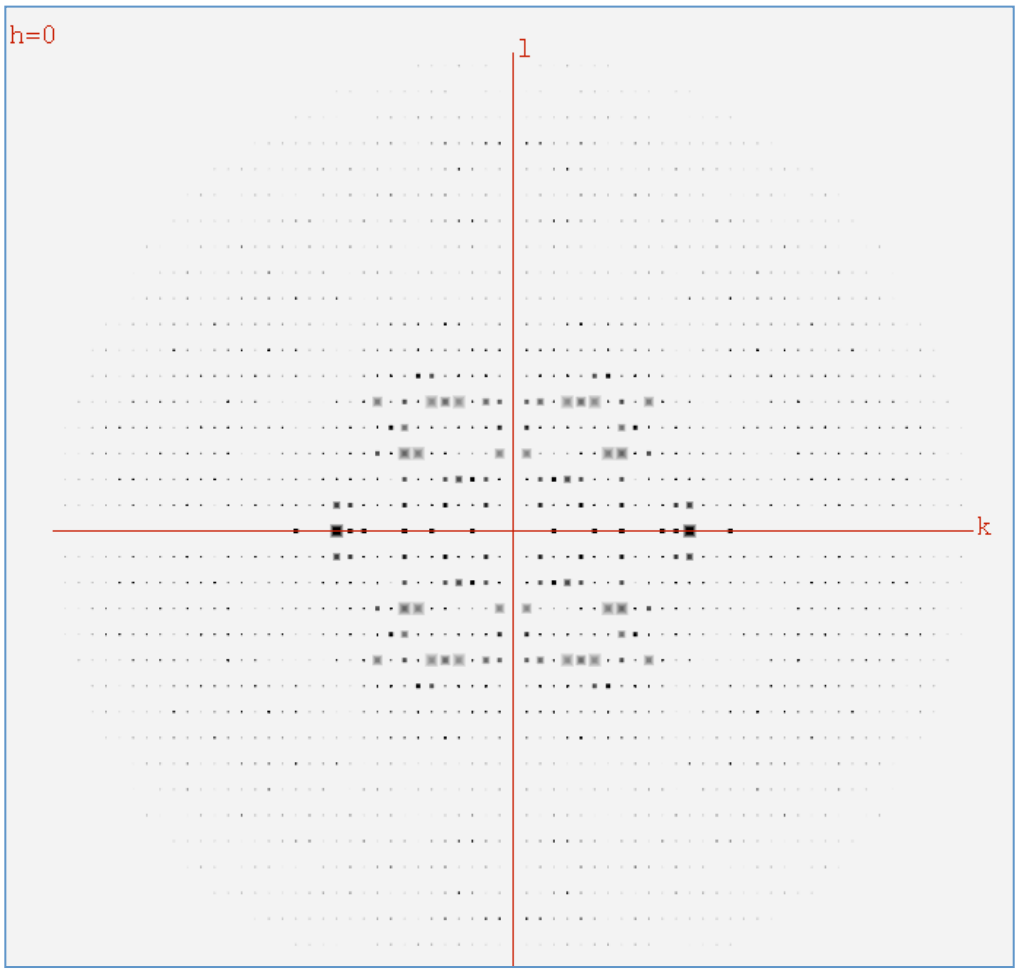


Figure 8.12 B: hkl view pseudo-precession image of zone 0kl from the NheB crystal. See more details view in Figure 8.12 C.

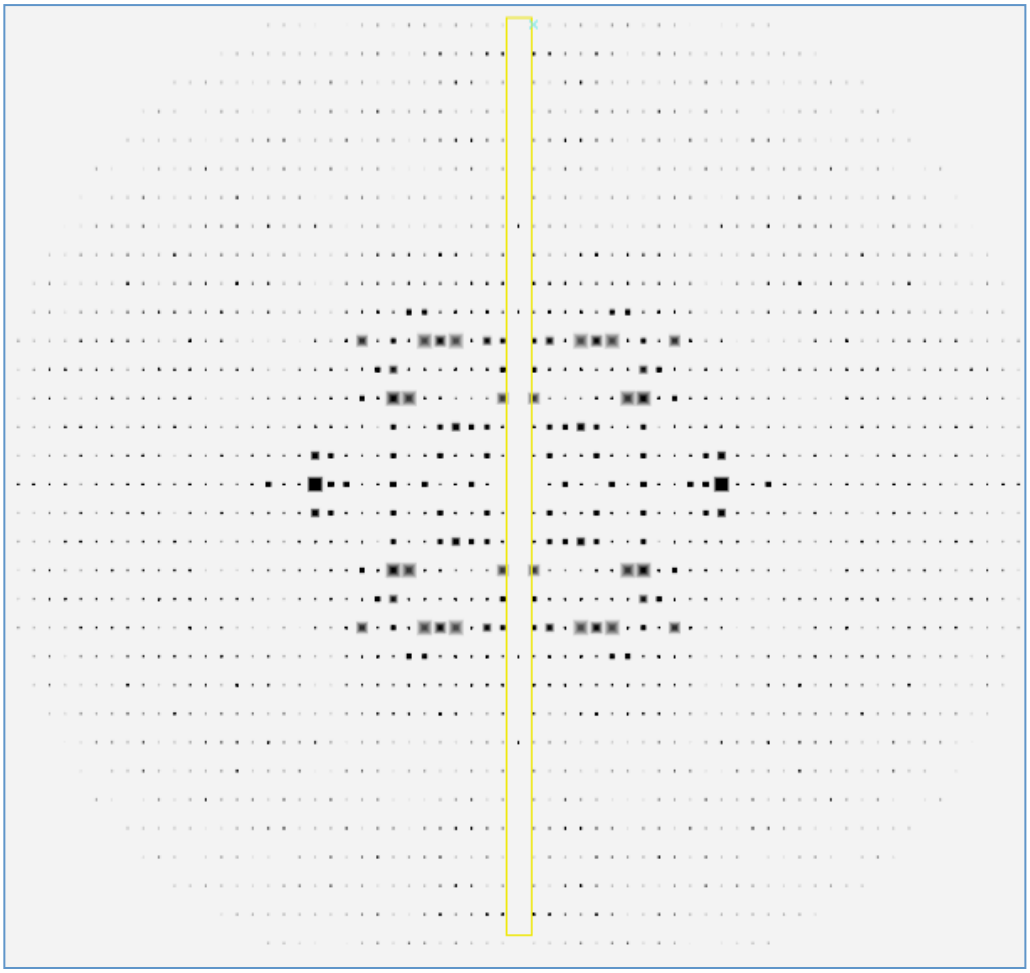


Figure 8.12 C: Close up of hkl view pseudo-precession image of zone 0kl from the NheB crystal.
The 00l axis is boxed and the only 006 and 009 are visible, corresponding to l= 3n.

To investigate the relationships of NheB molecules in the asymmetric unit, the self-rotation function using POLARRFN (CCP4: Supported Program) was run. The self-rotation function shows that there is no evidence for non-crystallographic symmetry in 6-fold axis or 3-fold axis. However, in 2-fold axis there is appears to be additional pseudo-non-crystallographic symmetry (marked as X in figure 8.13), which more related to *$P6_222$* space group or *$P6_422$* space group. This result also confirms that there are two monomers of NheB in the asymmetric unit. However, examination of a self-Patterson synthesis shows there is no evidence for pseudo-translational symmetry in the NheB crystal.

8.8 Attempts at NheB structure determination by molecular replacement

The programme of choice for molecular replacement was PHASER (CCP4: Supported Program) (McCoy et al, 2007). The best solution in molecular replacement is indicated by: the Log likelihood gain which is show how good the solution is compared to a random solution and the "Z-score" which shows the number of standard deviations above the mean value. The successful hit that indicates the best solution should have a high Log likelihood gain (> 30) and a high Z-score (> 5) (Gubbi et al, 2007; McCoy et al, 2007; McCoy et al, 2005; Storoni et al, 2004).

8.8.1 Molecular replacement in space group P6

Although the auto-indexing of NheB crystal showed that the crystal belongs to the hexagonal system and space group $P6_2$, for space group with 6-fold screw axes, the intensities do not give sufficient information to differentiate $6_1/6_5$, or $6_2/6_4$ axes. Therefore, It was decided to try to determine the structure for the NheB protein by molecular replacement in space group $P6_2$ or one of alternative space group (P6, $P6_1$, $P6_3$, $P6_4$ or $P6_5$) using the *B. cereus* HblB

molecule (Madegowda et al, 2008), which has 27% amino acid similarity to NheB, as a search model.

No satisfactory molecular replacement solution was found for NheB although the PHASER gives the single solution in space group $P6_2$. The Z-scores were 3.8, and 5.9 for the rotation and translation functions respectively. The overall log likelihood gain was 44.54. However, examination of electron density map calculated with the PHASER gave no useful additional information. The space group $P6_2$ was expected to be correct for NheB crystal from the systematic absences and so was likely to be give best solution.

8.8.2 Molecular replacement in space group P3, or $P3_1$, or $P3_2$

Because the MR failed in space group P6, it was suggested to try determination of the structure of NheB in the lower symmetry space group, which is P3, $P3_1$ or $P3_2$. The data were re-processed again at the Diamond Light Source using XDS (Kabsch, 2010) in space group P3. Before running the molecular replacement it was important to know the number of molecules that were likely to be present in the asymmetric unit of space group P3. This is estimated by the Matthews number (Matthews, 1968). The Matthews coefficient in space group P3 suggests that three or four molecules could be in the asymmetric, giving a crystal volume per protein mass (V_M) of 2.75 $Å^3$/Dalton and a solvent content of 55.20% (Matthews, 1968). The reprocessed data had an overall R_{merge} of 0.139 with a value of 0.65 in the outer resolution. The data were 99.7% complete with an I/σ of 7.7 overall. The statistics for X ray data processing are listed in (Table 8.2). The search model that was used was the same as that used for previous space group determinations - *B. cereus* HblB molecule (Madegowda et al, 2008). Unfortunately no convincing solution was suggested by PHASER in space group P3 or alternative space groups, $P3_1$ or $P3_2$ although the best solution (Z-score = 4.2 and 5.2 for the rotation and translation functions respectively)

was obtained for $P3_2$. However no additional information had been seen after examination of the electron density map that was calculated by PHASER.

The reasons for this disappointing result and failure of the molecular replacement is probably because no search model with high sequence similarity to NheB was available. The general rules to make molecular replacement work is that sequence similarity of the model to the target should be >30% (Rupp, 2010). But the sequence identities of NheA, HblB, and ClyA are only 22%, 27%, and 19% respectively. In addition the X-ray data are of limited resolution, which does not add the interpretation the map and refinement.

In the future it is hoped were to collect high-resolution data for NheB crystals or to find an alternative model with higher sequence similarity.

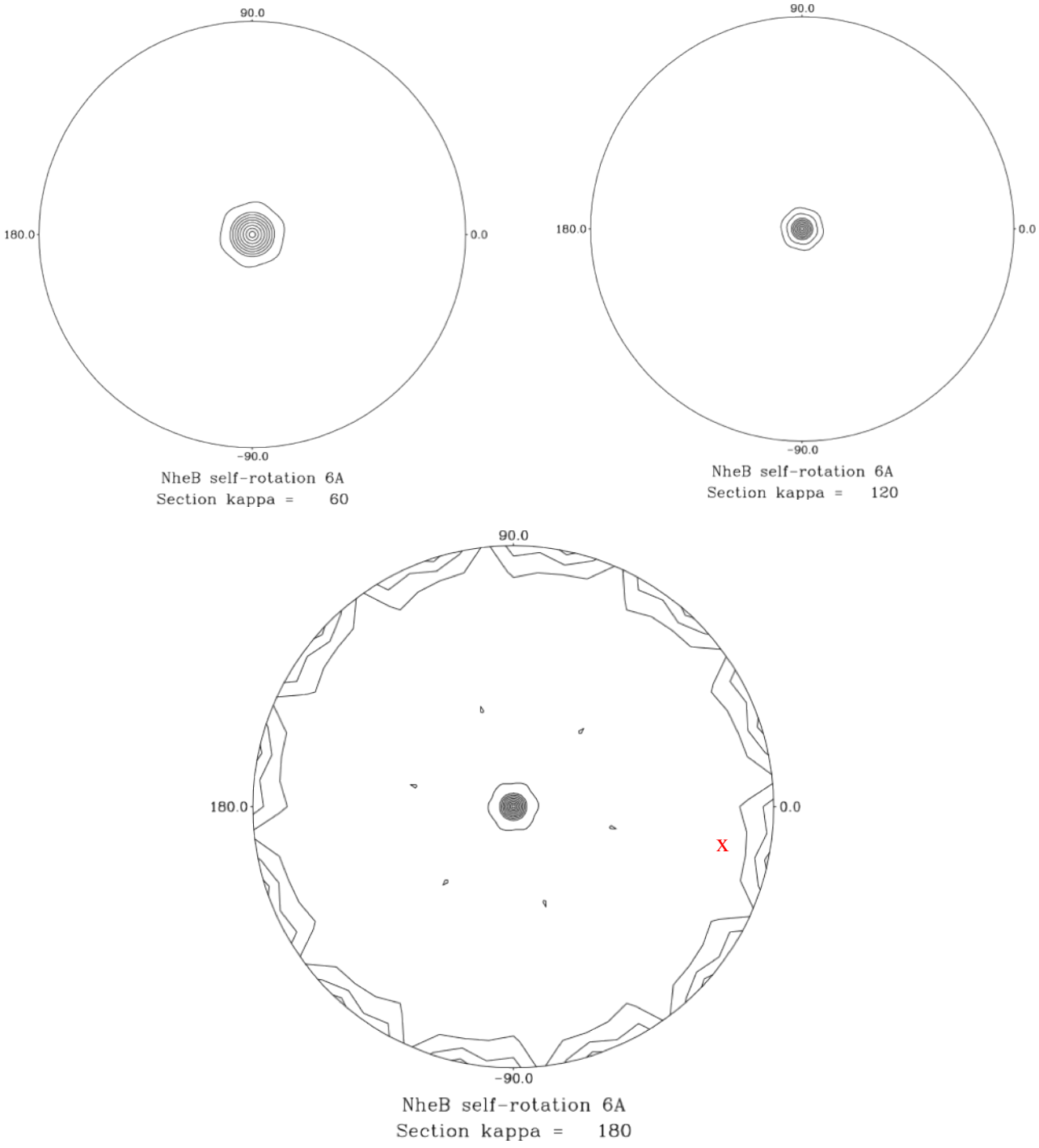


Figure 8. 13: Self-rotation function of NheB that was calculated using the program POLARRFN (CCP4: Supported Program). There is no evidence for non-crystallographic symmetry in 6- fold axis and 3- fold axis. There are 2-fold axes for pseudo- non-crystallographic symmetry (X labeling) that more related to $P6_222$ space group or $P6_422$ space group.

Table 8.1: Data collection statistics from the NheB crystal processed in space group P6.

Space group	**P6**		
Unit-cell parameters (Å, °)	*a*= 130.89 *b*= 130.89 *c*= 61.86 α=90 β=90 γ=120		
Crystal system	Hexagonal		
	Overall	**Inner Shell**	**Outer Shell**
Low resolution limit (Å)	65.44	65.44	3.43
High resolution limit (Å)	3.35	14.97	3.35
R_{merge}	0.140	0.045	0.614
Total No. of observations	66236	727	5140
Total No. Unique	8903	112	671
Mean ((I)/sd(I))	10.4	22.0	3.1
Completeness	99.7	96.6	99.8

Table 8.2: Data collection statistics from the NheB crystal processed in space group P3.

Space group	**P3**		
Unit-cell parameters (Å, °)	*a*= 130.69 *b*= 130.69 *c*= 61.76 α=90 β=90 γ=120		
Crystal system	trigonal		
	Overall	**Inner Shell**	**Outer Shell**
Low resolution limit (Å)	65.35	65.35	3.38
High resolution limit (Å)	3.29	14.73	3.29
R_{merge}	0.139	0.044	0.654
Total No. of observations	68831	741	5165
Total No. Unique	17776	195	1295
Mean ((I)/sd(I))	7.7	20.2	2.0
Completeness	99.0	96.2	100.0

8.9 Attempts to produce higher quality NheB crystals

To obtain higher quality NheB crystals in order to acquire higher resolution data, we need more optimization of the NheB conditions. Therefore, extra experiments of expression and purification of NheB were carried out. However, we noticed that the level of NheB expression from *Bacillus subtilis* strain JH642 decreased and became very low which did not allow to us to carry out the purification step. As the time for this project was running out, it was decided to make a trial to produce NheB from the pathogenic native strain, *Bacillus cereus* strain 0075/95.

8.9.1 Expression of NheB from a native strain (*Bacillus cereus* strain 0075/95)

8.9.1.1 Growth of *Bacillus cereus* strain 0075/95

Our collaborator Danh Phung in Norway carried out the growth of the pathogenic *B. cereus* strain 0075/95, as our laboratory does not have facilities to grow pathogenic bacteria (D. Phung, personal communication).

The cells were plated out from 70% glycerol stock of *B. cereus* strain 0075/95 onto a Petri dish and incubated overnight at 37 °C. A single colony was transferred to a sterile flask with 50 ml LB liquid media and 0.4% glucose. The primary culture was grown at 32 °C whilst being shaken at 150 rpm overnight.

20 litres of LB liquid media in 20 flasks were prepared. The primary culture was added to the secondary culture medium in a ratio of 3:100 (primary: secondary). Therefore 30 ml of primary culture was added to each 1000 ml of secondary culture medium. The cells were then also incubated at 32 °C whilst being shaken at 100 to 150 rpm for 5-6 hours. The culture was then spun in a cold centrifuge (4 °C) for 20 minutes at 15,344 g.

The supernatant was transferred to a large flask and precipitated with 65% $(NH_4)_2SO_4$ in a cold room overnight. The protein was kept in 65% $(NH_4)_2SO_4$ and sent by post to our laboratory in Sheffield to carry out purification experiments.

8.9.2 Trial purification of NheB from native strain, *Bacillus cereus* strain 0075/95

The precipitated media were spin down at 4 °C for 15 minutes at 70000 g, and the pellet dissolved in 50 mM Tris-HCl buffer at pH 8. Then the sample was dialyzed twice against 3L of the same buffer to remove $(NH_4)_2SO_4$.

The protocol for purification of NheB from native strain, *Bacillus cereus* strain 0075/95, was essentially the same as the protocol for purification of NheB from *Bacillus subtilis* strain JH642 (section 8.2).

The sample was applied onto a DEAE sepharose fast flow ion exchange column pre-equilibrated with loading buffer (25 mM Bis Tris at pH 8.0). It was assumed the protein would be found in the un-bound fractions, the un-bound proteins were collected and loaded onto a 5ml Phenyl Hp column pre-equilibrated with 25 mM Tris-HCl at pH 8. The NheB was eluted using a linear gradient of increasing 1.5 M $(NH_4)_2SO_4$ buffer. The fewer contaminant fractions were applied to a calibrated Superdex200 gel filtration column and the final purification step is shown by 12% SDS PAGE in Figure 8.14.

However, the purification of NheB from the native, *Bacillus cereus* strain 0075/95 had very little success and the NheB obtained was not of a sufficient purity or quantity for more crystallization trials.

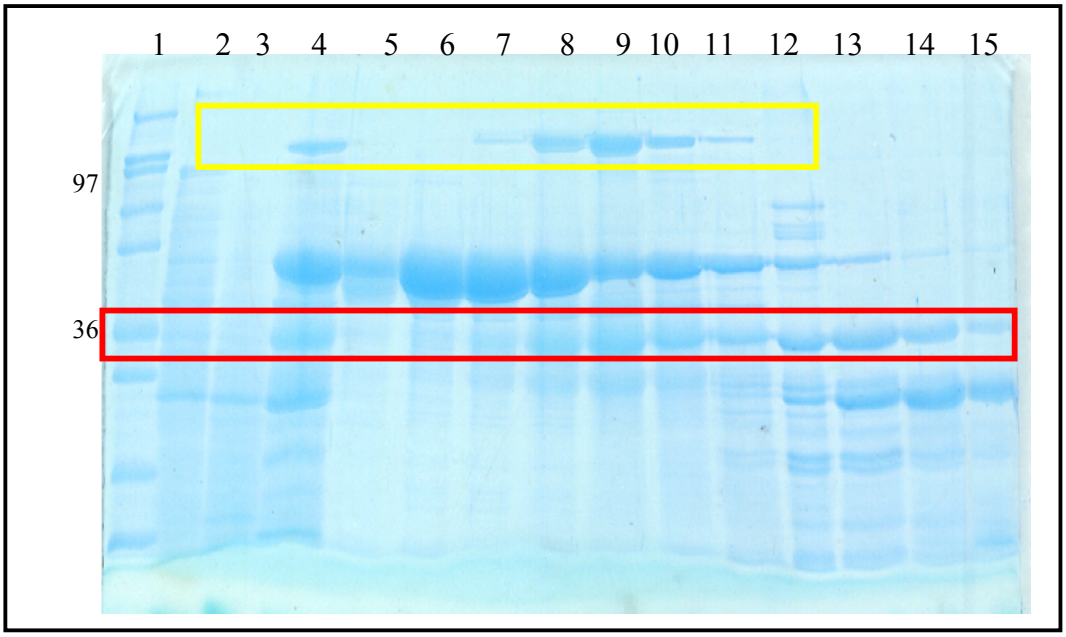


Figure 8. 14: 12% SDS-PAGE analysis of purification for NheB from native strain, *B. cereus* strain 0075/95.
Visualised by staining with Coomassie blue stain. Lane 1: Marker12 from top to bottom: 200, 116, 97, 66, 55, 36, 31, 21, 14 and 10 kDa; Lane 2: DEAE loading; Lane 3: Sample after DEAE; Lane 4: Sample after Phenyl Hp; Lane 5 to lane 15: fraction 16 to 26 the fractions collected from gel filtration column. Red rectangle represents monomer form of NheB and yellow rectangle represents dimer form of NheB.

