STRUCTURAL STUDIES OF THE BACTERIOPHAGE LAMBDA HOLIN AND THE *M. tuberculosis* SecA TRANSLOCASE

A Dissertation

by

GEORGE CHRISTOS WILLIAM SAVVA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2007

Major Subject: Biology

STRUCTURAL STUDIES OF THE BACTERIOPHAGE LAMBDA HOLIN AND THE *M. tuberculosis* SecA TRANSLOCASE

A Dissertation

by

GEORGE CHRISTOS WILLIAM SAVVA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee, Andreas Holzenburg
Committee Members, Ryland Young
James Sacchettini
Deborah Siegele
Head of Department, Vincent Cassone

December 2007

Major Subject: Biology

ABSTRACT

Structural Studies of the Bacteriophage Lambda Holin and the *M. tuberculosis* SecA Translocase.

(December 2007)

George Christos William Savva, B.S., University of Leeds, UK

Chair of Advisory Committee: Dr. Andreas Holzenburg

Double stranded DNA bacteriophages achieve release of phage progeny by disrupting the cell envelope of the host cell. This is accomplished by two phage-encoded proteins, the holin and the endolysin. In bacteriophage lambda, the S holin is a small three TMD membrane protein that creates a lesion in the inner membrane of the host at a specific time, programmed in its primary structure. Lesion formation permits the cytoplasmic endolysin R access to the murein cell wall for degradation and cell lysis. Although it has been shown that S oligomerizes in the membrane, the structural nature of this complex has not been elucidated. In this study the S holin was purified using a mild non-ionic detergent and the structure of a ring complex formed by the holin was determined by electron microscopy and single particle analysis at a resolution of 2.6 nm. Biochemical characterization of the rings suggests that such a complex might represent the assembly formed by S in the membrane.

Protein translocation in all organisms allows the export of proteins destined for localization outside the cytoplasm. In eubacteria, newly synthesized proteins are directed to the heterotrimeric membrane complex SecYEG by signals embedded in their sequence. The driving force through this complex is provided by the cytoplasmic ATPase SecA which combines ATP hydrolysis to mechanically insert proteins through the protein conducting channel. Using electron microscopy and single particle analysis we have obtained the structure of SecA from *M. tuberculosis*. The structure indicates that four SecA monomers assemble to form an elongated molecule with D2 symmetry. Docking of the EM map to the crystal structure of tb SecA confirms this arrangement of the subunits. This finding, that *M. tuberculosis* SecA forms a tetramer raises intriguing possibilities about SecA function.

DEDICATION

To George, Kathy and Jacky

ACKNOWLEDGEMENTS

I thank my advisor Dr. Andreas Holzenburg, for his support, guidance and friendship and for introducing me to the field of macromolecular EM.

I thank my committee members: Dr. Debby Siegele, Dr. Jim Sacchettini and Dr. Ry Young for their support and guidance throughout this study. I thank Dr. Sacchettini for allowing me to conduct research in his laboratory. I thank Dr. Young also for allowing me to work in his laboratory and for looking out for the well being of myself and other graduate students.

I thank members of the Sacchettini and Young groups for endless help and support, in particular, Dr. John Deaton, Brenley Macintosh, Rebecca White, Jill Dewey, and Sudharsan Sridharan.

I would also like to thank Dr. Arokiasamy Arulandu for introducing me to many of the biochemical techniques I used. I thank Dr. Douglas Struck for endless guidance and precious advice during the past few years. I thank Dr. Chris Gilpin for help in collecting the S105 cryo-EM dataset.

Finally, thank you to Dr. Arati Ramesh for her friendship, scientific discussions and advice, and for her support throughout this study. Thank you, Arati.

NOMENCLATURE

ATP	Adenosine 5'-triphosphate
APS	Advanced photon source
CMC	Critical micelle concentration
DNA	Deoxyribonucleic acid
DSP	Dithiobis (succinimidyl) propionate
DTT	Dithiothreitol
dNTP	Deoxyribonucleotide triphosphate
EBB	Empigen BB
EM	Electron microscopy
Hepes	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG	isopropyl β-D- thiogalactopyranoside
L	Liter
LB	Luria- Bertani
β-ME	β-mercaptoethanol
NMR	Nuclear magnetic resonance
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulfate
TB	Terrific broth

TMD	transmembrane domain
Tris	2-amino-2-hydroxymethyl-1,3-propanediol
X	Any amino acid