8.10 Future work on NheB protein

It is clear that, the knowledge of the crystallization conditions for NheB obtained during this work should be useful in future. However, the present data for NheB, which was collected to 3.5 Å, are of relatively low resolution and therefore it is important that more NheB crystals are obtained to provide high-resolution data.

If the molecular replacement fails, there are two methods that may help to solve the phase problem, which are MIR or MAD. In general, incorporation of SeMet into protein, followed by structure determination using MAD data set could be easier than MIR method. However given the difficulties involved in obtaining native protein, and the lack of protocol for SeMe incorporation in *B. subtilis* or *B. cereus,* selenomethionine incorporation could be very challenging. If however this not works then heavy atom soaking may work, due to the huge range of heavy atoms that are available to the researcher. Unfortunately, there was not sufficient time to attempt to solve the structure of NheB by these methods during this project.

References

Gubbi J, Parker M, Palaniswami M (2007) Solving protein structures using molecular replacement via protein fragments. In *Applications of Fuzzy Sets Theory*, Masulli FMSPG (ed), Vol. 4578, pp 627-634.

Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66:** 125-132

Kantardjieff KA, Rupp B (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12:** 1865-1871

Lindbäck T, Fagerlund A, Rodland MS, Granum PE (2004) Characterization of the *Bacillus cereus* Nhe enterotoxin. *Microbiology-SGM* **150:** 3959-3967

Madegowda M, Eswaramoorthy S, Burley SK, Swaminathan S (2008) X-ray crystal structure of the B component of hemolysin bl from *Bacillus cereus*. *Proteins-Structure Function and Bioinformatics* **71:** 534-540

Matthews BW (1968) Solvent content of protein crystals. *Mol Biol* **33:** 491-497

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40:** 658-674

McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **61:** 458-464

Rupp B (2010) *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, New York: Garland Science.

Storoni LC, McCoy AJ, Read RJ (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr D Biol Crystallogr* **60:** 432-438

Chapter 9: Electron Microscopy Studies

Electron microscopy is an important tool, which can provide essential low- and medium-resolution information on the size and quaternary structure of membrane proteins stabilized with detergent or in a lipid membrane. This Chapter describes the results from the electron microscopy studies on all the proteins, APEC ClyA, NheA and NheB, which have been worked on during this project. In addition I present here the first observations of the bacterium *B. cereus* strain NHV 0075/95 by scanning electron microscopy (SEM) and transmission electron microscopy (TEM). In this work I am grateful for the help and guidance of Prof. Per Bullough and Dr Svetomir Tzokov.

9.1 Electron microscopy of biological macromolecules

Electron microscopy (EM) is a powerful tool to analyse the ultra-structure of cells and tissues. Over the last three decades, electron microscopy has also been used to elucidate the structures of biological macromolecules. Electron crystallography is a term that describes the use of electron diffraction data and electron microscope images to determine the structure 2D crystals of biological macromolecules (Glaeser et al, 2007).

Electron microscopy or crystallography is useful for the structure determination of two-dimensional (2D) crystals of membrane proteins and for single particle analysis of proteins. Indeed electron microscopy is the only technique that has produced atomic structures of membrane proteins while embedded in their native biological lipids. In addition to the ability to determine membrane protein structures while contained in a native-like environment, electron microscopy can also measure phases, which are normally lost during X-ray crystallography. EM can also be used to determine structures by single particle analysis, which does not require proteins to be crystallised.

In addition to providing phases, electron microscopy has the great advantage that even samples that are only one molecule thick can be processed, something that is very difficult with X-ray crystallography. Indeed, electron microscopy requires very thin crystals. Electrons have 10000 times the scattering power of X-rays, and thus, thin samples that cannot diffract X-rays are able to scatter electrons.

All protein specimens in this project were prepared for EM by negative staining (see section 3.10.1 for protocol). In negative staining the sample is dried and embedded in a layer of electron-dense heavy metal salts, that provides high contrast for imaging in the electron microscope. The main advantage of negative staining is that it is quick and easy to use and does not require very specialized equipment

9.2 Study of APEC ClyA by electron microscopy

9.2.1 APEC ClyA pore formation by EM

Purified APEC ClyA was incubated with 1% ß-octylglucoside (ß-OG) in order to investigate APEC ClyA pore formation. Images of purified APEC ClyA protein were observed by transmission electron microscopy after staining in 0.75% uranyl formate, using a negative staining protocol (Ohi et al, 2004) as described in section 3. 10. 1.

From (Figure 9.1), the majority of the pores were circular from the top view and bind together in complex but uniformly sized clusters. The pores can be seen as spikes from the side view. The clusters may well be spherical assemblies of pore, not unlike virus particles. Seen from above some appear to have a central pore surrounded by 6 others. This may be consistent with 12- fold symmetry seen for K12 ClyA pores in the crystal structure (Mueller et al, 2009) or may reflect hexagonal close packing.

It is seen that different concentrations of APEC ClyA (0.1 mg/ml) and (0.025 mg/ml) in 25 mM Tris HCl at pH 6.8, 100 mM NaCl, 2.5mM $CaCl_2$, 0.05% NaN_3 and 1% ß-OG mixed with 1% of ß-OG gave similar results. However, using a different detergent (0.1% DDM) with APEC ClyA (0.1 mg/ml) produces more individual pores (Figure 9.2), which means in practice they should be easier to crystallize. For that reason, DDM was used in the 3D crystallization trials of APEC ClyA pores (See Chapter 4).

APEC ClyA pores are homogeneous in size and the majority of the pores appear to have an external pore diameter of approximately 110 Å and are 145Å long, measured from EM micrograph. These measurements were done by Digital Micrograph (GATAN software). The results presented here confirm earlier findings by Wyborn et al, (2004) who found the pores formed by APEC ClyA were more homogeneous than K-12 ClyA ones. Also the larger pore assemblies seen for the *E. coli* K-12 protein were absent (Wallace et al, 2000). This may be due to a higher variation of the number of ClyA subunits in the K-12 ClyA pore, like in some other toxins, e. g. the bacterial toxin pneumolysin (Tilley et al, 2005).

9.2.2 Two-dimensional crystallization trials of the trans-membrane APEC ClyA complex

To understand the structure and function of APEC ClyA in its real context in the lipid bilayer of the cell membrane, I attempted two-dimensional (2D) crystallization trials of membrane protein complexes since the physiological state of the protein may not always be reflected by the 3D crystal structure of the protein in detergent.

To attempt to obtain good 2D crystals APEC ClyA was combined with solubilized phospholipids and detergent, and then subsequently the detergent was removed.

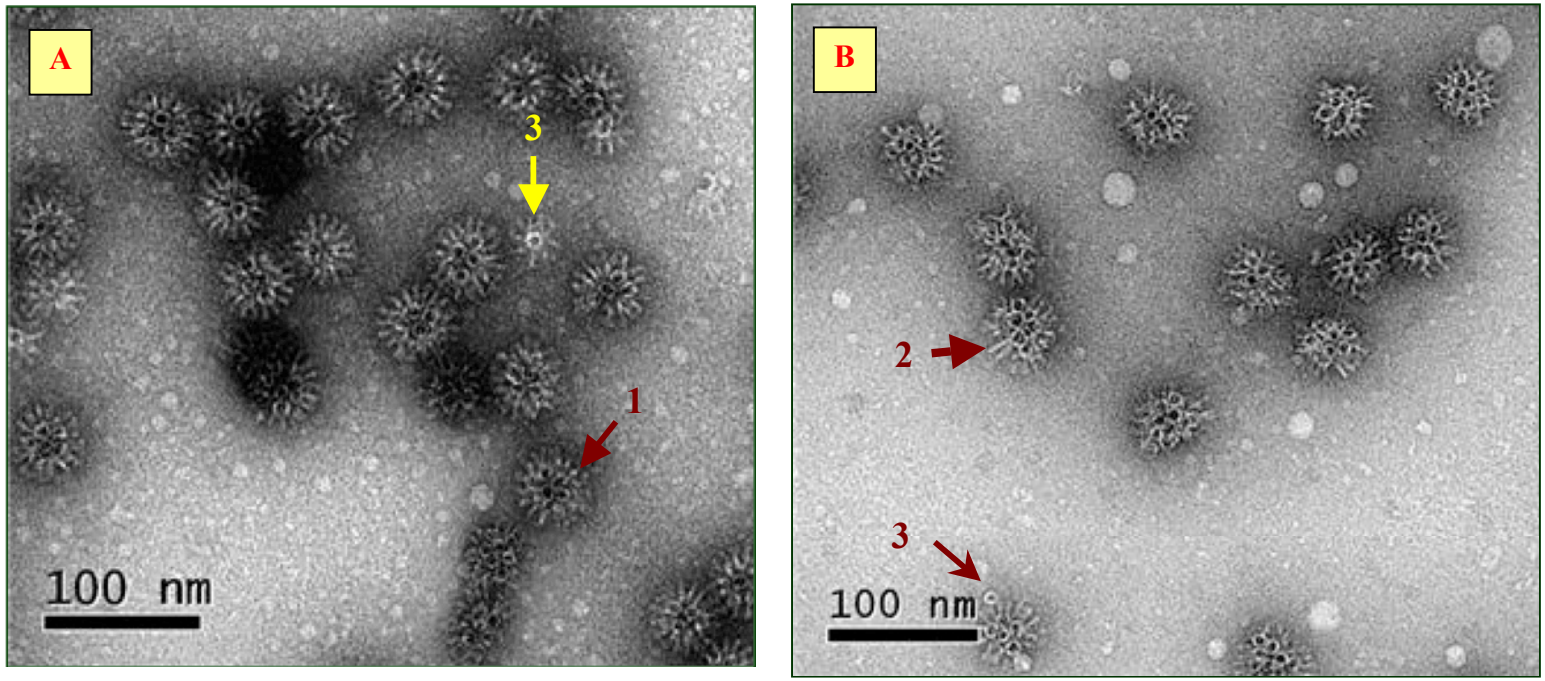


Figure 9.1: Electron micrograph of negative stained APEC ClyA in the presence of detergent (1% ß-OG).
(A) Using protein concentration of 0.1 mg/ml.
(B) Using protein concentration of 0.025 mg/ml.
The pore assemblies can be seen to be circular in the top views (1), while they are seen protruding spikes in the side view (2); (3) shows more isolated pores which were seen rarely.

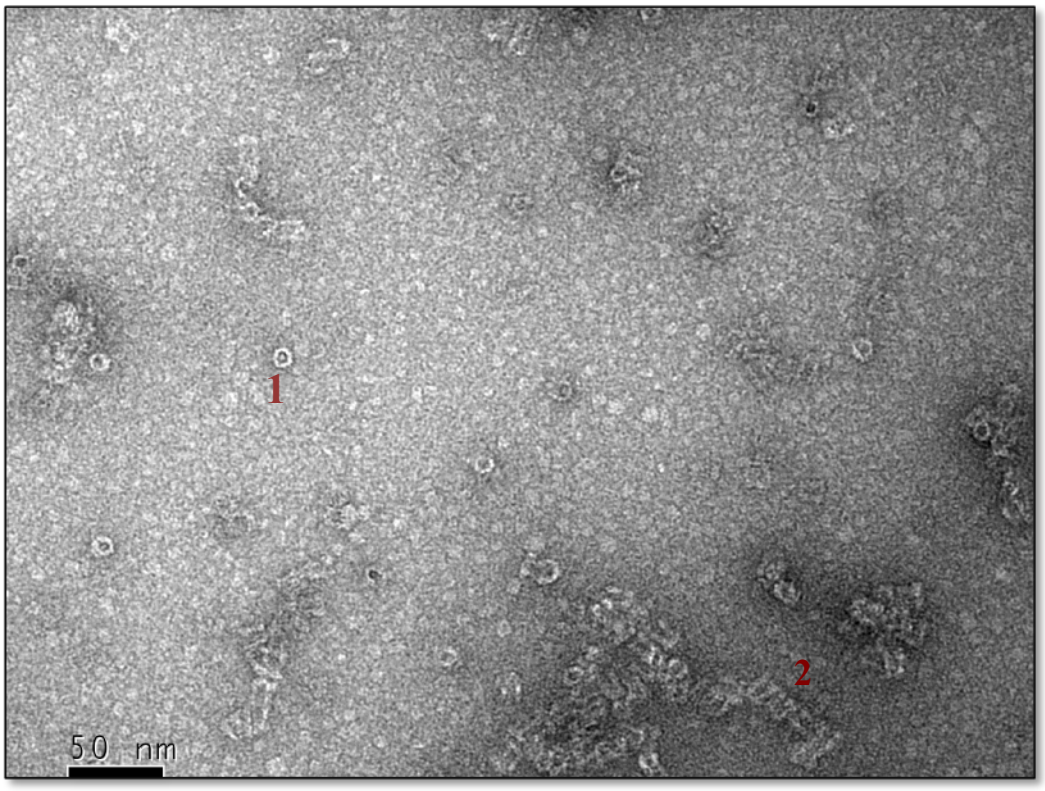


Figure 9.2: Electron micrograph of negative stained APEC ClyA in the presence of detergent (1% DDM).
(1) The top views of isolated pores. (2) The side view of the pores.

There are a number of parameters affecting the results of a 2D crystallization trial such as the lipid-to-protein ratio (LPR), the type of lipids and detergent, concentration of detergent, salt concentration, temperature, and pH (Abeyrathne et al, 2010).

In our trials, two kinds of lipids were used: Brain total lipid extract (BL) and *E. coli* total lipid extract (EL) which were prepared from lipid film (Avanti Polar lipids). We used method 1 for lipid preparation (see section 3.10. 4).

APEC ClyA (0.5 mg/ml) in the presence of ß-OG was mixed with phospholipids in different ratios and dialyzed against ß-OG free buffer (25 mM Tris HCl at pH 6.8, 100 mM NaCl, 2.5 mM $CaCl_2$, 0.05 % NaN_3) by the mini dialysis method for 20 days, changing the buffer every three days at 25 °C. These trials were set up as follows:

- **Trial 1 (as control)**

12.5 µl of 4.5 mg/ml of APEC ClyA was mixed with 1% of ß-OG but no lipid to give an LPR= 0. The sample was diluted with 25 mM Tris HCl at pH 6.8, 100 mM NaCl, 2.5 mM $CaCl_2$, 0.05 % NaN_3 buffer to obtain the final APEC ClyA concentration 0.5 mg/ml in 100 µl. of solution.

- **Trial 2**

The same as trial 1, except 5 µl of 4 mg/ml lipid was added. Thus, the LPR = 0.4.

- **Trial 3**

The same as trial 1, except 10 µl of 4 mg/ml lipid was added. Thus, the LPR = 0.8.

Electron micrographs in negative stain revealed that a possible small crystalline patch (Figure 9.3 A) was formed using an LPR = 0.4 with EL and an LPR = 0.8 with BL. This small crystal diffracted poorly (Figure 9.3 B). The small crystalline patch of APEC ClyA has approximate unit cell dimensions a=122.4 Å, b=114.0 Å, the cell angle (α) =120.9° as measured

from the real EM micrograph of the crystal. Unfortunately no large high quality 2D crystals suitable for electron data collection were found during this project.

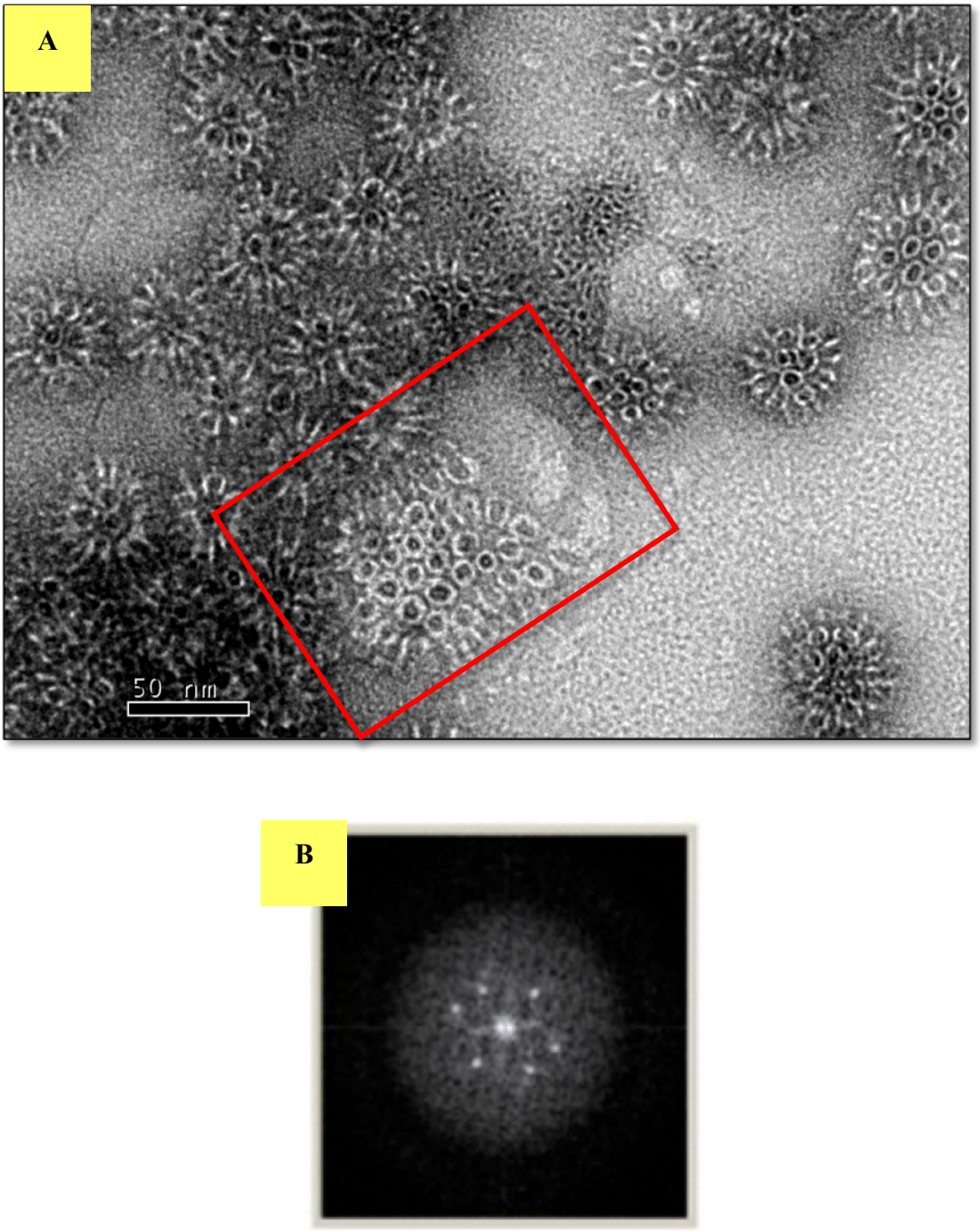


Figure 9.3: Electron micrograph of negative stained APEC ClyA in the presence of lipid.
(A) A small crystalline patch of APEC ClyA (0.5mg/ml) membrane complex (LPR = 0.4 with EL) in the red box. The protein may be starting to arrange itself in a flat lipid bilayer.
(B) Electron diffraction pattern of the crystalline patch.

9.3 Studying Nhe pore formation by electron microscopy

9.3.1 Study of Nhe pore formation in the presence of detergent by EM

Nhe proteins (NheA and NheB) were incubated with two types of detergent (ß-OG and DDM) in order to find evidence of pore formation. To examine the ability of each protein to form pores, each one was mixed separately with detergent. In addition both of them were combined together in the presence of detergent. The experiments are summarized as follows:

- **Trial 1:** NheA (0.17 mg/ml) with 0.5% DDM.
- **Trial 2:** NheB (0.13 mg/ml) with 0.5% DDM
- **Trial 3:** NheA (0.17 mg/ml) and NheB (0.13 mg/ml) with 0.5% DDM
- **Trial 4:** NheA (0.17 mg/ml) with 2% ß-OG
- **Trial 5:** NheB (0.13 mg/ml) with 2% ß-OG
- **Trial 6:** NheA (0.17 mg/ml) and NheB (0.13 mg/ml) with 2 % ß-OG

From electron micrographs (Figure 9.4), it can be concluded that NheA on its own with detergent is largely aggregated but with some V-shaped objects that may correspond to pores. Some of these resemble the ClyA pores observed before. This result may support the previous study by (Fagerlund, 2007), which stated that NheA on its own has a tendency to form pores.