TABLE OF CONTENTS

Page

ABSTRACT……..……..iii

DEDICATION……...……v

ACKNOWLEDGEMENTS……...……vi

NOMENCLATURE…….....vii

TABLE OF CONTENTS ……...……...ix

LIST OF FIGURES……...……xi

CHAPTER

I INTRODUCTION……...……...1

1.1 Introduction to macromolecular EM …….……...1

1.2 Introduction to the bacteriophage lambda holin ……...……..…….24

1.3 Introduction to the Sec translocation system.....34

II *IN VITRO* STRUCTURAL AND BIOCHEMICAL STUDIES OF THE BACTERIOPHAGE LAMBDA HOLIN ……...……...…44

2.1 Introduction…….…....44

2.2 Results…….……44

2.3 Discussion……...…….……82

2.4 Experimental procedures …….……...……...100

III STRUCTURAL STUDIES OF *M. tuberculosis* SecA TRANSLOCASE …...……110

3.1 Introduction ……..…….110

CHAPTER Page

3.2 Results ……..……...……110

3.3 Discussion ……....……..…..…….122

3.4 Experimental procedures..……...……125

IV CONCLUSIONS AND FUTURE DIRECTIONS …….129

4.1 Electron microscopy as a structural characterization method ..……...129

4.2 S105 forms rings of large diameter …….131

4.3 The structure of *tb* SecA …….….134

REFERENCES ……...136

VITA……...147

LIST OF FIGURES

FIGURE Page

1.1 Structure determination by EM and the CTF ….……...……...5

1.2 Macromolecular machines solved by single particle EM ……..13

1.3 MSA classification and angular reconstitution …….....20

1.4 Lambda lysis genes and gene products……...…….....27

1.5 Schematic representation of protein translocation in *E. coli*.……35

1.6 Sequence alignment of tb SecY and *M. jannaschii* SecYα.....……...41

2.1 Gel permeation chromatography of S ……....46

2.2 Electron microscopy of S105 gel permeation fractions…….….48

2.3 S105 purification in other non-ionic detergents....…….49

2.4 Oligomerization properties of $S105_{r59c}$...……...…….51

2.5 Single particle analysis of negatively stained S105……53

2.6 Cryo-EM and particle classification of S105……..55

2.7 Three-dimensional reconstruction of S105 ring dimer……...56

2.8 S105 remains soluble in the absence of detergent……..58

2.9 $S105_{c51s}$ does not require DTT to form soluble rings …….……60

2.10 S105 forms rafts of rings at low detergent concentration……...61

2.11 Circular dichroism spectroscopy of S alleles …….63

2.12 DSP crosslinking of S in *E.coli* membranes ……..65

2.13 DSP crosslinked S105 gel-permeation results and EM ……..67

FIGURE Page

2.14 Enrichment of S IMVs from *E. coli* C43(DE3)……..69

2.15 Flotation gradient centrifugation of S IMVs ……..71

2.16 Liposome reconstitution and sucrose gradient flotation assay ……...74

2.17 Liposome reconstitution of S105 from EBB ……..75

2.18 Liposome reconstitution of S105 from DDM…….76

2.19 Cysteine accessibility and nanogold labelling of S105 DDM rings ……...78

2.20 Protease digestion of S alleles ……81

2.21 Model for S mediated lysis ……98

3.1 Electron micrographs of *tb* SecA……..112

3.2 Single particle classification and 3D reconstruction of *tb* SecA ……..114

3.3 Comparison of EM map to tb SecA crystal structure ……...116

3.4 Expression studies of tb *secYE* ……119

3.5 Purification of EBB solubilized SecYE……121

CHAPTER I

INTRODUCTION

1.1 INTRODUCTION TO MACROMOLECULAR EM

1.1.1 Introduction to structural biology

In light of the rapid sequencing of genomes including the human genome, a vast amount of information has been gathered that will help in the understanding of cellular processes, identify new genes involved in disease and provide an understanding of the evolutionary relationships between organisms (1). Molecular and cell biologists, biochemists and biophysicists can use the data to manipulate and probe the functions of the thousands of genes and gene products from different organisms and contribute from their own specialization to a greater comprehension of biology. A detailed understanding of the molecular mechanism of these gene products which are mostly proteins, nucleic acids or complexes of the two could not have been achieved without studying what makes these macromolecules unique in their function, i.e. their three-dimensional structure. Determined by the primary structure for proteins and the forces that drive folding such as electrostatic and hydrophobic interactions within the polypeptide and with its environment (2), the structures of proteins provide detailed information on the mechanics, catalysis and interactions with other ligands and can provide a starting point

---

This dissertation follows the style of *Journal of Biological Chemistry.*

for further biochemical and biophysical characterization. Furthermore structural knowledge of macromolecules involved in disease, inherent or foreign, can aid in the design of drugs that prevent their disease-causing attributes (3).