APEC ClyA is a homo-oligomeric pore forming toxin and needs only one protein to achieve maximum cytotoxic activity (Wallace et al, 2000). In contrast, the Nhe complex is a hetero-oligomeric pore forming toxin and requires three proteins to obtain maximum cytotoxic activity (Granum et al, 1999). An initial measurement of one possible NheA pores indicated that the NheA pore might be a more tapered cone, longer and bigger than APEC ClyA pores (Figure 9.5).

On the other hand, NheB on its own in the presence of detergent forms ribbon-like aggregates (Figure 9.6) with a thickness of about 50 Å. The ribbon-like aggregates can be several hundred Å long.

A combination of the two proteins in the ratio 1 NheA: 1 NheB in the presence of 0.5% DDM produces network like aggregates (Figure 9.7).

Here, it should be pointed out that in this experiment the third Nhe protein, NheC was not used. This was because it is known that NheA and NheB are necessary and sufficient for pore formation and cell lysis (Lindbäck et al, 2010). However, NheC is necessary for full cytotoxicity in some cell types (Vero cells) (Lindbäck et al, 2010). It is believed that NheC may enhance pore formation, either by combining NheA and NheB together or by enhancing conformational changes (Lindbäck et al, 2004).

9.3.2 Study NheA and NheB pore formation in the presence of lipids by EM

In order to investigate the possibility of NheA and NheB forming pores in the presence of lipids, pore formation trials were carried out. In these trials, brain total lipid extract was used which was prepared from lipid film (Avanti Polar lipids). Method 1 was used for lipid preparation (see section 3.10. 4).

NheA (0.5 mg/ml) and NheB (0.5 mg/ml) in the presence of DDM were mixed with phospholipids in different ratios and dialyzed against DDM free buffer (25 mM Tris HCl at pH 7.5, 100 mM NaCl, 2.5 mM $CaCl_2$, 0.05% NaN_3) by the mini dialysis method for 20 days, changing the buffer every three days at 19 °C. The trials were set up as follows:

- **Trial 1:**

6.4 µl of 7.8 mg/ml NheA was mixed with 0.15 % DDM. The sample was diluted to obtain the final NheA concentration of 0.5 mg/ml in 100 µl of solution, and then 10 µl of 2 mg/ml lipid was added. Thus, the LPR = 0.4.

- **Trial 2:**

The same as trial 1, except 20 µl of 2 mg/ml lipid was added. Thus, the LPR = 0.8.

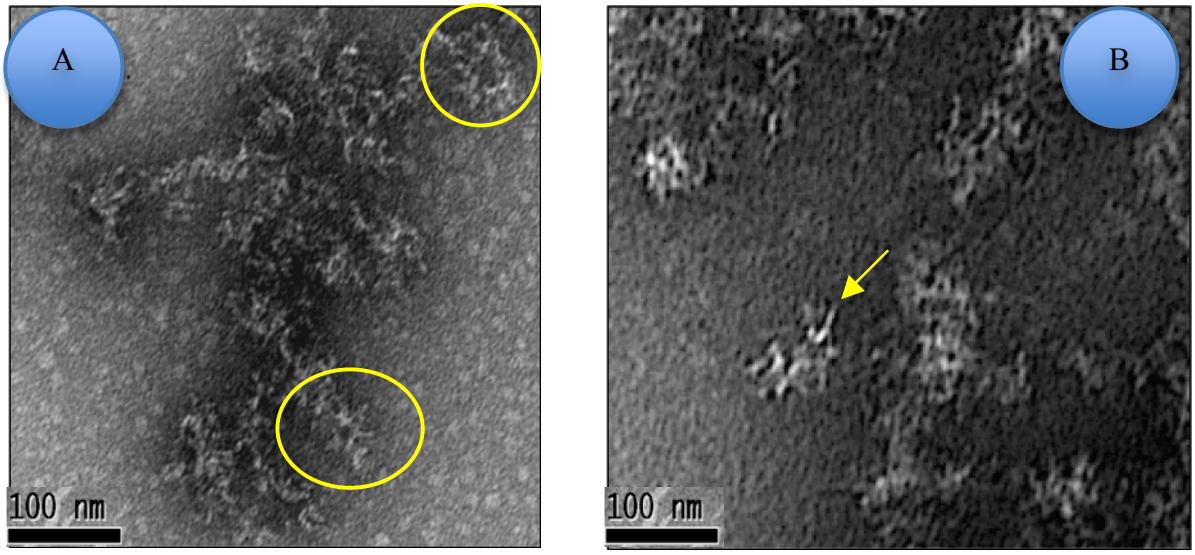


Figure 9.4: Electron micrograph of negative stained NheA and NheB in the presence of detergent (0.5% DDM).

(A) 0.17 mg/ml NheA with 0.5% DDM. There are possible completed pores (yellow circle).

(B) Side view of NheA pore (yellow arrow).

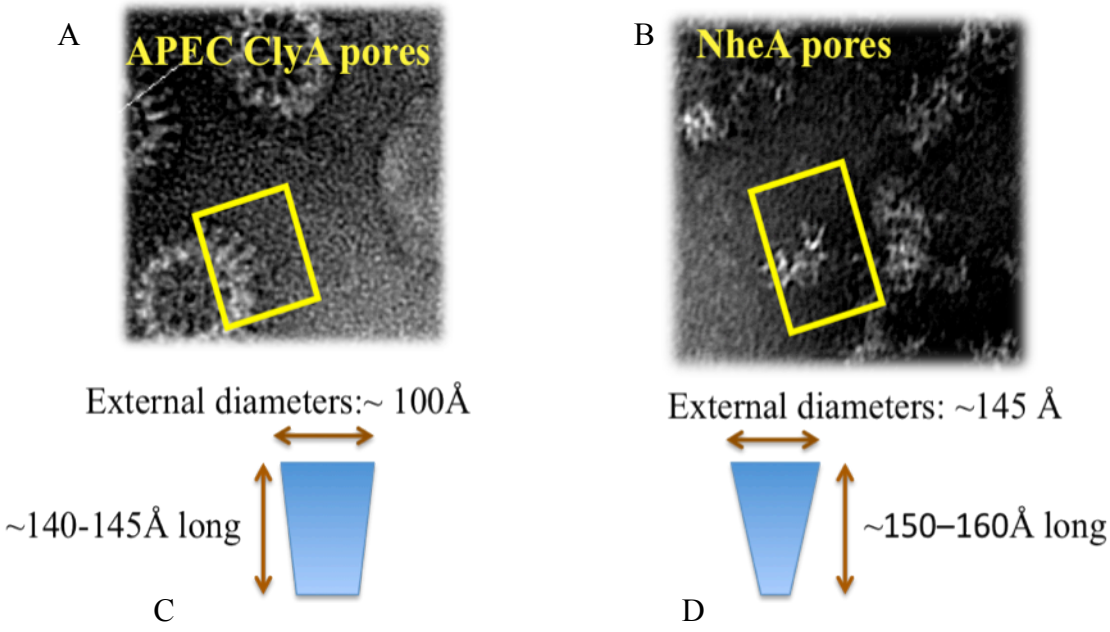


Figure 9.5: Evaluation of the dimensions of NheA and APEC ClyA pores.

A and B are electron micrographs of negative stained APEC ClyA and NheA respectively. C and D show the approximate dimensions and length of the pores.

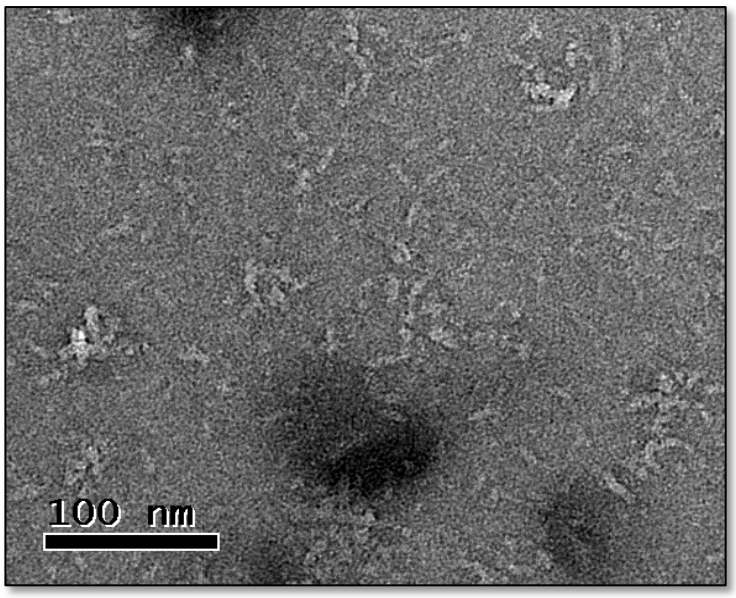


Figure 9.6: Electron micrograph of negative stained NheB in the presence of detergent (0.5% DDM). 0.13 mg/ml NheB with 0.5 % DDM produces ribbon-like aggregates.

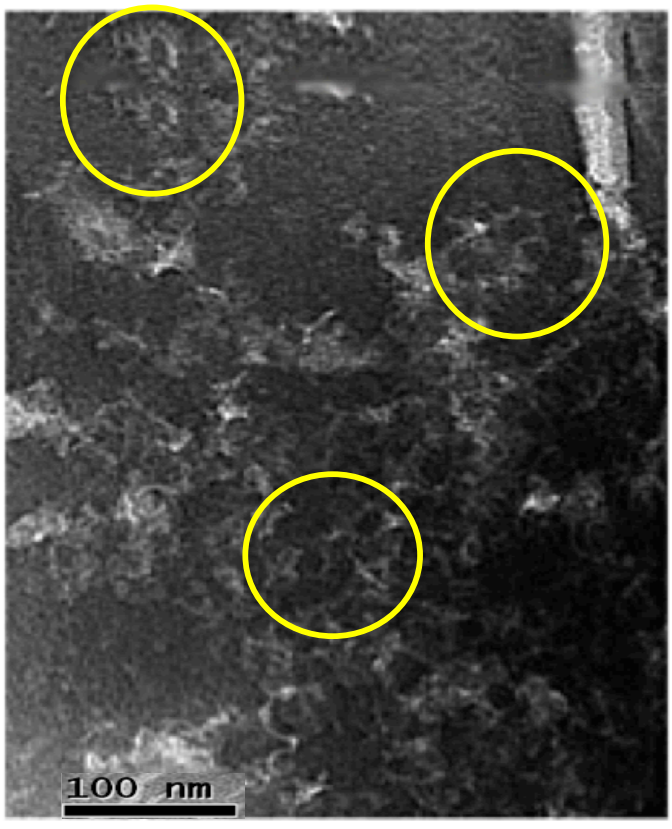


Figure 9.7: Electron micrograph of negative stained a mixture of NheA and NheB in the presence of detergent (0.5% DDM). There are net work-like aggregates which may contain pore-like structures (yellow circles).

- **Trial 3:**

89 µl of 0.5 mg/ml of NheB was mixed with 0.15 % of DDM, and then 10 µl of 2 mg/ml lipid was added. Thus, the LPR = 0.4.

- **Trial 4:**

50 µl of trial 1 and 50 µl of trial 2 to get mixed NheA and NheB with the LPR = 0.4

In the electron micrographs (Figure 9.8), it can be seen that NheA does not form a specific structure with the lipid bilayer of the membrane. On the other hand, NheB forms structures that look like boats and rings with different sizes (Figure 9.9), and it also forms structure that look like pores (Figure 9.10). The boats structures have a tiny lollipop like structures projecting from them. The end of lollipop is approximately 50 Å in diameter and the stick is about 60 Å in long (Figure 9.11). These structures, which will simply be called "boats" in the rest of this Chapter are completely novel in view of two experienced electron microscopists (Prof. P. Bullough, and Prof. H. Saibil, personal communications).

A combination of NheA and NheB in the presence of lipid gives unusual looking networks, which look like assemblies of pore-like structures (Figure 9.12)**.**

9.3.3 NheA and NheB pore formation in the presence of liposomes by EM

Method 2 for lipid preparation was used in these experiments (see section 3.10.4) in order to find the optimal conditions for Nhe pore formation.

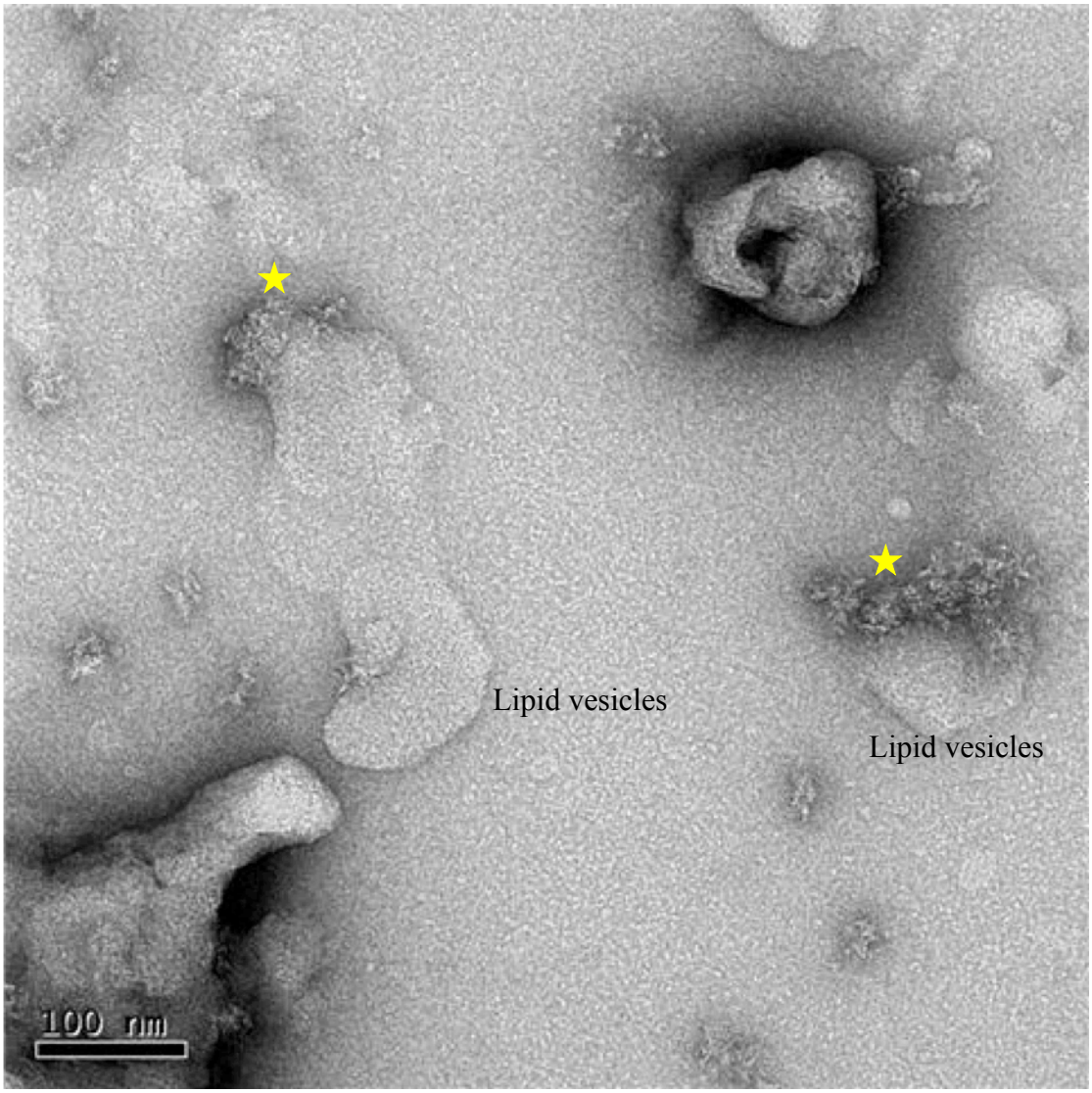


Figure 9.8: Electron micrograph of negative stained NheA in the presence of lipid preparation (method 1 was used for lipid preparation as described in section 3.10.4). NheA does not form a specific structure with lipid vesicles.
Brain Lipid in LPR:0.4, NheA (0.5 mg/ml).

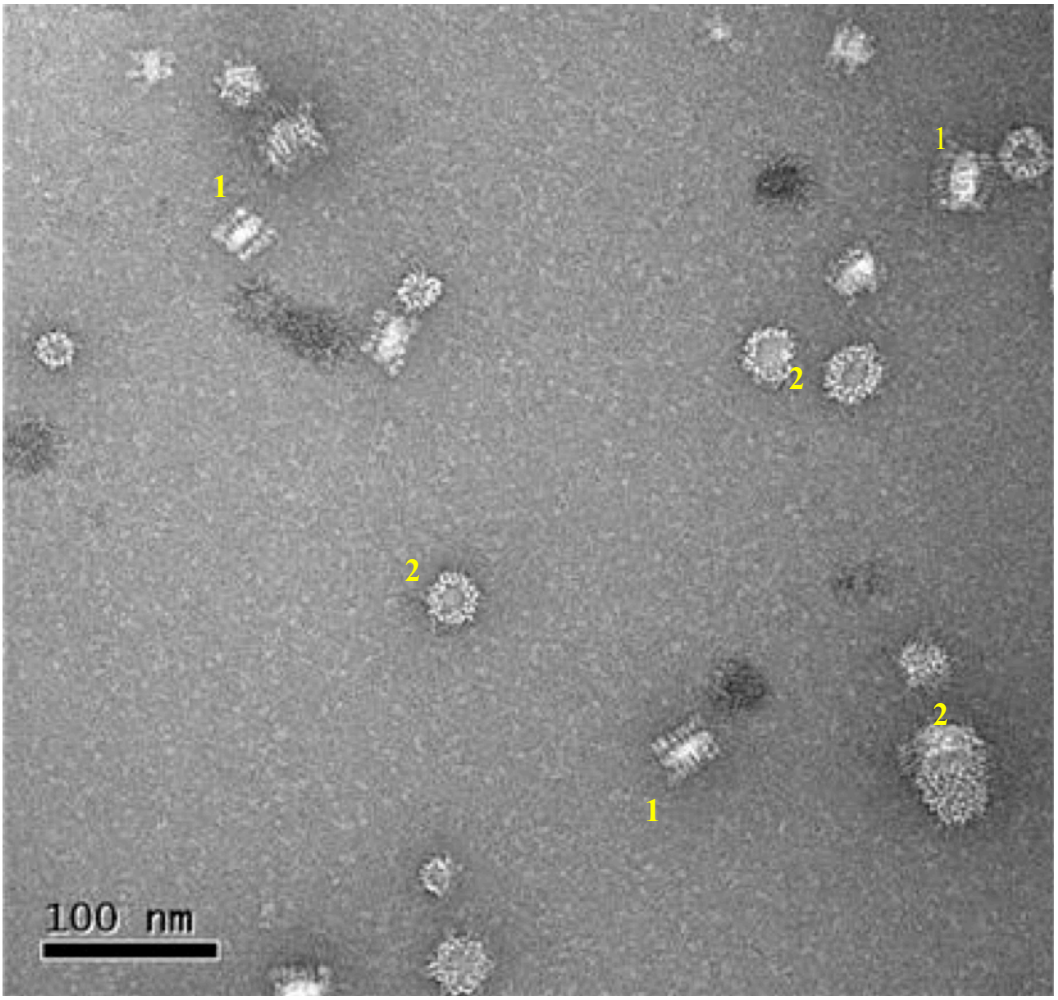


Figure 9.9: Electron micrograph of negative stained NheB in the presence of lipid preparation (method 1 was used for lipid preparation as described in section 3.10.4).
NheB forms boats (1) and rings (2). This result was observed using LPR:0.4; NheB (0.5 mg/ml).

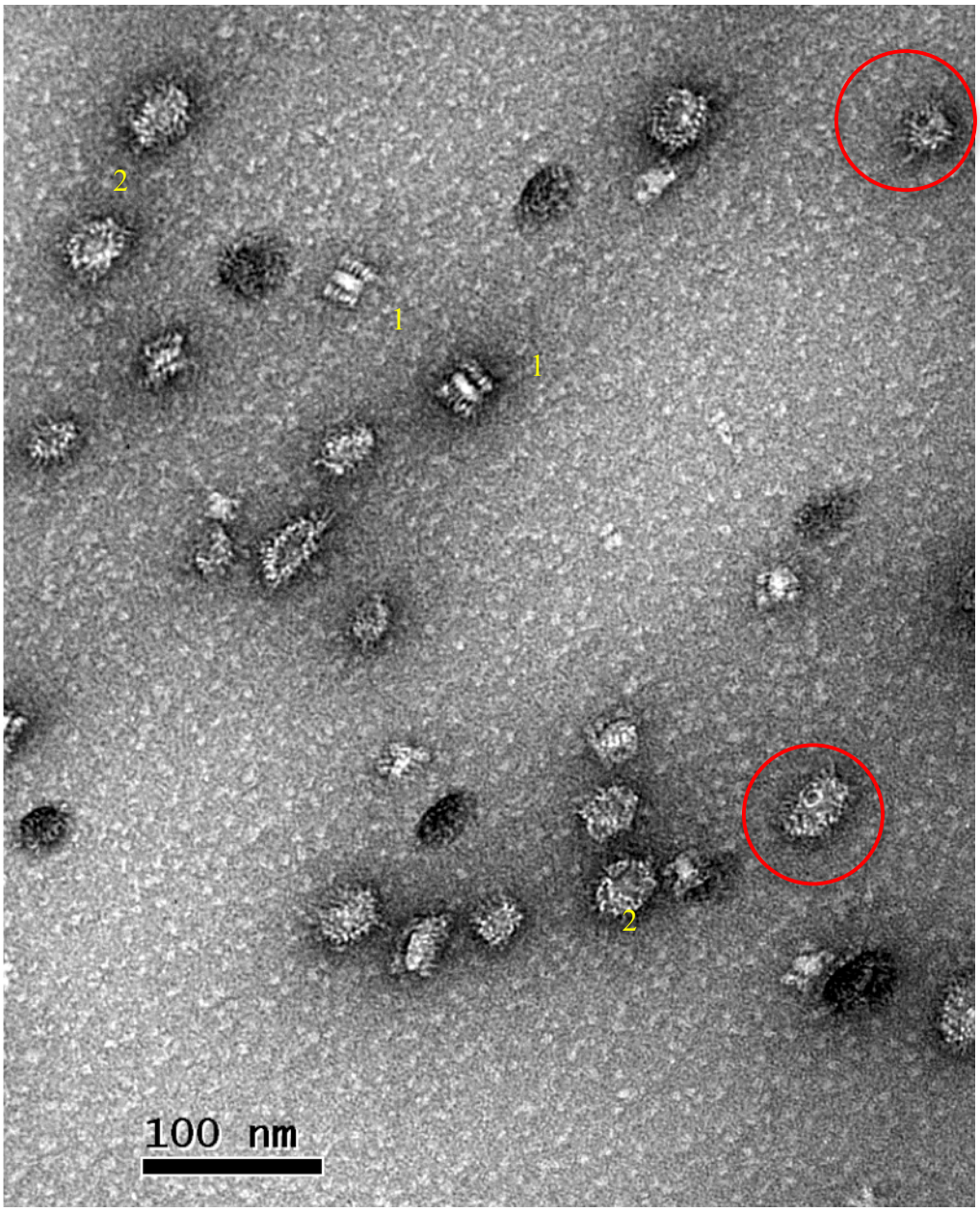


Figure 9.10: Electron micrograph of negative stained NheB in the presence of lipid preparation (method 1 was used for lipid preparation as described in section 3.10.4).
NheB forms boats (1) and rings (2) as well forms structure look like pores (red circles). This result was observed using LPR:0.4; NheB (0.5 mg/ml).

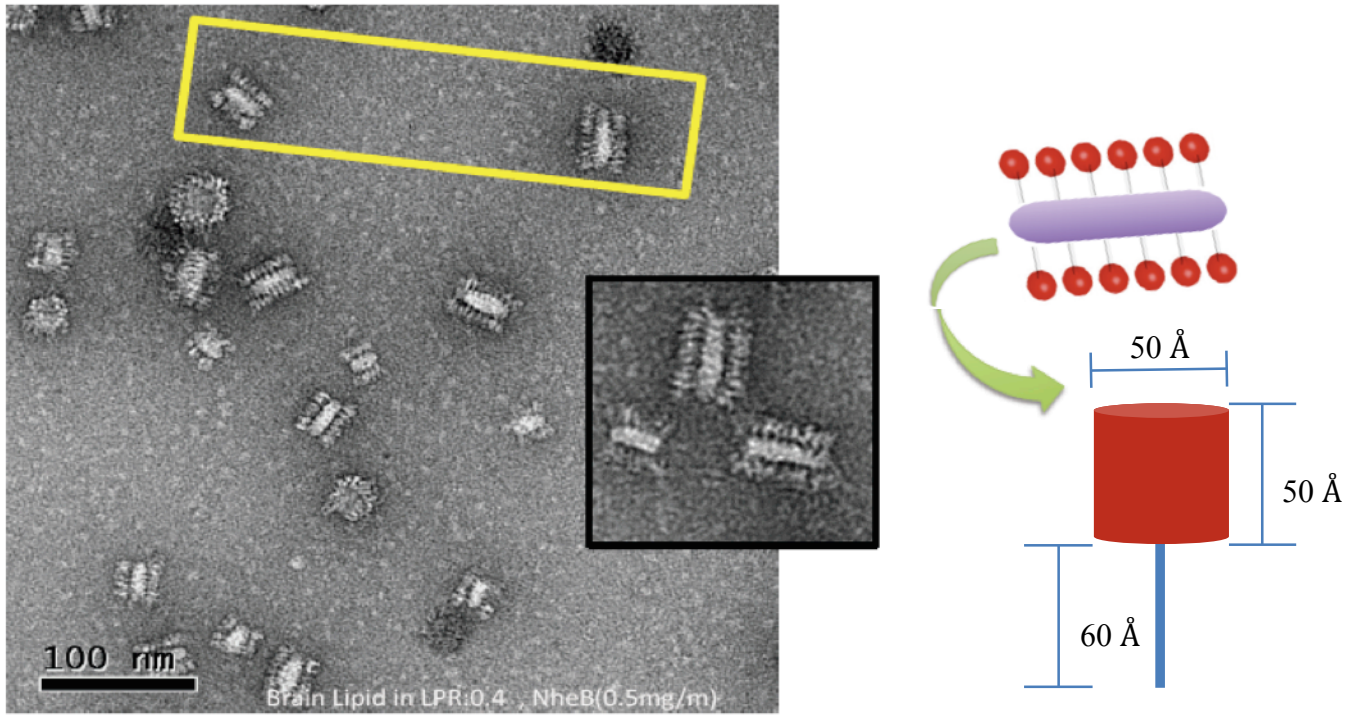


Figure 9.11: Electron micrograph of negative stained NheB in the presence of lipid preparation (method 1 was used for lipid preparation as described in section 3.10.4).
NheB forms strange structures that look like lollipops that shown in a yellow box and close up of this structure is shown in black box. This result was observed by using LPR: 0.4 and LPR: 0.8 ; NheB (0.5 mg/ml).

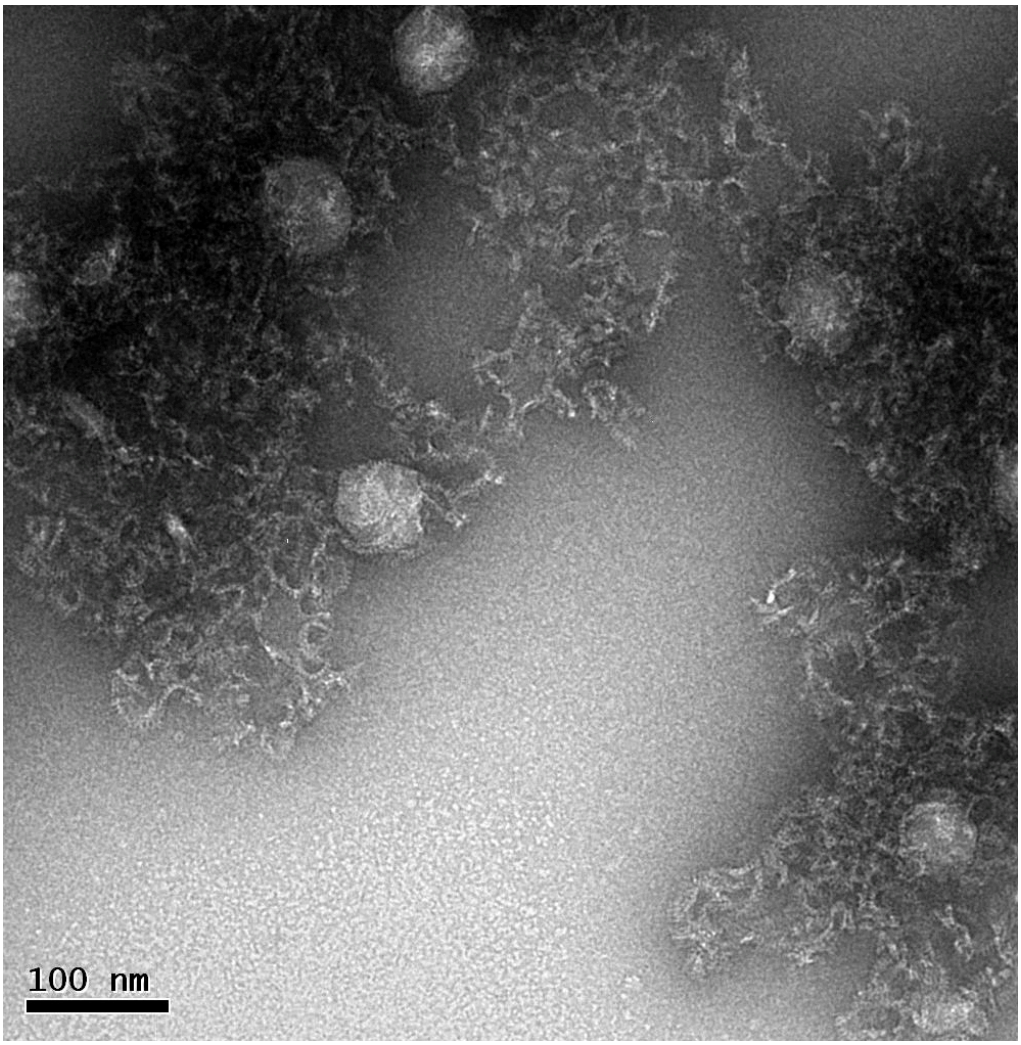


Figure 9.12: Electron micrograph of negative stained a combination of NheA and NheB in the presence of lipid preparation (method 1 was used for lipid preparation as described in section 3.10.4).
NheA and NheB form net-like assemblies of possible partial pore-like structures. This result was observed using LPR:0.4. NheA (0.5 mg/ml) and NheB (0.5 mg/ml).

Liposomes from Brain Total Lipid Extract from Avanti Polar Lipids were mixed with the protein. The NheA (final concentration 0.1 mg/ml), NheB (final concentration 0.1 mg/ml) and NheC (final concentration 0.05 mg/ml) were mixed with liposomes in different ratios of protein to lipid and for different lengths of time. The experiments are summarized in (Table 9.1). The controls were made from liposomes (0.4 mg/ml) alone (Figure 9.13). From the electron micrographs it can be observed:

NheA with liposomes shows the protein does not appear to form specific structures associated with the liposome (Figure 9.14). However, NheB with liposomes preparation (in ratio 1:1) formed rings of different sizes and formed once again structures looking like boats with oars. As well, it can be seen NheB forms structure look like "hairy" or "spikey" ring- like structures in projection (Figure 9.15 A). This result confirmed our previous observations that mixing NheB with lipid leads to formation of boat like structures and rings (see section 9.2.2). After incubation of NheB with liposomes preparation (in ratio 1: 1) for 20 days, the big rings were observed as well as boats and smaller rings (Figure 9.15 B and C). Increasing the ratio of lipid to protein (3:1) apparently led to formation of more ring and boat structures (Figure 9.16 A and B).

However it was also found NheB alone without liposomes was able to form rings and boats (Figure 9.17), but not as much as in NheB with liposomes. Since we were able to see NheB alone forming rings and boats, is possible some hydrophobic materials were co-purified with the NheB protein. It may also be that the protein formed boats and rings over time without the need of lipid.

Incubation of NheA and NheB with liposomes in different ratios for one week showed formation of some small ring structures but the big rings were not seen (Figure 9.18 A, B, and C) which may be mean that some interaction between the two proteins was possibly happening. It was also possible to observe some boats but very rarely comparired to NheB with lipid alone (Figure 9.15 and 9.16). However, significantly some pore structure with

internal diameter 7.5 to 10 nm could be seen (Figure 9.18 A, B and C). Similar results were observed when NheB and lipid were incubated overnight, and later mixed with NheA for 30 min. We saw some smaller rings, but most of the bigger rings were not seen. After incubation of NheA and NheB with liposomes, the formation of the pores increased (Figure 9.18 D, E, and F). However, we found that mixing NheA with NheB without liposomes gave amorphous aggregates, with no formation of rings or boats or pores.

- Incubation of NheA and NheB and NheC (in ratio 10:10:1 respectively) with liposomes gives similar result to NheA and NheB with liposomes, as there were some medium rings and pore- like structures (Figure 9.19 A, B, and C). This means that in our electron microscopy experiments the third component NheC did not appear to have any significant effect on the interaction of the other two proteins (NheA and NheB). This is consistent with (Lindbäck et al, 2010) who found that NheA and NheB are necessary and sufficient for pore formation and cell lysis. Nevertheless addition of NheC is important for full cytotoxicity in some cell types (Vero cells) (Lindbäck et al, 2010).

Table 9.1: Experimental conditions for Nhe pore formation by EM in the presence of liposomes.

The conditions
Lipid alone
NheA (0.1mg/ml) + lipid (1L:1P)
NheA (0.1mg/ml) + lipid + 1A8 (1L:1P)
NheB (0.1mg/ml) + lipid (1L:1P)
NheB (0.1mg/ml) + lipid + (3L:1P)
NheB (0.1mg/ml) + lipid + 1E11 (1L:1P)
NheB (0.1mg/ml) + lipid + 2B11 (1L:1P)
NheB (0.1mg/ml) +lipid + secondary antibody(1L:1P)
NheA (0.1mg/ml) + NheB (0.1mg/ml) + lipid (1L:1P)
NheA (0.1mg/ml) + NheB (0.1mg/ml) + lipid (1L:2P)
NheA (0.1mg/ml) + NheB (0.1mg/ml) + lipid (1L:3P)
NheA (0.1mg/ml) + NheB (0.1mg/ml) + lipid (3L:1P)
NheA (0.1mg/ml) + NheB (0.1mg/ml) + NheC (0.05 mg/ml) + lipid (1L:1P)
NheA (0.1mg/ml) + NheB (0.1mg/ml)
NheA (0.1mg/ml) + NheB (0.1mg/ml) + 1A8
NheA (0.1mg/ml) + NheB (0.1mg/ml) + lipid +1A8(1L:1P)
NheA (0.1mg/ml) + NheB (0.1mg/ml) + lipid +1E11(1L:1P)
NheA (0.1mg/ml) + NheB (0.1mg/ml) + lipid +2B11(1L:1P)
NheB (0.1mg/ml) + lipid + Crosslinking by Glutaraldehyde + 1E11
NheB (0.1mg/ml) + lipid + Crosslinking by Glutaraldehyde + 2B11
NheB (0.1mg/ml) + lipid + Crosslinking by Glutaraldehyde + NheA
NheB (0.1mg/ml) + lipid + Crosslinking by Glutaraldehyde (control)
NheB alone
NheB (0.1mg/ml) + 2B11
NheB (0.1mg/ml) + 1E11

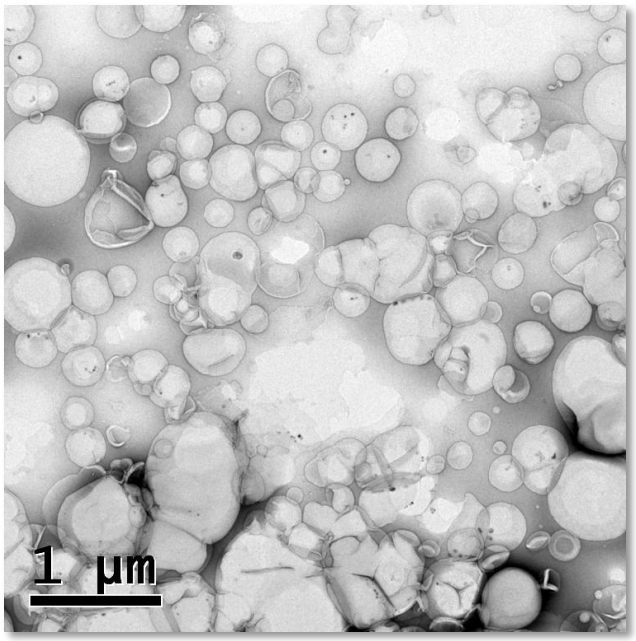


Figure 9.13: Electron micrograph of negative stained control electron micrograph shows liposome vesicles (method 2 was used for lipid preparation as described in section 3.10.4).

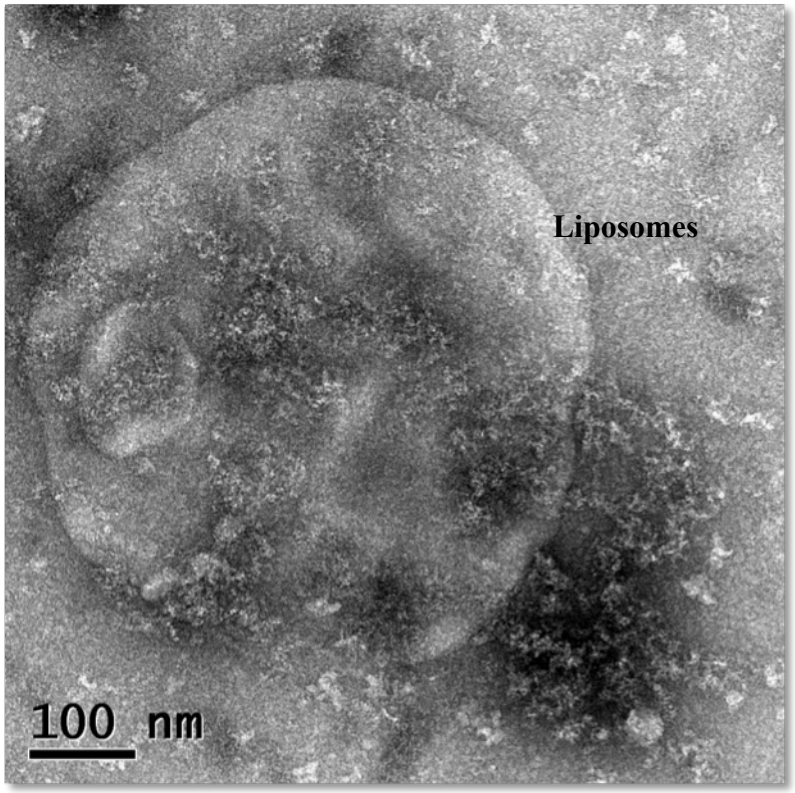


Figure 9.14: Electron micrograph of negative stained NheA in the presence of liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4).
NheA does not form a specific structure with liposome vesicles. This result was observed using the protein to lipid ratio (1:1), (0.1 mg/ml). This was observed after one week and the same result was observed after 20 days.

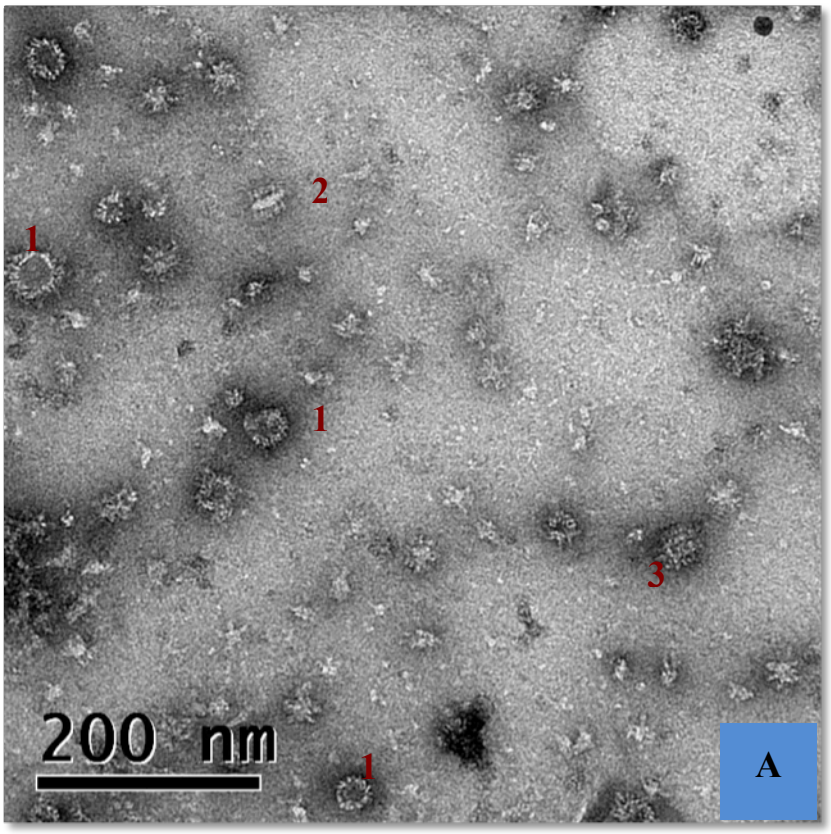


Figure 9.15 A: Electron micrograph of negative stained NheB in the presence of liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4).
NheB forms ring structures of different sizes in projection (1) and forms boats (2). NheB forms structure look like "hairy" or "spikey" ring (3). This result was observed using the protein to lipid ratio (1:1), NheB (0.1mg/ml). This was observed after one week.