Structural biology is comprised of techniques that strive to determine the structure of macromolecules at the highest resolution, i.e. to a resolution where individual atoms and their bonds can be discerned. The two main methods that are widely used for atomic resolution structure determination are x-ray crystallography and nuclear magnetic resonance spectroscopy (NMR). The former of the two techniques has been used to solve approximately 37,500 of the 44,320 structures that are available in the protein database bank (PDB) while NMR has been used to solve just under 6,500 of these (PDB statistics from www.rscb.org). Although both of these techniques can provide detailed structural information, they both have their respective advantages and limitations. X-ray crystallography requires highly ordered crystal formation by the macromolecules in order to produce diffraction patterns to high resolution. Although this is generally not a problem due to the ability to rapidly screen hundreds of conditions that may favor crystal growth, not all molecules are amenable to crystallization. In particular, multiprotein and protein-nucleic acid complexes are difficult to crystallize due to heterogeneity, conformational variability and other factors. On the other hand crystallography can deal with a relatively broad size range providing homogeneity is present. NMR is ideal for smaller proteins below 40 kDa and does not require ordered assemblies. However both crystallography and NMR are “protein hungry” requiring milligram amounts of protein for structure determination. Standardized overexpression

procedures for producing large amounts of protein can provide enough protein in most cases but not all proteins can be overexpressed successfully. Solubility issues associated with incorrect folding as well as toxicity to the expressing host seriously limit the protein yields. In particular, overexpression of membrane proteins is often problematic (4). To exemplify this, only 256 high resolution structures of membrane proteins have been solved to date with many of these being homologues from different organisms. The development of *in vitro* translation methods as well as the availability of host strains and expression vectors that favor the solubility of problematic proteins may help overcome these protein limitations (4).

A third structural determination technique that aims to bridge the gap between the molecular and cellular level is that of electron microscopy (EM). Currently the PBD contains 146 structures solved by EM, of which some are coordinates produced by fitting pre-existing atomic structures or nucleic acids chains into parts of EM density maps. The range and types of macromolecule categories though reveals the specialization of this technique. Complete viral capsids, ribosomes and integral membrane proteins are all macromolecules that are either too large, too difficult or available in insufficient quantities to be studied by crystallography and NMR. Many more proteins have been studied by EM and only a fraction of the structures are available in the PDB. Furthermore the imaging and 3D reconstruction of complete cells and organelles at nanometer resolutions by EM promises and exciting future and anticipates the day when the whole proteome of organisms can be examined in its native context (5).

1.1.2 Electron microscopy for the determination of macromolecular structure

The transmission electron microscope (TEM) has offered invaluable information to the field of cell biology by providing images of the architecture of cells and their organelles, tissues and even viruses (6). Despite these achievements, the full potential and power of the electron microscope had not been fully appreciated. It was not until the late 1960's that biologists, theoreticians and mathematicians began expanding on a new role for the TEM in the study of macromolecules. The groundbreaking research by DeRosier and Klug (1968) describing the 3D reconstruction of the T4 bacteriophage tail from electron micrographs illustrated that information in the electron micrographs could be extracted, processed mathematically and used to produce information that was originally unappreciated visually. By using the theory and tools developed for x-ray crystallography the authors showed how these could be applied to electron microscopy. Fourier transforms of the tail images produced reflections that arose from the periodic densities in the helical tail (Figure 1.1a, b). Moreover because electrons are focused to produce an image, the phase information is also present in the images and the respective Fourier transforms. These reflections could then be isolated from the noise, thus producing images with an improved signal to noise ratio (Figure 1.1c). Furthermore because of the helical symmetry of the tail, a single projection contained all the information required for a 3D reconstruction. This groundbreaking research set the foundations for electron crystallography and single particle EM that are discussed in more detail below. EM has advanced rapidly over the last 20 years in terms of improvement in electron optics, computational methods and in experimental techniques