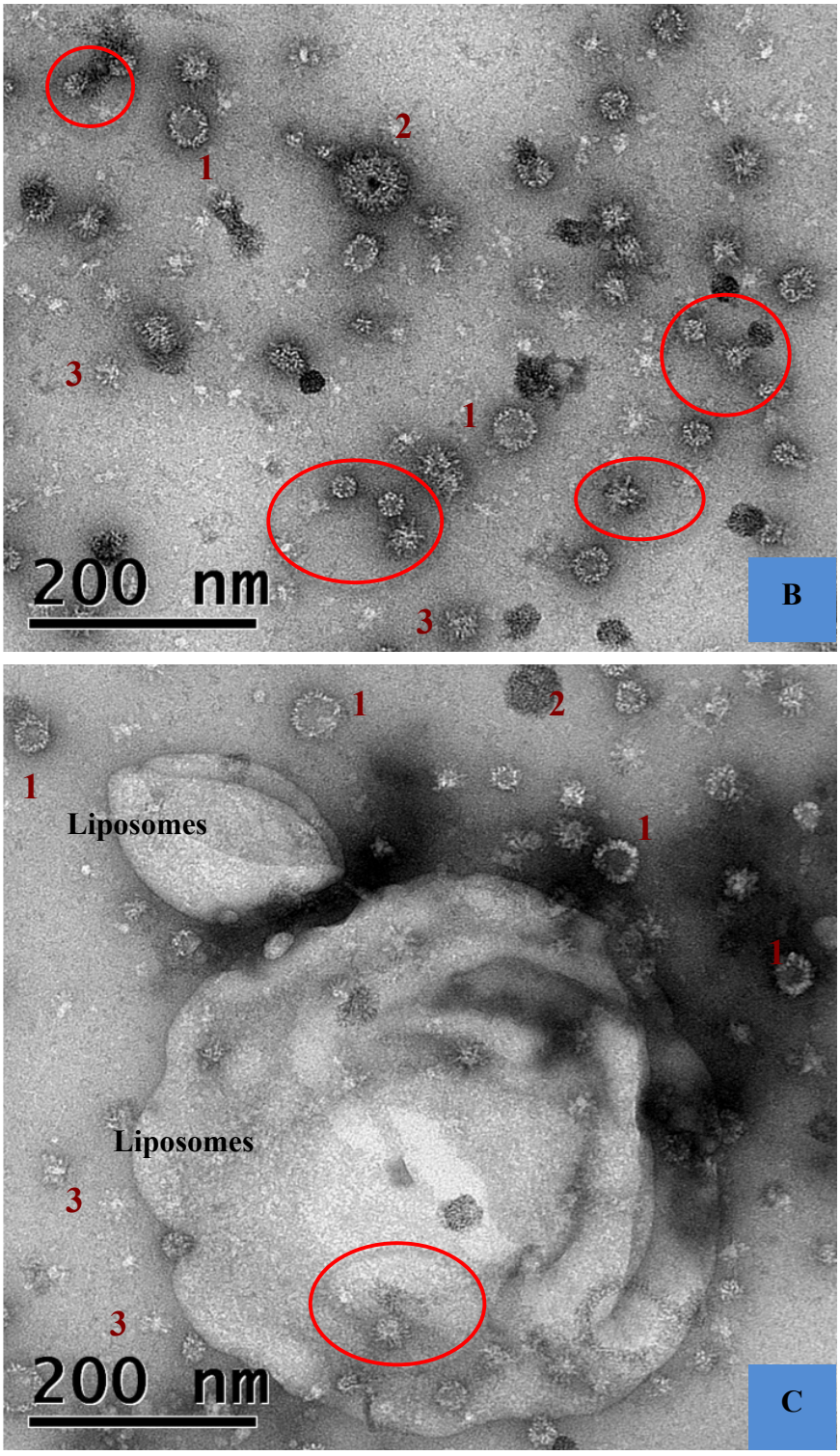


Figure 9.15 B and C: Electron micrograph of negative stained NheB in the presence of liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4). NheB forms rings structures of different sizes (1). Some rings look like "hairy" or "spikey" spheres (2). Structures looking like a boat (3). Small rings and pore-sized assemblies with protein densities protruding out from the central body (red circles) were also observed. This result was observed using the protein to lipid ratio (1:1), NheB (0.1mg/ml). This was observed after 20 days.

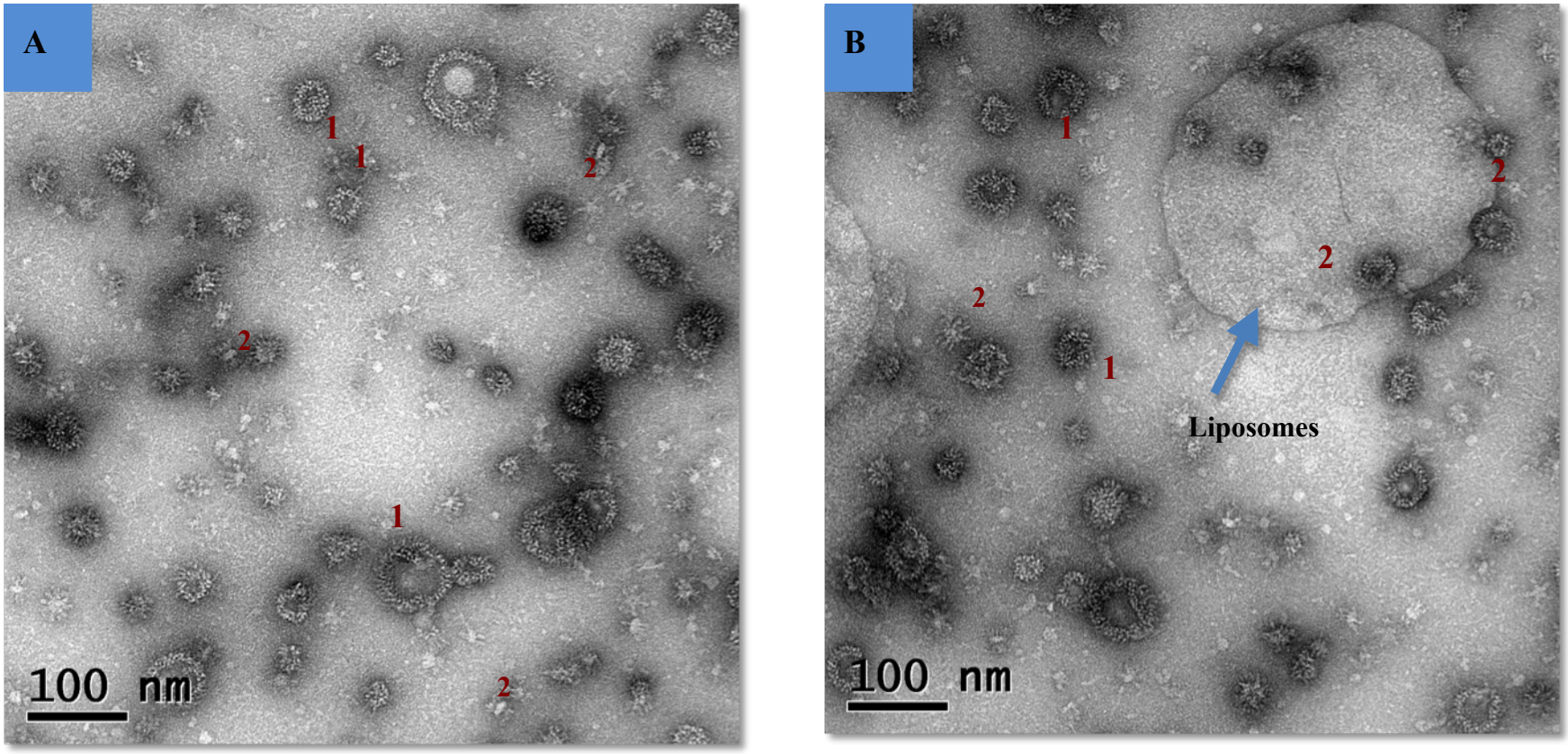


Figure 9.16 A and B: Electron micrograph of negative stained NheB in the presence of liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4).
NheB forms ring of different sizes (1) and form boats (2). This result was observed using the protein to lipid ratio (1:3); NheB (0.1 mg/ml). This was observed after one week and similar results were observed after 20 days

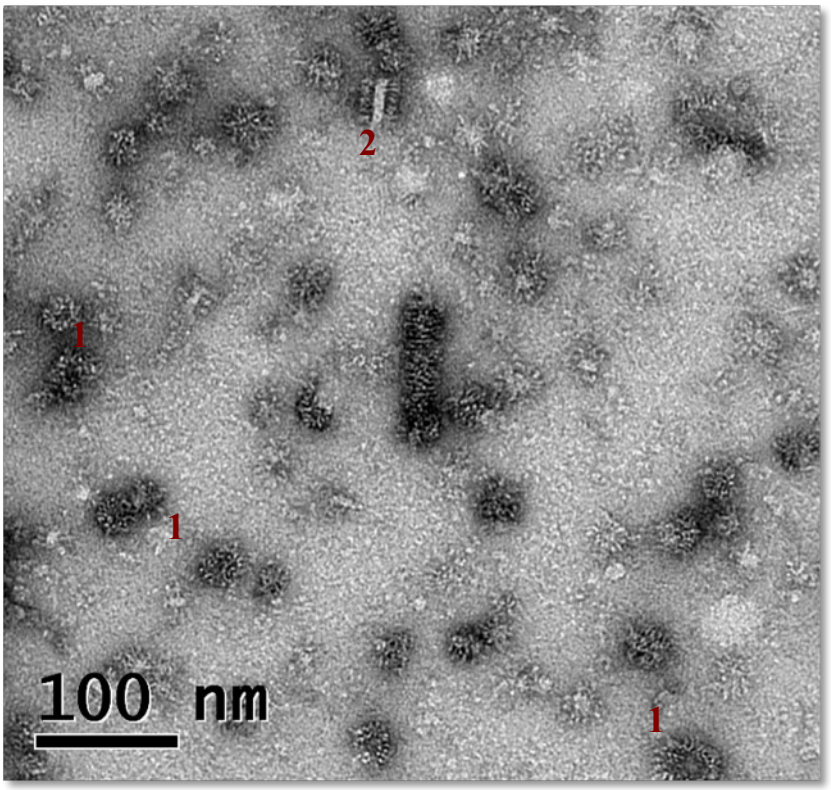


Figure 9.17: Electron micrograph of negative stained NheB.
Without liposomes preparation NheB forms rings in different sizes (1) and forms boats (2). This result was observed using NheB 0.1mg/ml.

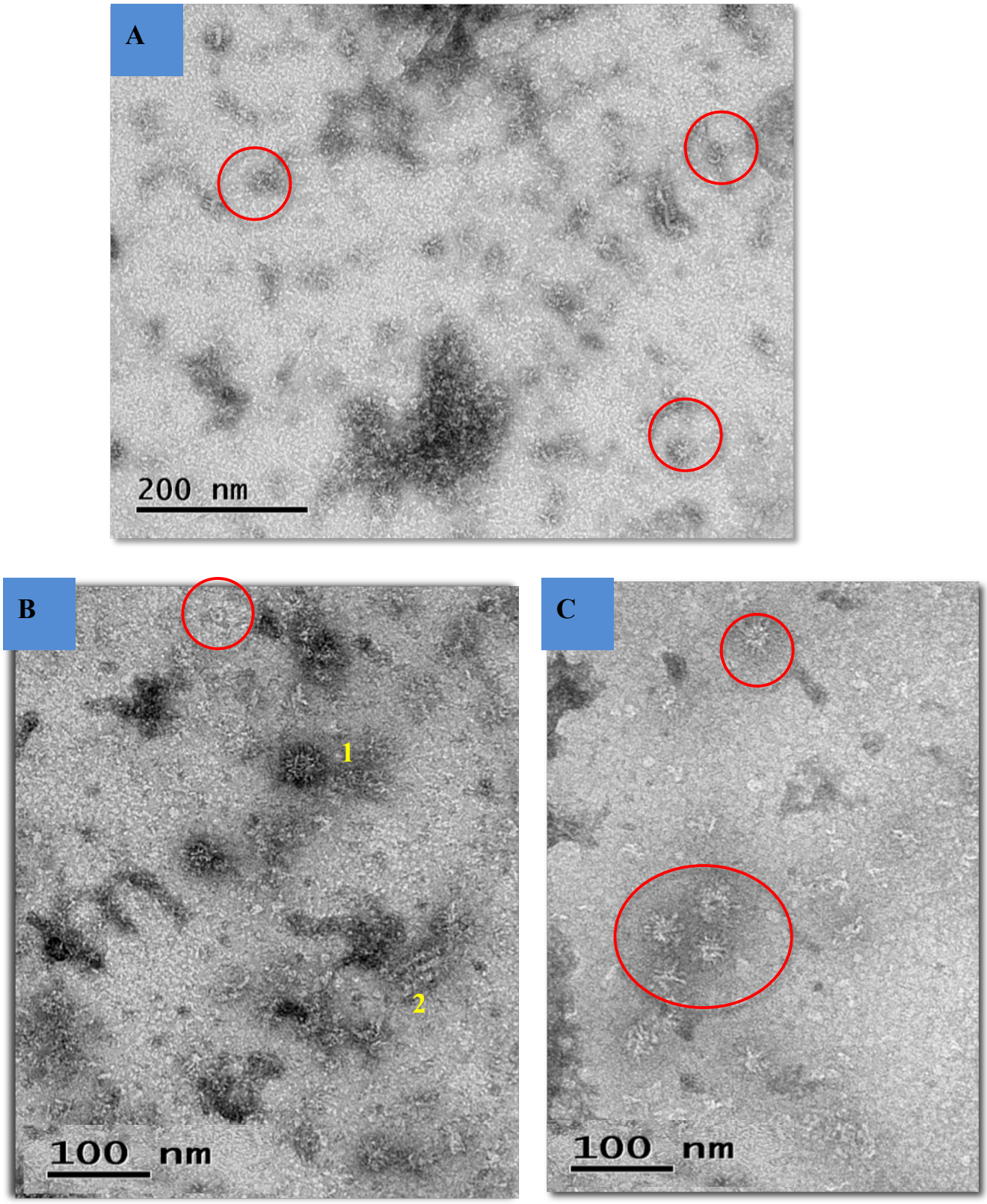


Figure 9.18 A, B, and C: Electron micrograph of negative stained a combination of NheA and NheB in the presence of liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4). NheA and NheB form pore structures (red circles). There are some small rings (1) but no big rings. Some boats (2) can be seen very rarely. This result was observed using the protein to lipid ratio (1:1); NheA (0.1 mg/ml) and NheB (0.1 mg/ml). This result was seen after one week.

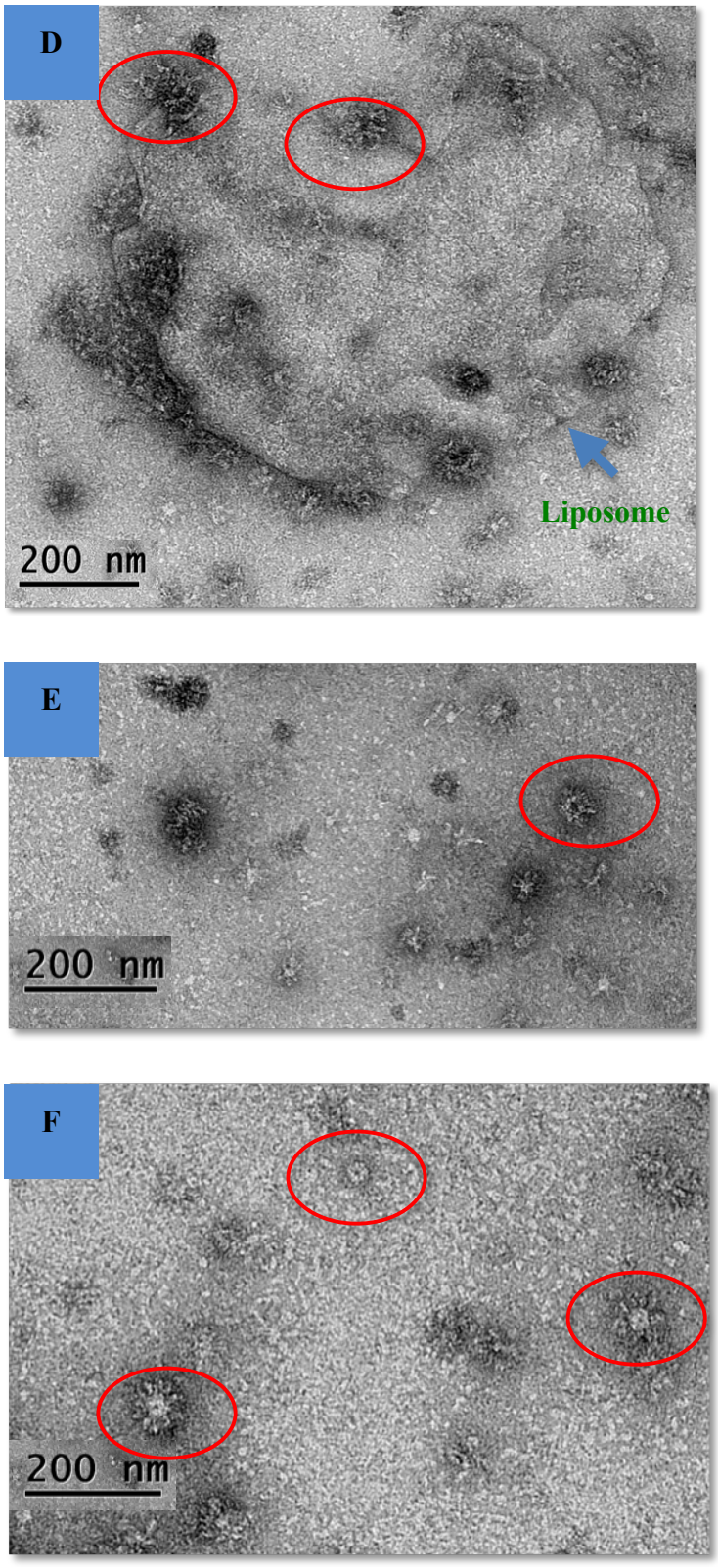


Figure 9.18 D, E, and F: Electron micrograph of negative stained a combination of NheA and NheB in the presence of liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4). NheA and NheB form pore like structures (red circles). Few rings and the boats are seen. This result was observed using the protein to lipid ratio (1:1); NheA (0.1 mg/ml) and NheB (0.1 mg/ml). This result was seen after 20 days.