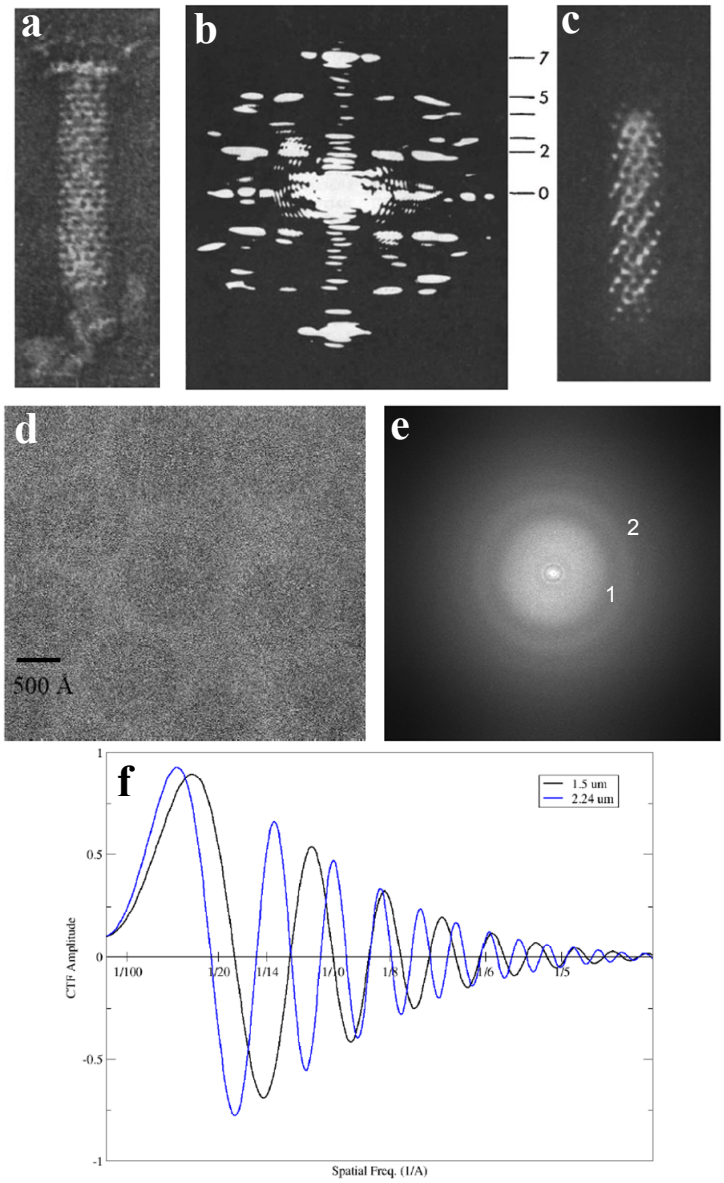

Figure 1.1 Structure determination by EM and the CTF
Structure determination of the T4 bacteriophage tail by EM (DeRosier and Klug, 1968) showing the tail image (a), Fourier transform (b) and a one sided projection of the tail after removing noise (c). The effect of defocus on the contrast transfer function (CTF) showing the oscillations of the CTF at different spatial frequencies. Image of herpes simplex virus (d), the corresponding fast Fourier transform (FFT) (e) and the CTF amplitude plotted against resolution (f). When the CTF crosses the x-axis no information is transferred to the image (zeroes of the CTF) and when it becomes negative the contrast is reversed. Numbers in the FFT show the dark circles that correspond to the zeroes. Unless corrected information beyond the first zero is unreliable.

aimed at determining macromolecular structure. The role of EM as a complementary technique alongside X-ray crystallography and NMR is discussed as well as the unique capabilities that make it an indispensable tool in the postgenomic era.

1.1.3 Image formation in the electron microscope

The image formed on photographic film or the viewing screen of the electron microscope is a result of the interaction of electrons with the specimen and the optical properties of the electron microscope. As electrons approach the atoms that comprise the specimen, they are scattered by the attraction to the nucleus of the atoms (7). Electrons that interact briefly with the specimen without losing energy are termed “elastically scattered” electrons whereas electrons that interact with atomic electrons and lose energy are known as “inelastically scattered electrons”. Two contrast mechanisms contribute to the image formed. Amplitude or scattering contrast results from scattering of electrons by the specimen and is affected by the atomic number (Z) of the atoms in the specimen. The elastic scattering increases with Z, so staining with heavy metal salts allows for better contrast in the images (7). The second contrast mechanism arises from changes to the defocus of the objective lens that results in the interference of unscattered and elastically scattered electrons which are shifted in phase. This contrast mechanism is referred to as phase contrast.

Inelastically scattered electrons that have lost energy do not contribute to high resolution information in the image and are not affected by the atomic number of the specimen (7). Furthermore they cannot be focused and contribute to noise and

background in the image. Finally these electrons are responsible for most of the radiation damage associated with the electron beam. Their contribution becomes more apparent when specimens are unstained and the elastic component is low or in specimens with increased thickness leading to an increase of the inelastic electron population. The chromatic aberration of the objective lens i.e. electrons of different wavelengths being refracted differently, leads to the focal planes of the inelastic and elastic components being different thus causing a blurriness in the image. Modern microscopes can now filter out the inelastic electrons leading to an increase in signal in both unstained and thick specimens, which is especially important in cryo-electron tomography (5,8).

The image formed in the electron microscope is affected by the contrast transfer function (CTF) of the microscope that describes the effect of the spherical aberration coefficient (*Cs*) of the microscope, the defocus (*Δf*) and the accelerating voltage on how the contrast is transferred to the image (8). The CTF can have deleterious effects on the image by affecting the contrast at different spetial frequencies, therefore making the image not a true representation of the specimen. Ideally the contrast should remain unchanged across different resolutions, but, due to the CTF, the contrast oscillates depending on the defocus used (Figure 1.1 d, e, f). In addition at certain spatial frequencies, no amplitude is transferred leading to missing information at that resolution (8). The effect on the image can be a complete reversal of the contrast, so that a protrusion on the specimen appears like a cavity or hole. Correction for the effects of the CTF is now routinely performed when high resolution information is to be extracted from electron micrographs and is discussed later in this chapter (9).

1.1.4 EM structure determination techniques

Electron crystallography

The use of electrons to study two-dimensional arrays of macromolecules or 2D crystals is known as electron crystallography. As in x-ray crystallography, scattering of waves (electrons instead of x-rays) by the lattice produces a diffraction pattern which contains the amplitude information of the array (7). However diffraction data need not be collected on the microscope at the diffraction plane but can be calculated from the image directly by a fast Fourier transform (FFT). In addition to the amplitude data, the image FFT contains the phase information of the lattice, and a 2D projection free of noise can be retrieved from the image. If the diffraction pattern had been acquired directly from the TEM (at the back focal plane) the phase information would have been lost. The 2D crystal can then be tilted in the TEM by +/- 60º to collect images from different angles and the 2D projections combined and back-projected to produce a 3D density of the macromolecule. This density is called a "Coulomb potential" since the electrons are scattered by the protons and not the electrons as in x-ray diffraction (7).

As in x-ray crystallography, the quality of the crystal ultimately determines the resolution of the data. Typically well ordered large (in the region of a few 1000 unit cells) 2D crystals are required which compared to the millions used from 3D crystals is quite small. So typically the amount of protein required for 2D crystals is much less than that required for 3D. However the molecules in 2D crystals are more radiation prone since a smaller area is being subjected to the beam. Because screening for crystals requires an electron microscope, the time spent screening is significantly larger. In an

ideal situation screening of a hundred 3D crystal conditions with a light microscope would take approximately 10 minutes. This is the minimum amount of time required to screen a single condition for 2D crystals although in reality it is often longer. Ultimately this limits the number of conditions that can be tried and remains a challenge in electron crystallography.

One of the main advantages of electron crystallography is in the study of membrane proteins because they can be crystallized and imaged within the lipid bilayer (10). This in turn allows membrane proteins to be captured in their most native state. Moreover, the bilayer offers an advantage as it acts as a 2D plane that aids the formation of the 2D lattice. The first membrane protein structure was solved by 2D crystallography and was another landmark in macromolecular EM. Henderson and Unwin (1975) solved the structure of bacteriorhodopsin (BR) from the purple membrane of *Halobacterium halobium* (11).This membrane containes naturally existing 2D lattices of BR which functions as a light driven ion pump in photosynthesis. The 7Å resolution structure also provided the first direct confirmation that membrane proteins consist of α-helices that span the membrane width, almost perpendicular to the plane of the bilayer.

Since then the structures of a number of membrane proteins have been solved at atomic resolution by electron crystallography including bacteriorhodopsin (12), the plant light harvesting complex-II (13) and aquaporin-1 (14). Recently a 1.9 Å structure of aquaporin-0 was solved by electron crystallography that also allowed visualization of the lipid molecules interacting with the protein (15).

Although 2D crystallography has been primarily used for membranes proteins it

has also allowed for soluble protein structures to be determined. Because initial crystal growth requires a nucleation point and an expansion of that nucleation in two dimensions, success in 2D crystallization has been enhanced by providing a surface for such a nucleation to occur. Lipid monolayers at the air-water interface have been especially successful as crystallization surfaces. Interactions between the phospholipid head groups and the macromolecules can range from general electrostatic (16) to specific ligand interactions (17). More recently the use of Ni-NTA modified lipids has allowed the crystallization spectrum using the lipid monolayer to expand to potentially any protein that carries a His_6 tag (18). The immobilization of the macromolecules at the lipid monolayer also facilitates the crystallization process by inducing unique orientations of the molecules in respect to the ligand and by providing a high local concentration in the order of 100 mg/ml from a dilute protein sample of only a few micrograms/milliliter (4).

Electron crystallography offers a unique advantage for the determination of membrane protein structures and for soluble proteins that cannot be purified in milligram amounts, but the difficulty in obtaining good crystals combined with the longer screening times limits the speed with which structures can be solved. To reach atomic resolutions (<3Å) may take years of optimization and data collection. In recent years the number of membrane proteins solved by X-ray crystallography has outnumbered those solved by EM. The ability to screen hundreds of conditions quickly combined with robotic automation for crystallization trials and the availability of various detergents has produced fruitful results (4). However electron crystallography offers examination of the

protein in its most native environment, the lipid bilayer.