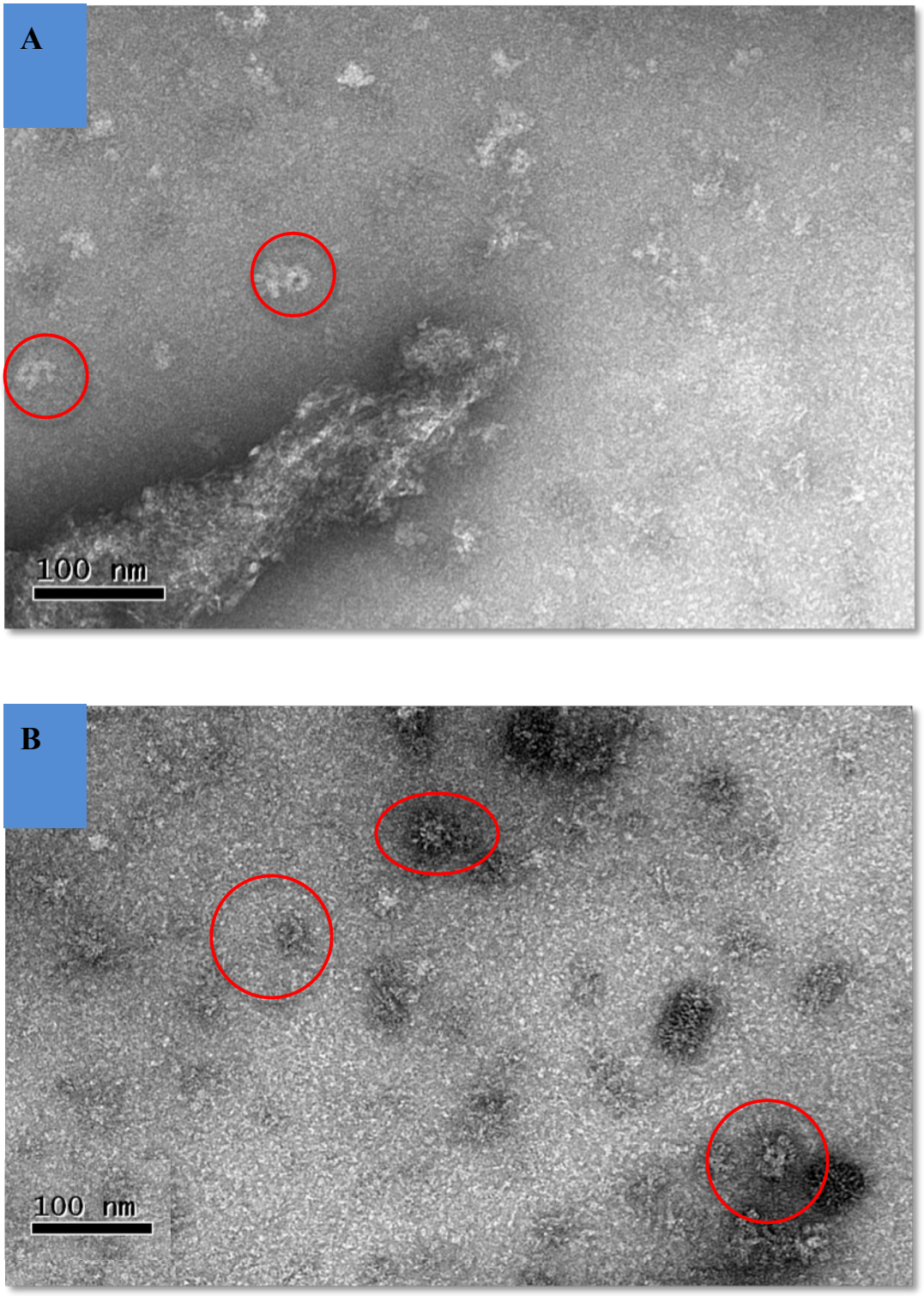


Figure 9.19 A and B: Electron micrograph of negative stained a combination of NheA, NheB and NheC in the presence of liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4). Pore-like structures can be observed (red circles). This result was observed using the protein to lipid ratio (1:1); NheA (0.1 mg/ml) and NheB (0.1 mg/ml) and NheC (0.05 mg/ml). This result was seen after one week.

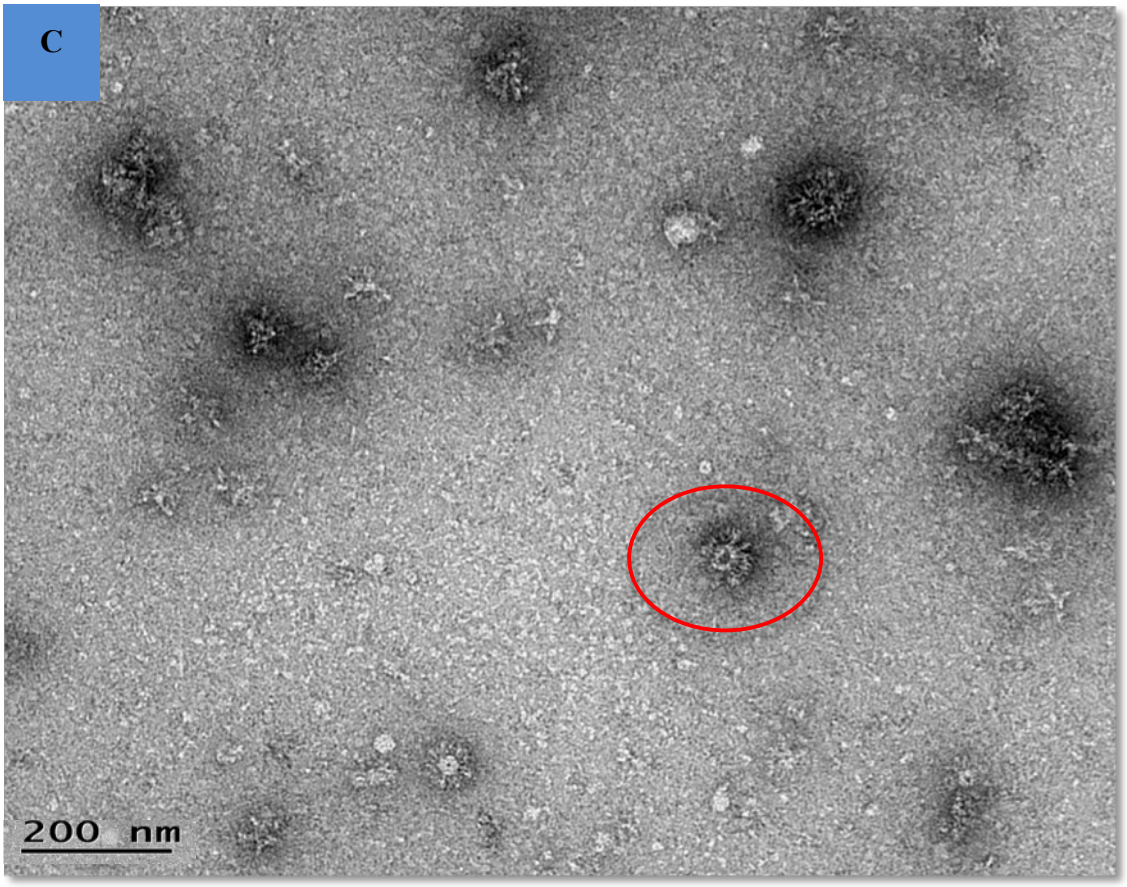


Figure 9.19 C: Electron micrograph of negative stained a combination of NheA, NheB and NheC in the presence of liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4). Pore structures can be observed (red circle). This result was observed using the protein to lipid ratio (1:1); NheA (0.1 mg/ml) and NheB (0.1 mg/ml) and NheC (0.05 mg/ml). This result was seen after 20 days.

9.3.4 Immuno-gold labelling

The aim of immuno-gold labelling is to allow identification of the proteins by labelling them with specific antibodies which are themselves labeled with a gold conjugated antibody which is highly visible in EM. We used three antibodies in our EM studies: MAb 1A8 (0.2 mg/ml) against NheA, MAb 2B11 (0.2 mg/ml) and MAb1E11 (0.2 mg/ml) against NheB. These antibodies have been kindly given from Dr. Erwin Märtlbauer (Ludwig-Maximilians-Universitat, (Dietrich et al, 2005), see table 3. 1 and section 3.10.2 for the protocol. The secondary antibody was Anti-Mouse IgG–Gold antibody produced in goat with the particle size of 5 nm (from SIGMA ALDRICH):

NheA with liposomes was incubated with primary antibody 1A8 and secondary immunogold antibody for 20 min and 1 h but give no specific binding of the secondary antibody. However, there was a strong immunogold labeling to the amorphous background material, which was everywhere (Figure 9.20).

NheB and liposomes were incubated for 20 min and 1h with primary antibody 1E11 and and secondary immunogold antibody. This did not produce any clear structures and no specific immunogold labeling (Figure 9.21). However, NheB and liposomes after incubation for 20 min and 1h with primary antibody 2B11 and secondary immunogold antibody showed formation of rings and the boats (Figure 9.22 A, B, C, and D). On the other hand, NheB alone with either 1E11 or 2B11 did not produce clear structures or specific binding of immunogold (Figure 9.23). However, for NheB and liposomes after incubation for 20 min with secondary antibody, clear ring formation and boats were seen. It therefore appears that the primary antibody 1E11 had a neutralising effect, possibly disrupting the ring and boat structures when added to NheB. 2B11 might have the same effect, although some rings can see after the addition of 2B11.

NheA and NheB and liposomes with either with 1A8 or 1E11 and secondary immunogold antibody did not form specific structure or show any specific

binding (Figure 9.24 A and B). NheA and NheB with liposomes and 2B11 and secondary immunogold antibody form pores look like structure but without specific binding (Figure 9.25 A and B) did do so.

In order to investigate the disruptive effect of the primary antibodies on NheB structure formation, cross-linking experiments using 50 mM glutaraldehyde solution were carried out (see section 3.10.3 for the protocol). Glutaraldehyde solution has revealed to be an effective crosslinking agent for the stabilization of biomaterials (Damink et al, 1995).

The controls were made for NheB with liposomes cross-linked by 50 mM glutaraldehyde solution (Figure 9.26). When NheB with liposomes was cross-linked, then incubated with primary antibodies 1E11 or 2B11, the boats and rings can be seen (Figure 9.27 A and B) without specific immunogold labeling to them. When NheB liposomes were cross-linked and then incubated with NheA, we were also able to see some the big ring formation (Figure 9.28).

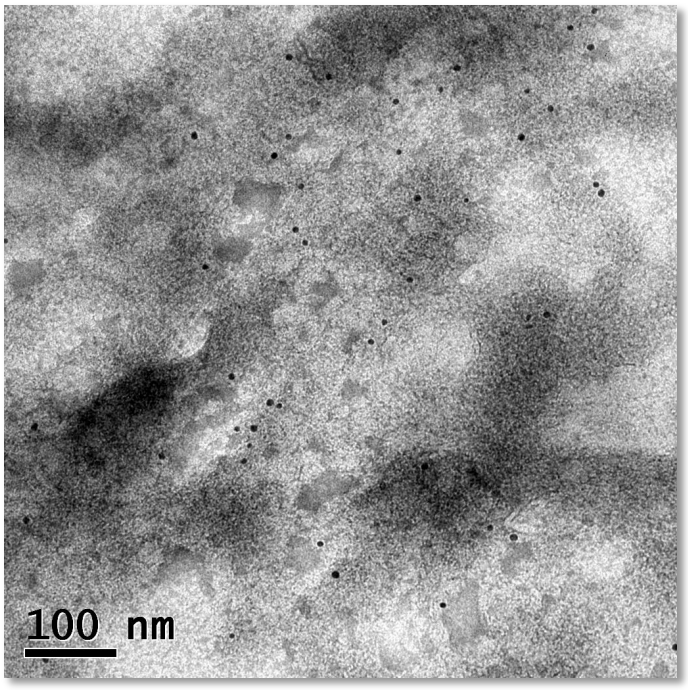


Figure 9.20: Electron micrograph of negative stained immunogold labeled NheA in the presence of liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4). The primary antibody 1A8 (after incubation for 20 min) was used**.** There is no formation of clear structures and there is a random labeling (black spots) to the to the amorphous background material. NheA (0.1 mg/ml).

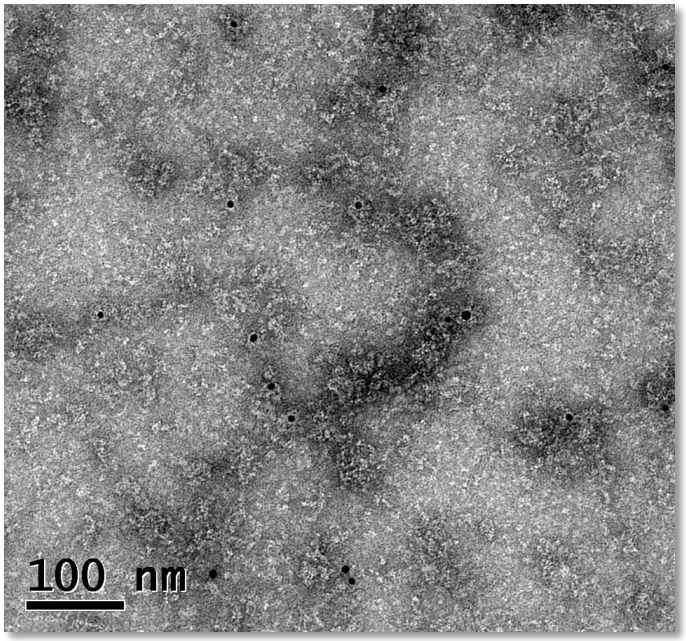


Figure 9.21: Electron micrograph of negative stained immunogold labeled NheB in the presence of liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4). The primary antibody 1E11 (after incubation for 20 min) was used. There is no formation of clear structures and there is a random labeling (black spots) to the to the amorphous background material. NheB (0.1 mg/ml).

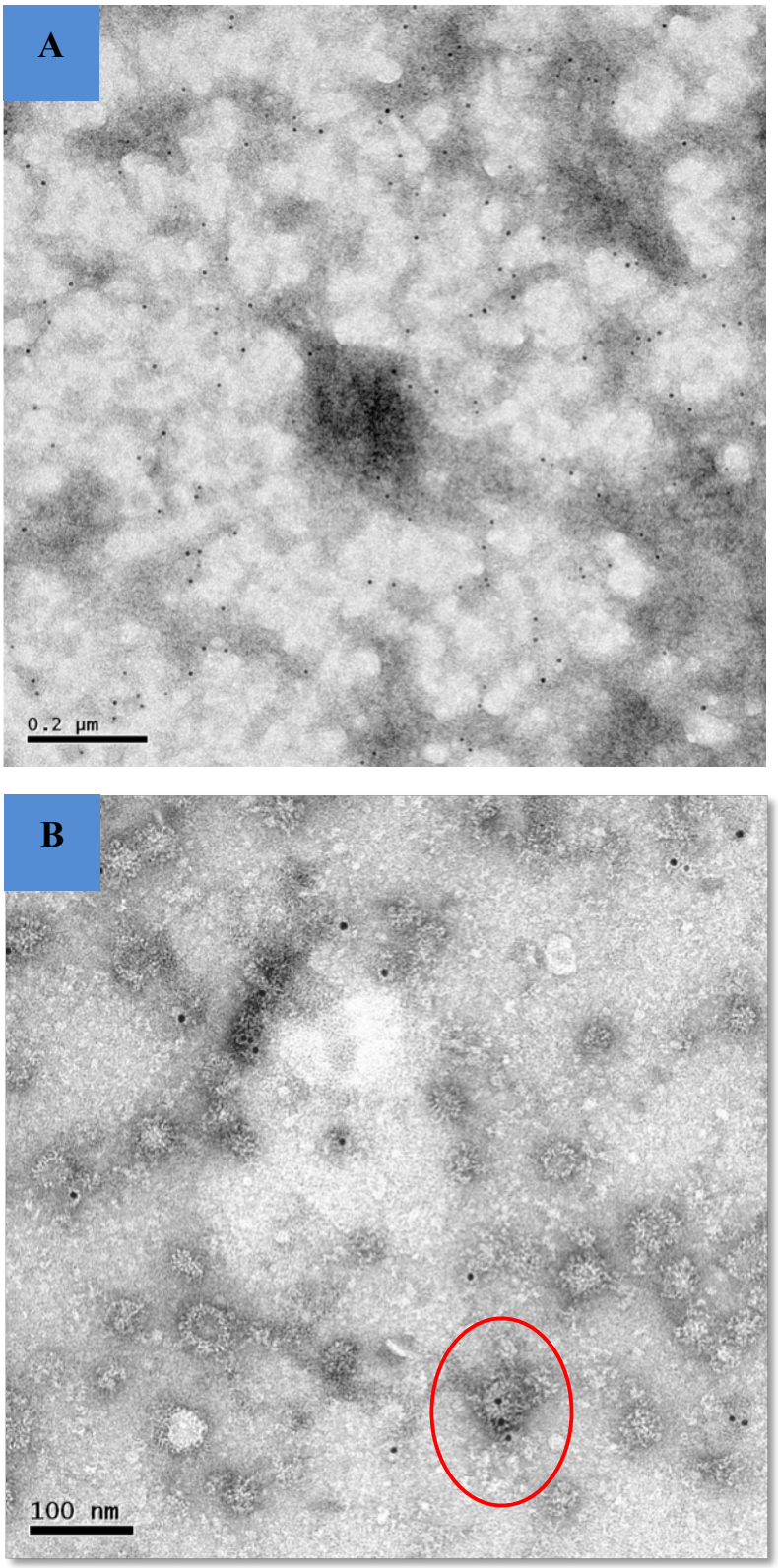
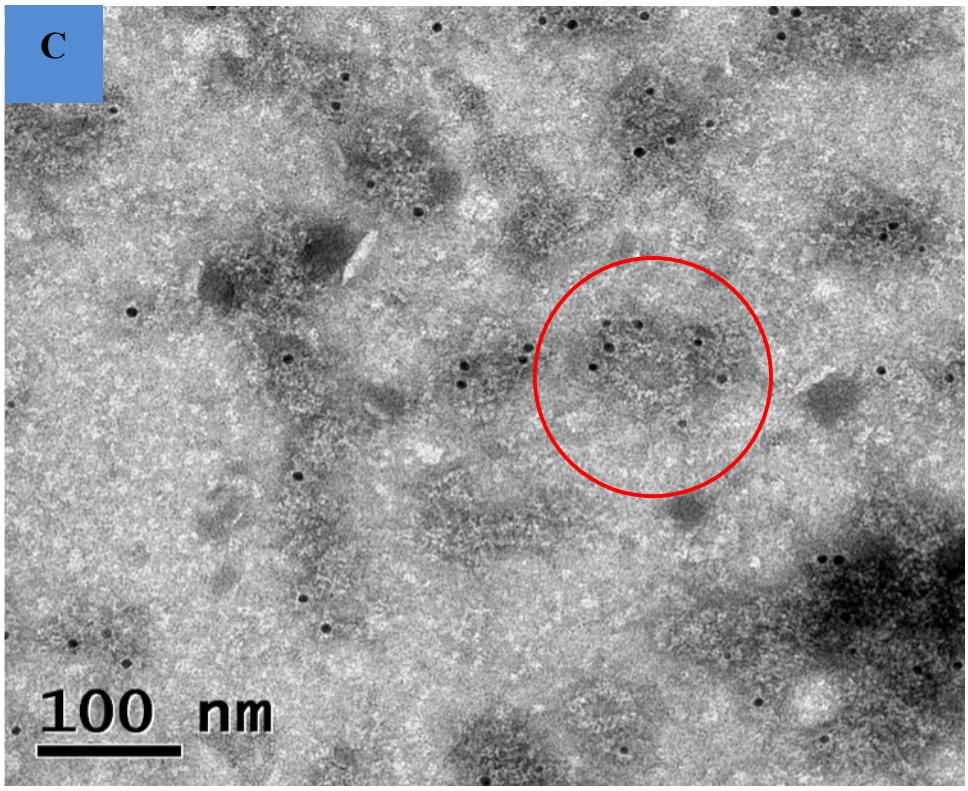


Figure 9.22 A and B: Electron micrograph of negative stained immunogold labeled NheB in the presence of liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4). (A) The primary antibody 2B11(after incubation for 20 min) was used. There is still a random labeling (black spots). (B) There is formation of rings of NheB (0.1 mg/ml; red circle).



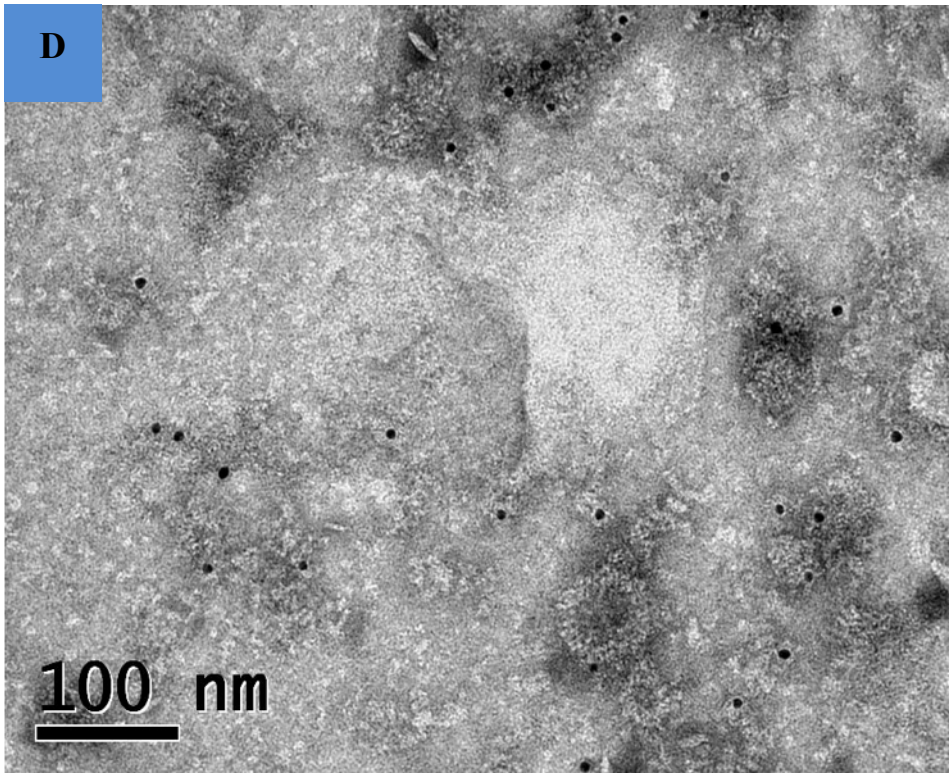


Figure 9.22 C and D: Electron micrograph of negative stained immunogold labeled NheB in the presence of liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4). The primary antibody 2B11 (after incubation for 1h) was used**.** There is formation of rings with possible specific immunogold labeling (black spots) to them (red circles). However, still there is also strong labelling of the amorphous background material. NheB (0.1 mg/ml).

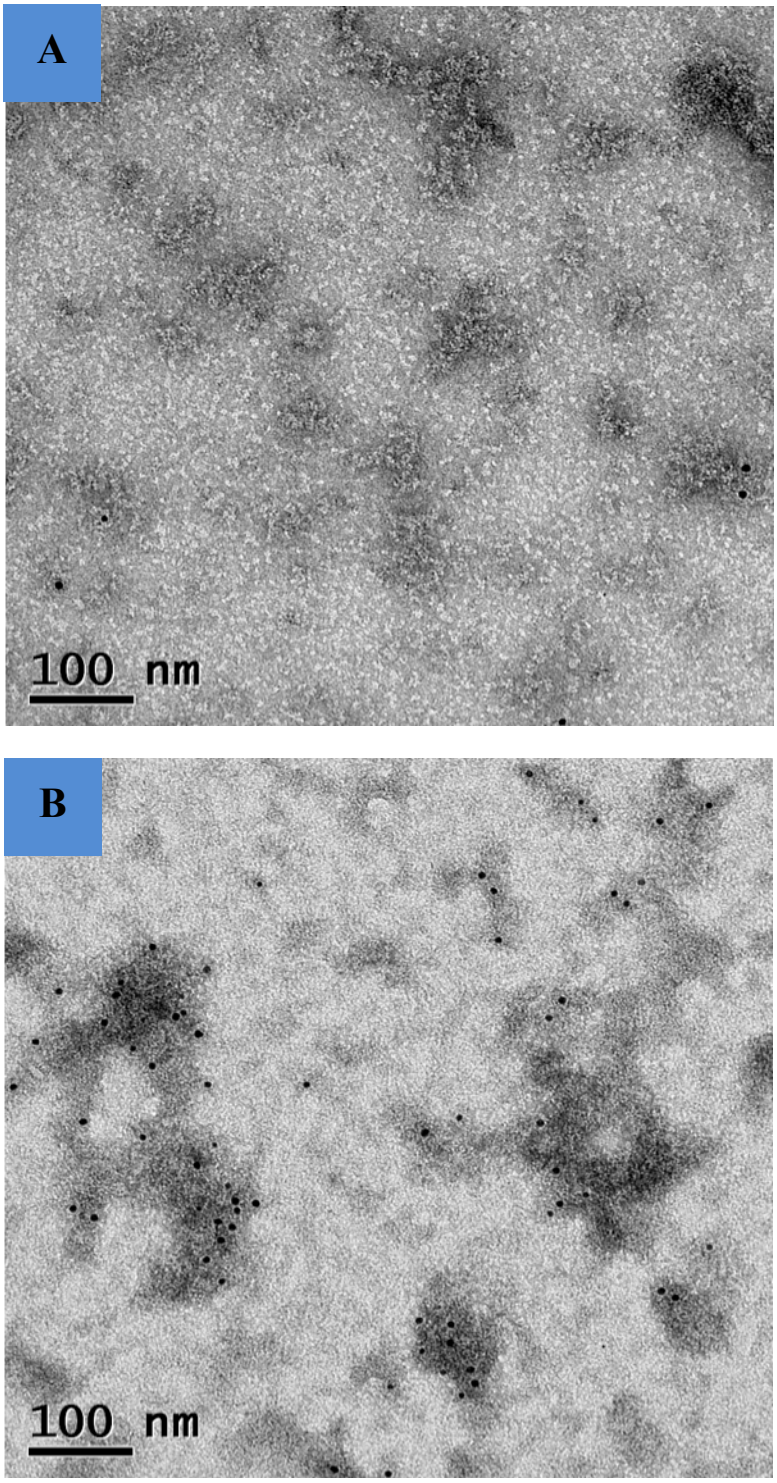


Figure 9.23: Electron micrograph of negative stained immunogold labeled NheB without liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4). The primary antibody 1E11 (A) and 2B11 (B) after incubation for 1h were used**.** There is no formation of specific structures. However, still there is also random labelling (black spots) of the amorphous background material. NheB (0.1 mg/ml).

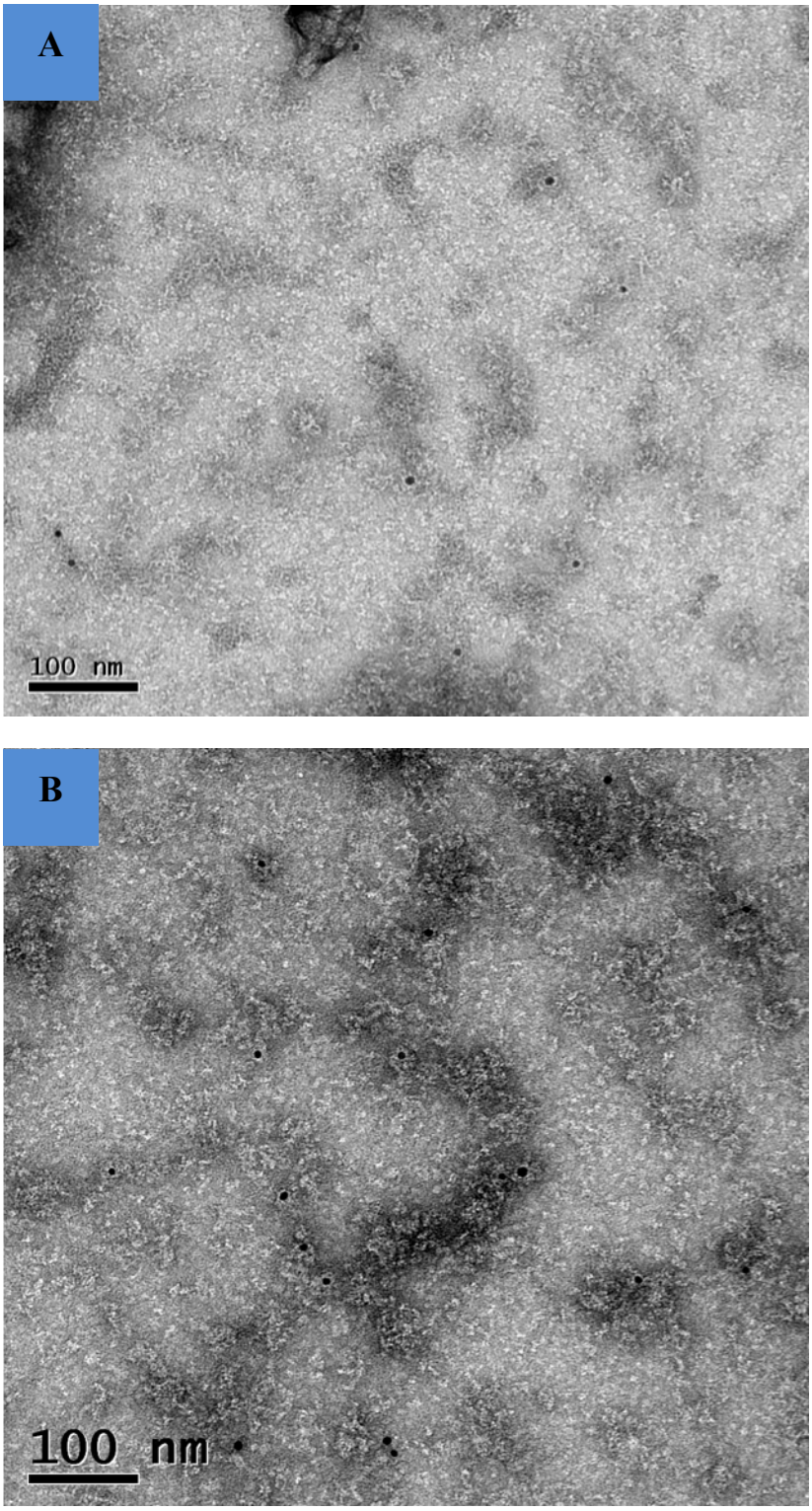


Figure 9.24 A and B: Electron micrograph of negative stained immunogold labeled NheA and NheB in the presence of liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4). (A) The primary antibody 1A8 (after incubation for 1h) was used. (B) The primary antibody 1E11 was used (after incubation for 1h). There is no formation of any clear structures and no specific labeling (black spots). NheA (0.1 mg/ml) and NheB (0.1 mg/ml).

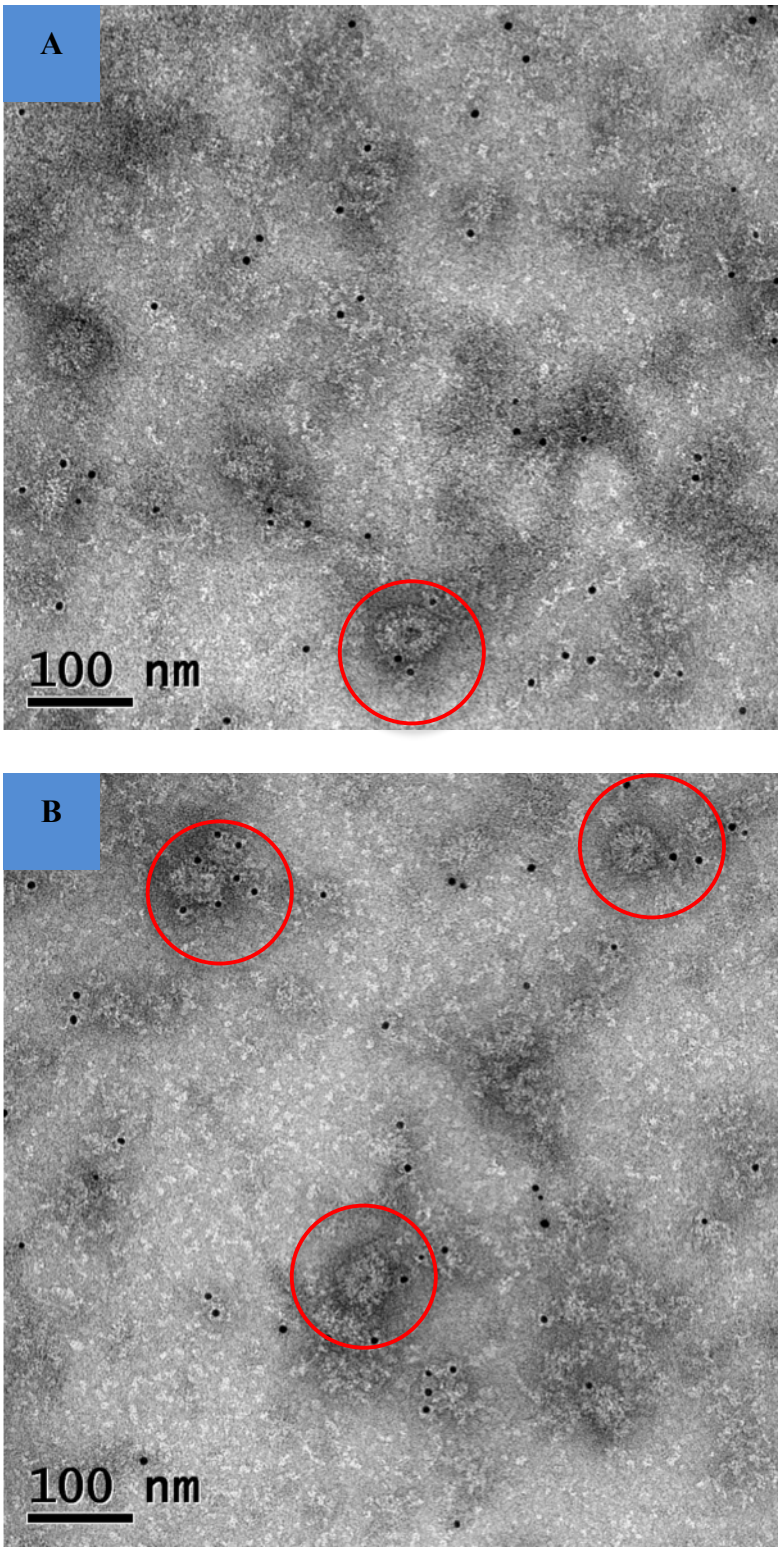


Figure 9.25 A and B: Electron micrograph of negative stained immunogold labeled NheA and NheB in the presence of liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4). The primary antibody 2B11(after incubation for 1h) was used. There are some structures that look like rings (red circles). However, there is also a strong labeling (black spots) to amorphous background material. NheA (0.1 mg/ml) and NheB (0.1 mg/ml).

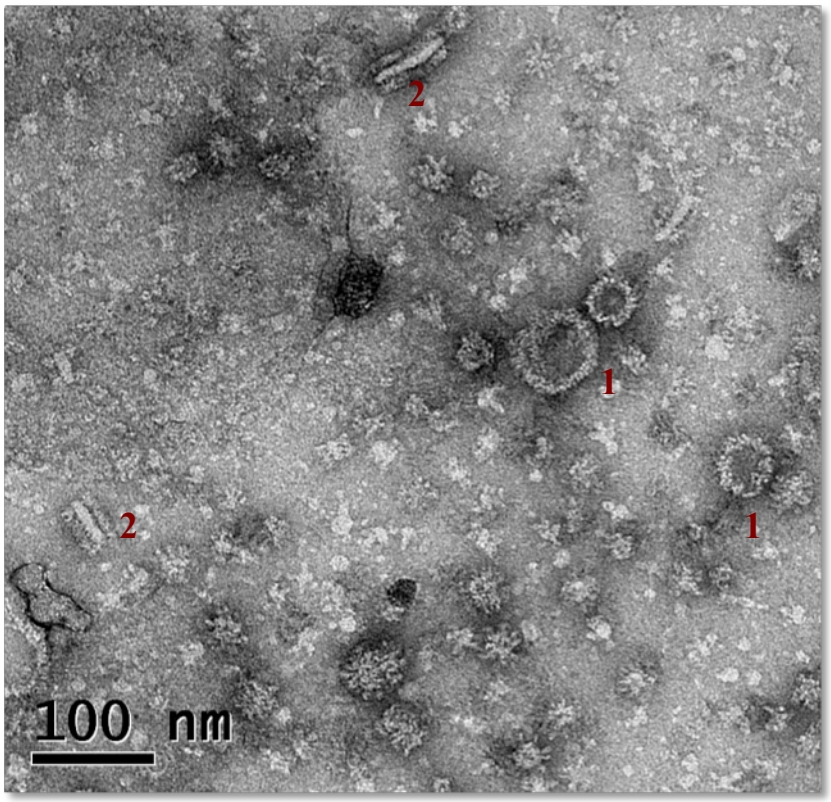


Figure 9.26: Electron micrograph of negative stained NheB with liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4), cross-linked by 50 mM glutaraldehyde.
There is formation of rings (1) and boats (2). NheB (0.1 mg/ml).

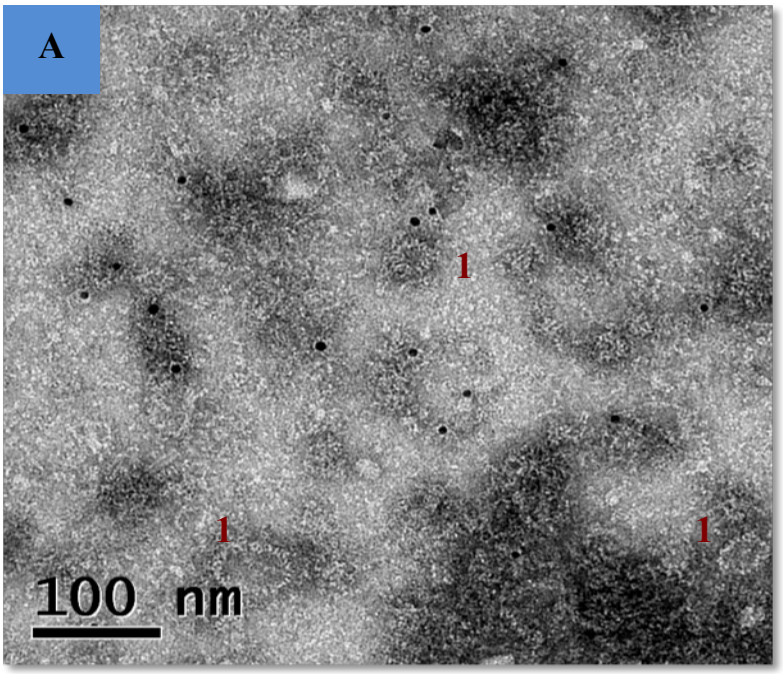


Figure 9.27 A: Electron micrograph of negative stained NheB with liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4), cross-linked by 50 mM glutaraldehyde and then incubated in 1E11. There is formation of rings (1). There is no specific immunogold labeling. However, still there is a random binding to amorphous background material. NheB (0.1 mg/ml).

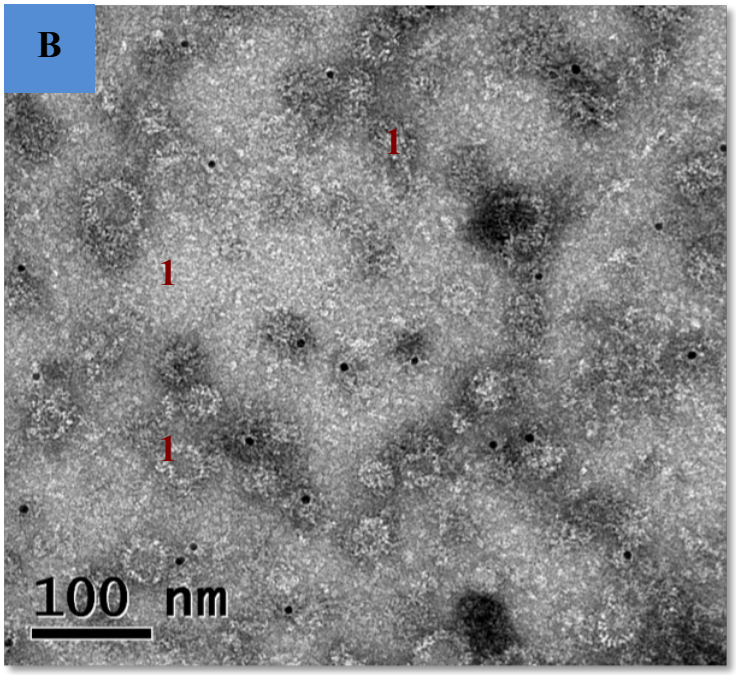


Figure 9.27 B: Electron micrograph of negative stained NheB with liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4) cross-linked by 50 mM glutaraldehyde and then incubated in 2B11. There is formation of rings (1). There is no specific immunogold labeling. However, still there is a random binding to amorphous background material. NheB (0.1 mg/ml).

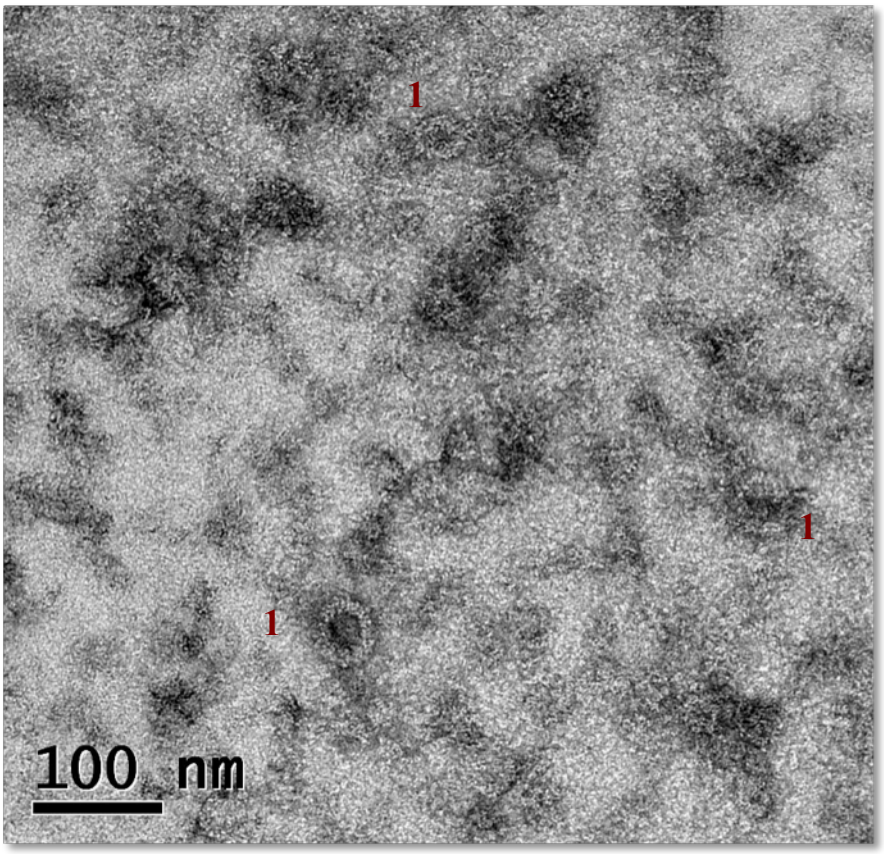


Figure 9.28: Electron micrograph of negative stained NheB with liposomes preparation (method 2 was used for lipid preparation as described in section 3.10.4) cross-linked by 50 mM glutaraldehyde and then incubated in NheA. There is formation of rings (1). NheB (0.1 mg/ml).