Single particle analysis

The difficulty associated with crystal growth has lead to the development of alternative methods for structural determination by EM. The single particle method, as the name implies, utilizes single images of macromolecules that are dispersed and randomly oriented on a carbon support film or embedded in amorphous ice (19). This technique removes the need for crystals but after raw data acquisition it becomes computationally demanding. In one way, the protein is crystallized *in silico* (8). The general principle behind single particle EM is that all the molecules in the dataset represent various orientations of the same 3D object. These projections (images of the particles) can then be separated into subgroups or classes that contain particles of the same orientation. The signal to noise ratio of the data can then be increased by averaging the members of each class (8). Finally, the relative angular orientation of each class can be determined and these class averages used to calculate the 3D map of the molecule. Two techniques exist for single particle reconstruction, the random conical tilt method (20) and the angular reconstitution method (21), the former of which requires a single stage tilt in the electron microscope to be performed. Another advantage of single particles is the ability to extract structural information from large, otherwise un-crystallizable complexes. A prime example is the ribosome that until recently defied crystallization. However, single particle EM has provided and is still providing information of the process of translation (22-24) and in a recent publication the ribosome was captured in co-translational

translocation in complex with SecYEG (25) (Figure 1.2a). Another example of a macromolecular complex solved by single particle analysis that would not have been amenable to crystallization is the multimeric assembly of the membrane inserted pneumolysin. Using cholesterol rich liposomes the authors were able to visualize and obtain a 3D reconstruction of the membrane bound and membrane inserted forms of this pore forming toxin including part of the membrane, thus enabling visualization of the membrane deformation around the pore (26) (Figure 1.2b).

Due to noise present in the data, a limitation of single particle EM is the ability to correctly align and classify the molecules. For this reason smaller particles (<200 kDa) are more difficult to align precicely. However examples of small macromolecules being studied by single particle methods are becoming more common due to advances in electron optics that increase the contrast and computational methods that increase alignment efficiency (27-29).

The resolution obtained by single particle EM is generally lower than that obtained from regular arrays and depends on a variety of factors. The symmetry of the molecule is especially important since having redundancy in subunit arrangement reduces the amount of data required and will allow for a higher signal to noise ratio to be obtained. This is why highly symmetric particles such as icosahedral viruses (30) and chaperones with dihedral symmetry (31) have been studied to better than 10 Å resolution. Another factor influencing resolution is the homogeneity of the molecules. The reconstruction process generally assumes that each projection is derived from the