9.4 The first observation of *B. cereus* strain NHV 0075/95 by SEM & TEM

B. cereus is a gram-positive bacteria, and causes many food poisoning cases in humans due to the formation of spores which survive inadequate cooking.

By using scanning electron microscopy (SEM) and transmission electron microscopy (TEM) the first experiments were carried out to observe *B. cereus* strain NHV 0075/95. This was in order to find more information about this strain, which does not carry the *hbl* genes and just produces the Nhe pore forming toxins. No other toxins have been recognized and secreted by *B. cereus* strain NHV 0075/95 (Lund & Granum, 1996).

The specimen was prepared according to the techniques of (Bozzola & Russell 1999) techniques using the Philips XL-20 SEM and the FEI Tecnai G2 Spirit TEM in the Department of Biomedical Science, University of Sheffield, UK. This work was done with the assistance of Dr. Christopher J Hill.

In (Figure 9.29) the surface structures of *B. cereus* strain NHV 0075/95 can be seen by SEM. The bacterium is a rod shaped structure, about 3 μm long and 0.7 μm wide and appears very similar to other strains of *B. cereus.*

Internal cellular structures of *B. cereus* strain NHV 0075/95 can be observed by TEM (Figure 9.30 A). The cell wall and many organelles such as the genetic material, mesosome, and ribosomes can be seen.

Studying internal structures of *B. cereus* strain NHV 0075/95 revealed a strange ring- shaped structure located behind the cell wall (Figure 9.30 B). This structure has never been noticed in any strains of *B. cereus* or other bacteria before (Prof. Anne Moir, Prof. Jeff Green and others, personal communication). So what is it and what is function? If you look closely, you can see evidence for secreted materials around this structure and it is tempting to speculate that this structure may be related to a new secretion

system in the bacterium. This observation could be open a new research window for microbiologists in the future.

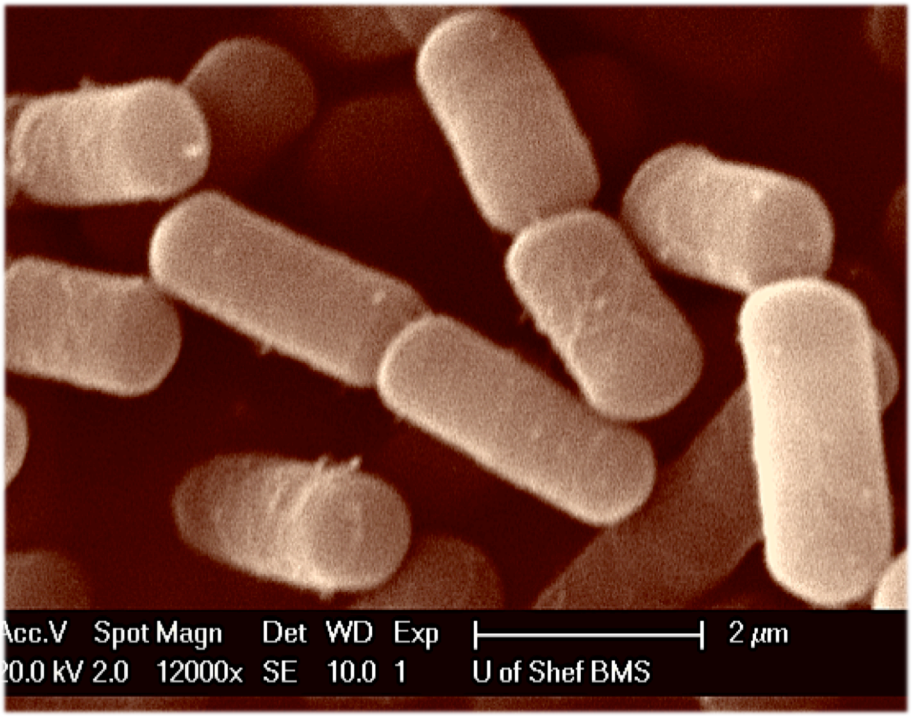


Figure 9.29: Scanning electron micrograph of *B. cereus* strain NHV 0075/95. *B. cereus* strain NHV 0075/95 is rod shaped strcuure and , about 3µm long and 0.7 µm wide.

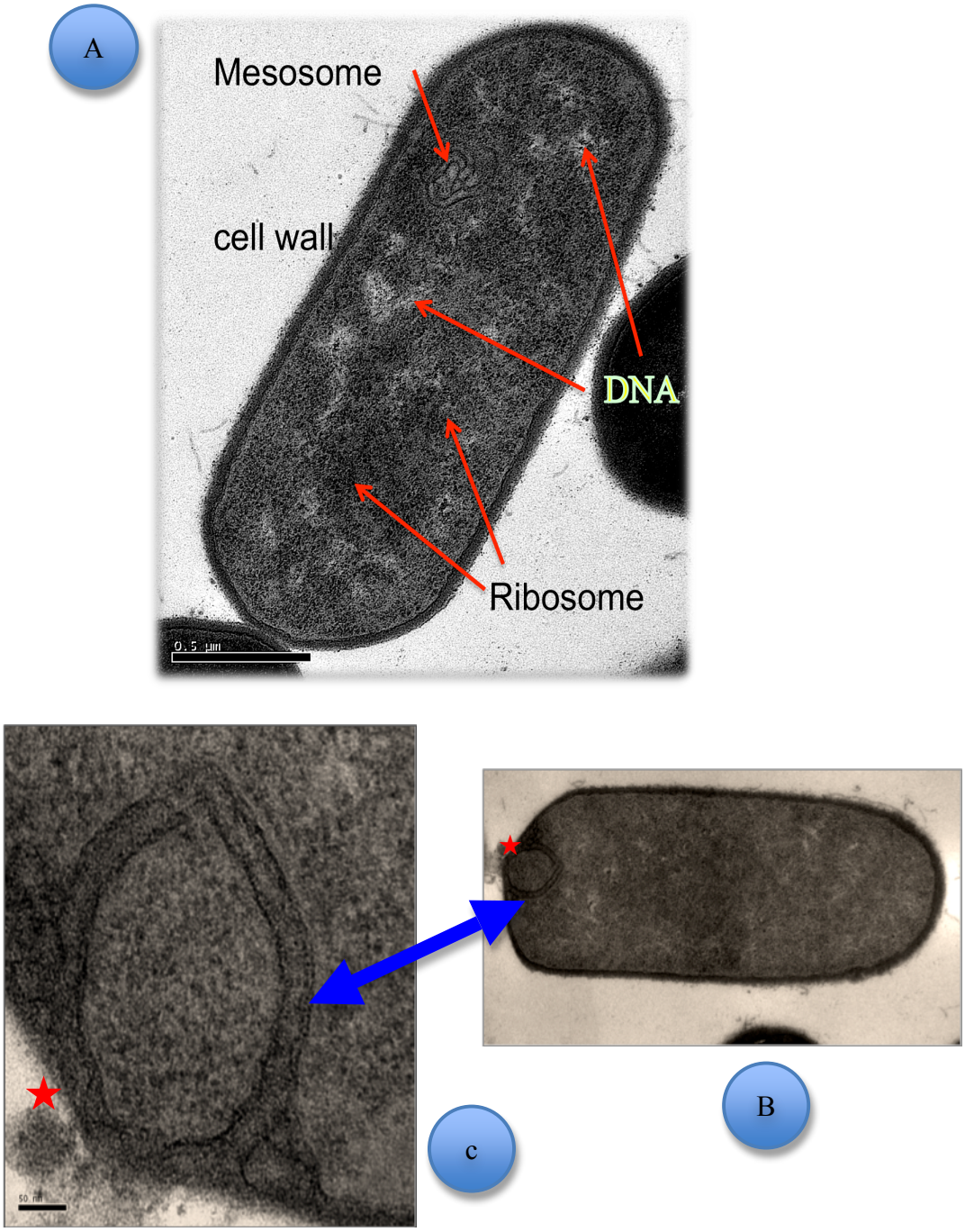


Figure 9.30: Transmission electron micrograph of *B. cereus* strain NHV 0075/95: (A) showing various organelles.
(B) and (C) There is a ring-like structure has never been seen in *B. cereus* before. It looks like a ring in shape and has unknown structure and function (blue pointer). The red stars show the secretion material around this structure.

References

Abeyrathne PD, Chami M, Pantelic RS, Goldie KN, Stahlberg H (2010) Preparation of 2D crystals of membrane proteins for high-resolution electron crystallography data collection. In *Methods in Enzymology*, Jensen GJ (ed), Vol. 481, pp 25-43.

Bozzola J, Russell LD (1999) *Electron microscopy*, 2nd edn.: Jones & Bartlett Publisher.

Damink L, Dijkstra PJ, Vanluyn MJA, Vanwachem PB, Nieuwenhuis P, Feijen J (1995) Glutaraldehyde as a cross-linking agent for collagen-based biomaterials. *J Mater Sci Mater Med* **6:** 460-472

Dietrich R, Moravek M, Burk C, Granum PE, Martlbäuer E (2005) Production and characterization of antibodies against each of the three subunits of the *Bacillus cereus* nonhemolytic enterotoxin complex. *Appl Environ Microbiol* **71:** 8214-8220

Fagerlund A (2007) *Bacillus cereus* Hbl, Nhe and CytK cytotoxins; mecanism, regulation, and secretion. PhD Thesis, Norwegian School of Vetlerinary Science, Norway

Glaeser R, Downing K, DeRosier D, Chiu W, Frank J (2007) *Electron Crystallography of Biological Macromolecules*: Oxford University Press.

Granum PE, O'Sullivan K, Lund T (1999) The sequence of the non-haemolytic enterotoxin operon from *Bacillus cereus*. *FEMS Microbiol Lett* **177:** 225-229

Lindbäck T, Fagerlund A, Rodland MS, Granum PE (2004) Characterization of the *Bacillus cereus* Nhe enterotoxin. *Microbiology-SGM* **150:** 3959-3967

Lindbäck T, Hardy SP, Dietrich R, Sødring M, Didier A, Moravek M, Fagerlund A, Bock S, Nielsen C, Casteel M, Granum PE, Maertlbauer E (2010) Cytotoxicity of the *Bacillus cereus* Nhe enterotoxin requires specific binding order of its three exoprotein components. *Infect Immun* **78:** 3813-3821

Lund T, Granum PE (1996) Characterisation of a non-haemolytic enterotoxin complex from *Bacillus cereus* isolated after a foodborne outbreak. *FEMS Microbiol Lett* **141:** 151-156

Mueller M, Grauschopf U, Maier T, Glockshuber R, Ban N (2009) The structure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism. *Nature* **459:** 726-730

Ohi M, Li Y, Cheng Y, Walz T (2004) Negative staining and image classification - powerful tools in modern electron microscopy. *Mol Microbiol* **20:** 191-199

Tilley SJ, Orlova EV, Gilbert RJC, Andrew PW, Saibil HR (2005) Structural basis of pore formation by the bacterial toxin Pneumolysin. *Cell* **121:** 247-256

Wallace AJ, Stillman TJ, Atkins A, Jamieson SJ, Bullough PA, Green J, Artymiuk PJ (2000) *E. coli* hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy. *Cell* **100:** 265-276

Wyborn NR, Clark A, Roberts RE, Jamieson SJ, Tzokov S, Bullough PA, Stillman TJ, Artymiuk PJ, Galen JE, Zhao L, Levine MM, Green J (2004) Properties of haemolysin E (HlyE) from a pathogenic *Escherichia coli* avian isolate and studies of HlyE export. *Microbiology* **150:** 1495-1505

Chapter 10: Summary, Conclusion and Future work

This chapter summarizes and describes the major findings of this thesis on the APEC ClyA protein from *E. coli* and the Nhe proteins from *B. cereus*.

Pore Forming Toxins (PFTs) are one of the major groups of bacterial toxins and have been classified into two divisions according to the category of the secondary structure utilized to insert into the phospholipid bilayers of plasma membrane (Gilbert, 2002; Tilley et al, 2005):

1) α PFTs

These form α-helical channels. Pore formation by α–PFTs involves three steps: binding to a specific receptor, formation of the pore and insertion in the membrane. e.g Colicins, and Diphtheria toxin (Tilley et al, 2005).

2) β PFTs

These can form β sheet structures such as β barrels. This type of toxin requires an important refolding step during insertion in the membrane so they can form trans-membrane β barrels. e.g. the α-Hemolysin family, which are also known as repeat in toxin (RTX) toxins, Aerolysin toxins, Anthrax toxin, and Cholesterol-dependent cytolysins (Tilley et al, 2005).

E. coli K12 ClyA revealed the first structure of a novel family of cytolytic toxins (Wallace et al, 2000) and it is classified as α-helical pore-forming (Mueller et al, 2009). APEC ClyA protein is a cytolytic toxin with similar characteristics to K-12 ClyA. Having successfully optimized the conditions for expression and purification of APEC ClyA, crystals of probable APEC ClyA pore form were grown in different concentrations of PEG400 (10-15%), 50 mM Tris HCl at pH 8.5, 100 mM $MgCl_2$ and detergent in different forms, which diffracted poorly to >10Å resolution. Further methods to

improve the diffraction of APEC ClyA pore form crystals were attempted and the best diffraction obtained was only to 7 Å by synchrotron radiation in the presence EMP. Using hemolytic activity experiments we were able to prove that the APEC ClyA is in active form and in folded conformation.

Recently, the X-ray crystal structure of the *E. coli* K-12 ClyA pore form was published (Mueller et al, 2009) and this structure had a 12-meric pore. High sequence homology (about 75% identity) between *E. coli* K-12 ClyA and APEC ClyA, suggested that they would share the same large conformational changes on pore formation.

Bacillus cereus is gram-positive, rod shaped bacterium, widespread in the environment, which can be isolated from soil and contaminated food. Due to the ability of *B. cereus* to form spores it is involved in two types of food poisoning: the emetic and diarrheal types. *B. cereus* bacteria have three putative enterotoxins: hemolysin BL (Hbl), non-hemolytic enterotoxin (Nhe) and cytotoxin K (CytK). Both Hbl (HblL1, HblL2, HblB) and Nhe (NheA, NheB, NheC) are tripartite pore forming toxins. Hbl and Nhe are encoded by three genes cotranscribed as operons: *hblCDA* encodes Hbl proteins L2, L1 and B, and *nheABC* encodes NheA, NheB and NheC. There is sequence similarity between the three components in each complex and between the proteins of Nhe and Hbl (Fagerlund et al, 2008).

Despite low sequence similarities between Hbl and ClyA, the 3D crystal structure of Hbl component B from *B. cereus* shared significant structural homologies with *E. coli* ClyA (Madegowda et al, 2008).

This thesis presents the first structure of an Nhe protein, component A from *B. cereus*. NheA and a SeMet-derivative have been expressed successfully and purified to homogeneity using several purification protocols including an anion exchange column, hydroxylapatite chromatography and gel filtration. NheA and SeMet NheA were successfully crystallized using PEG3350 as a precipitant by the sitting-drop vapour-diffusion method. A full diffraction data set was collected using synchrotron facility to 2.05 Å resolution from a single crystal of NheA. The Matthews coefficient

suggested that 8 molecules could be in the asymmetric unit, giving a V_M of 2.20 Å^3/Dalton and a solvent content of 44.52% (Matthews, 1968). The structure of NheA was solved using SeMet MAD data sets, which was collected from a SeMet NheA crystal.

NheA has mostly an α-helical bundle architecture. It is an elongated, rod shaped structure with dimensions 95 Å X 40 Å X 20 Å. It consists of seven α-helices, known as α A, α B, α C, α F, α G, α D, α E. NheA has two sub-domains the "Tail" domain, and the "Head" domain, which includes the "ß tongue" which contains two strands ß1 and ß 2. A third ß strand, ß3, forms the C-terminal and forms hydrogen bonds to the ß tongue.

The structure of NheA reveals a strong similarity in tertiary structure with HblB from *B. cereus* and ClyA from *E. coli*. The major interesting differences between them are in the orientation of head domains of the proteins. The head domain is turned downward and interacts with the tail domain in NheA and HblB (the interdomain angles are ~20° and ~30°, respectively), however it is turned upward in ClyA with only small interactions with its tail domain (~120° interdomain angle).

The structural and functional similarities among Nhe, Hbl and ClyA appear to indicate that they belong to a superfamily of pore-forming toxins. Furthermore, the different orientations of the head domains in ClyA, NheA and HblB may suggest a degree of flexibility in the β tongue region but could also represent different conformations in these toxins. As more 3D crystal structures of other toxins from different species are determined, it is likely that the number of ClyA homologues identified will increase and so it will become apparent if ClyA, Hbl and Nhe are really member of a novel superfamily of pore-forming toxins.

The elongated amphipathic beta sheet of NheA molecules is reminiscent of transmembrane sequences in some toxins for which structures have been reported, such as αHL (Song et al, 1996), and LukF (Olson et al, 1999). However, these are based on β sheet rather than α helical architectures. Therefore could Nhe form a transmembrane β-barrel like *S. aureus* α-toxins

and could the Nhe be considered as a possible link between the α and β pore forming toxins?

This project also involved the successful expression, purification and crystallization of the second, NheB, component of Nhe toxin from *B. subtilis* strain JH642. A full diffraction data set was collected under cryoconditions using synchrotron radiation to 3.5 Å resolution from a single crystal of NheB. The space group was determined to be primitive hexagonal ($P6_2$ or $P6_4$) with unit cell dimensions a = 130.89 Å, b = 130.89 Å, c=61.86Å and ϒ=120°. However, the attempts within this thesis to determine the structure of NheB by molecular replacement in space group $P6_2$ and $P3_2$ or one of their alternative space groups were unsuccessful. There are two considerations, which may explain the unsuccessful results: The first is the lack of a homology model with high sequence similarity with NheB, and the second is the lack of a major high- resolution data.

Observations of NheA and NheB by electron microscopy show that NheA does not form specific structures with lipid. NheB can form different size of rings and structure that looks like a boat. It can form long sticks in the presence of detergent. NheA and NheB together give strange network materials, which looks like an assembly of pores in the presence of detergent, and form small rings and a pore like structure with internal diameter 7.5 to 10 nm. However, adding NheC not did change that result which was not surprising as it found that NheA and NheB are sufficient for pore formation and cell lysis (Lindbäck et al, 2010). However, addition of NheC is important for full cytotoxicity in some cell types (Vero cells) (Lindbäck et al, 2010).

Many questions concerning the mechanism of Nhe pore formation remain unanswered and require extensive biochemical and structural studies.

Future work should focus on further expression, purification and crystallization trials of the NheB protein. The main goal of future study on Nhe toxin should be the obtaining of detailed, high-resolution data which

will enable description of the transmembrane assembly of the Nhe pore complex by X-ray crystallography or cryo-EM.

References

Fagerlund A, Lindback T, Storset AK, Granum PE, Hardy SP (2008) *Bacillus cereus* Nhe is a pore-forming toxin with structural and functional properties similar to the ClyA (HIyE, SheA) family of haemolysins, able to induce osmotic lysis in epithelia. *Microbiology* **154:** 693-704

Gilbert RJC (2002) Pore-forming toxins. *Cell Mol Life Sci* **59:** 832-844

Lindbäck T, Hardy SP, Dietrich R, Sødring M, Didier A, Moravek M, Fagerlund A, Bock S, Nielsen C, Casteel M, Granum PE, Maertlbauer E (2010) Cytotoxicity of the *Bacillus cereus* Nhe enterotoxin requires specific binding order of its three exoprotein components. *Infect Immun* **78:** 3813-3821

Madegowda M, Eswaramoorthy S, Burley SK, Swaminathan S (2008) X-ray crystal structure of the B component of hemolysin bl from *Bacillus cereus*. *Proteins-Structure Function and Bioinformatics* **71:** 534-540

Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* **33:** 491–497

Mueller M, Grauschopf U, Maier T, Glockshuber R, Ban N (2009) The structure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism. *Nature* **459:** 726-730

Olson R, Nariya H, Yokota K, Kamio Y, Gouaux E (1999) Crystal structure of Staphylococcal LukF delineates conformational changes accompanying formation of a transmembrane channel. *Nat Struct Bio* **6:** 134-140

Song LZ, Hobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux JE (1996) Structure of Staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **274:** 1859-1866

Tilley SJ, Orlova EV, Gilbert RJC, Andrew PW, Saibil HR (2005) Structural basis of pore formation by the bacterial toxin Pneumolysin. *Cell* **121:** 247-256

Wallace AJ, Stillman TJ, Atkins A, Jamieson SJ, Bullough PA, Green J, Artymiuk PJ (2000) *E. coli* hemolysin E (HlyE, ClyA, SheA): X-ray crystal structure of the toxin and observation of membrane pores by electron microscopy. *Cell* **100:** 265-276