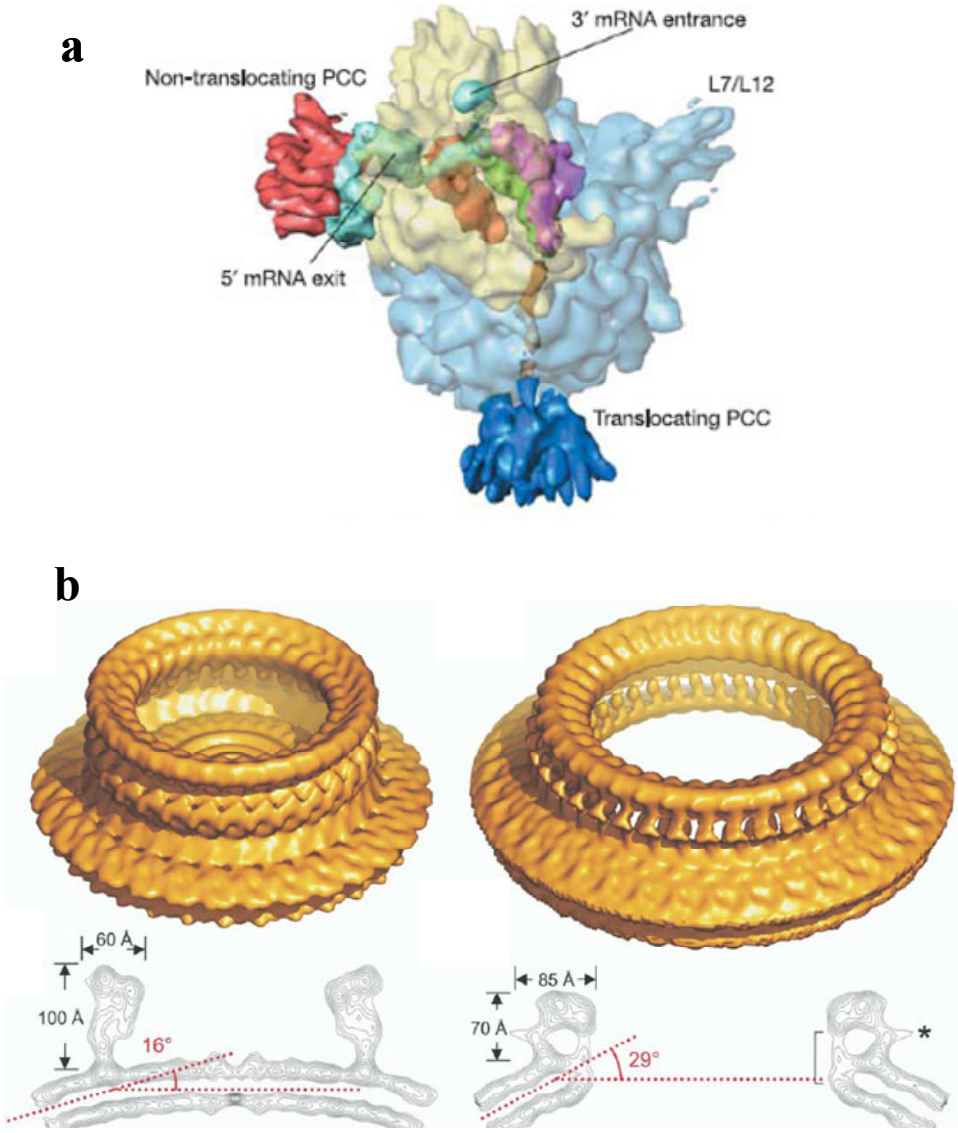

Figure 1.2 Macromolecular machines solved by single particle EM
(a) The structure of the 70s ribosome in translational arrest during translocation of a nascent polypeptide through the SecYEG protein conducting channel (PPC) from Mitra *et al*. 2005. (b) Structures of prepore (left) and pore (right) forms of the cholesterol dependent cytolysin pneumolysin captured along with the lipid bilayer (Tilley *et al.* 2005)

same molecule but unlike in a crystalline lattice that forces a certain conformation to the molecules, single molecules are not restricted in any way. Ligand binding and other co-factors that may be present in the protein preparation may induce conformational changes that will affect the alignment and subsequent reconstruction. This is why samples have to be biochemically homogeneous, and if possible, exist in a stable conformation. An example is ATP binding and hydrolysis by GroEL. It was found that large domain movements of GroEL occur upon ATP binding but these could not be captured by crystallography or EM due to hydrolysis of ATP and could not be propagated by the use of non-hydrolysable analogues. A single point mutation abolishes ATP hydrolysis and the ATP bound structure was solved by cryo-EM to a resolution of 9 Å (32).

Although heterogeneity in a protein population is generally undesirable it cannot always be avoided, especially in large multiprotein complexes. Single particle EM offers an advantage in such cases as computational methods for dealing with more than one conformational population are already in use (26,33) and are discussed in more detail below. Heterogeneity also arises from specimen preparation and imaging. Negative staining can give rise to undesirable flattening or distortion effects but can also contribute to heterogeneity be means of uneven staining. Radiation damage of the molecules also has a large impact factor especially when approaching the subnanometer barrier. Low dose electron microscopy (< 10e-/$Å^2$) is now routinely carried out to minimize radiation damage but as a consequence a lower contrast results and thus a larger number of particles are required (34).

As mentioned previously, a 3D reconstruction requires a certain amount of molecules to achieve a target resolution. This number has been calculated in several publications and was found to be dependent on the symmetry, size and homogeneity of the protein. A 10 Å reconstruction of the asymmetric 70S ribosome requires ~ 30,000 particles but it is estimated that to achieve atomic resolution (~3Å) it will require a dataset in the order of 10^6 particles (19). For highly symmetric proteins like GroEL, a 6 Å structure has been achieved from just under 40,000 particles and represents the current highest resolution structure obtained by single particle EM (31). The collection of large datasets has prompted the development of high-throughput automation procedures and software for collecting, extracting and processing large datasets. Using automated image capturing software that controls the specimen stage and CCD camera, the group of Bridget Carragher could collect 280,000 particles of GroEL over a period of 25 hours with minimal human intervention and reconstruct a 3D map to better than 8Å (35). Such technologies will be required, if single particle EM is to routinely achieve subnanometer resolutions.

Cryo-electron tomography

Tomography refers to the assembly of images from sections (Greek=tomi) of a 3D object to reconstruct that 3D object (36). Although tomography of whole cells and organelles has been primarily from images of thin-sections where the specimens have been chemically fixed, embedded in resins and sectioned physically on microtomes, the preservation of biological material limited the resolution of the reconstruction. In recent

years developments in specimen preparation and electron optics have allowed the field of cryo-electron tomography (cryo-ET) to expand and produce 3D maps of whole cells at nanometer resolution without physical sectioning (5,36). By rapidly freezing the specimen (with a thickness up to 10μm) in liquid ethane, the same procedure used for sample preparation in cryo-EM, the specimen is kept in its hydrated environment in amorphous ice. Specimens are then imaged in a TEM equipped with a cryo-stage that maintains liquid nitrogen temperatures throughout the imaging process. The TEM has depth of field of several hundred nanometers and thus a 2D image of an unstained specimen represents the densities along the Z-axis (parallel to the electron beam). Thus by tilting the specimen and collecting images at as many angles as possible, the resulting 2D projections can be back-projected to produce a 3D map of the specimen. The same principle applies to all EM reconstruction techniques including electron crystallography and single particles. Because each specimen is unique and pleomorphic, unlike when imaging single particles of the same molecule, contrast enhancement by averaging is not possible. Thus the maximum contrast has to be extracted during the imaging process. The large thickness of cryo-ET samples (usually >200 nm) results in many inelastic scattering events that give rise to image blurring (7). The use of energy filters that prevent these electrons from contributing to the image makes it possible to achieve maximum contrast. The use of a field emission gun (FEG) that produces a more coherent beam allows information to high resolution to be transferred with minimal attenuation. Both of the above microscope properties make it possible to perform cryo-ET at high resolution. Once a 3D map or tomogram is obtained, one can segment the densities to

study the part of the map that is of interest. Large protein complexes (>500 kDa), such as ribosomes, chaperones and proteasomes, can be visualized directly and enhanced by extracting and averaging their densities. Efforts are being made to create a macromolecular library of known structures that can be used to probe the location of these macromolecules within the tomogram (5). Some examples of structural studies using cryo-ET include visualization of the chemotaxis receptor arrays in the inner membrane of *E. coli* (37), microtubules in mammalian cells (38) and eukaryotic flagella (39). These examples indicate the broad application spectrum of cryo-ET in the structural study of macromolecules in their native environments.

Single particle reconstruction methodology and principles

Once suitable quality micrographs have been collected for the macromolecule of interest, either by negative stain or cryo-EM, particles are extracted from the micrographs either manually or semi-automatically. In the latter case the user selects 10-20 projections as references to be used by the software for recognition of protein particles in the micrographs (9). The accuracy of automated selection sometimes depends on the contrast and shape of the molecules. At this point, depending on the resolution required, a correction or compensation for the contrast transfer function (CTF) has to be made. For target resolutions beyond typically 15 Å, a CTF correction is usually performed on either the micrographs or the raw particles. In the single particle analysis software EMAN (9) the raw particles from a single micrograph are loaded and an FFT of the average is generated. The positions of the zeroes can then be used to determine the

defocus value used to capture the image. A series of steps is followed to match the experimental power spectrum curve to a theoretical curve whose shape will be determined by the CTF parameters (defocus, CTF envelope function, acceleration voltage and spherical aberration). Once a fit is found, the phases of the CTF are flipped to positive values thus correcting the contrast reversal caused by the CTF. If the target resolution lies beyond the position of the first zero, one can band-pass filter the data to ignore spatial frequencies affected by the CTF (7). In practice, when CTF correction is performed, micrographs taken at different defocus values are used so the information at the zeroes of one micrograph is compensated by the information from another micrograph. When the CTF is dealt with by band-pass filtering, micrographs from the same defocus are used so that the contrast features are similar. Once CTF correction is performed the particles from different micrographs can be merged and processed as a whole.

Particles are then normalized by scaling the pixel values so that differences in greylevels and background amongst the projections do not interfere with the reconstruction process. Following normalization, the particles are then centered inside their respective boxes and then subjected to classification. Different algorithms exist for separating the particles into classes. One approach is multivariate statistical analysis (MSA) (40,41). In this procedure the data, which basically consist of a number of pixels of different densities, is compressed into eigenvectors. Variations in the pixels that correspond to variation in the images give rise to different eigenvectors whose analysis is used to cluster images into groups with characteristic features. The products of

eigenvector analyses are eigenimages which describe the density distribution in the particle population (Figure 1.3a). Eigenimage inspection can reveal information of the molecule's symmetry and if heterogeneity in size exists in the population as indicated by concentric circles in the eigenimage (42). Eigenvector selection and eigenimage inspection were used in this study to separate S105 rings of various diameters (Chapter II). Once the particles are classified the members of each class are averaged to increase the signal to noise ratio. At this point some programs will improve the class averages by realigning the raw particles to the class averages in a process called multireference alignment (MRA) (40).

The next step is a calculation of the Euler angles of each class. Three angles define the orientation of a 2D projection in 3D space, namely alpha, beta and gamma. Given that a particle is intersected by three axes X, Y, Z then gamma corresponds to a rotation around the Z-axis. Beta refers to a rotation around the new Y axis and alpha is a final in plane rotation around the new Z-axis. Determination of the Euler angles of each class is based on the common lines theorem that states that any two 2D transforms of projections of the same 3D object must share a common line at their intersection (43) (Figure 1.3b). To find this common line, brute force cross-correlation between every one-dimensional line in the 2D transforms of each projection is carried out. Once a common line is found the two transforms and hence the projections are fixed in 3D space and allowed only to rotate around the central line or section. A third transform or projection is added and once its common line is found with respect to either of the first

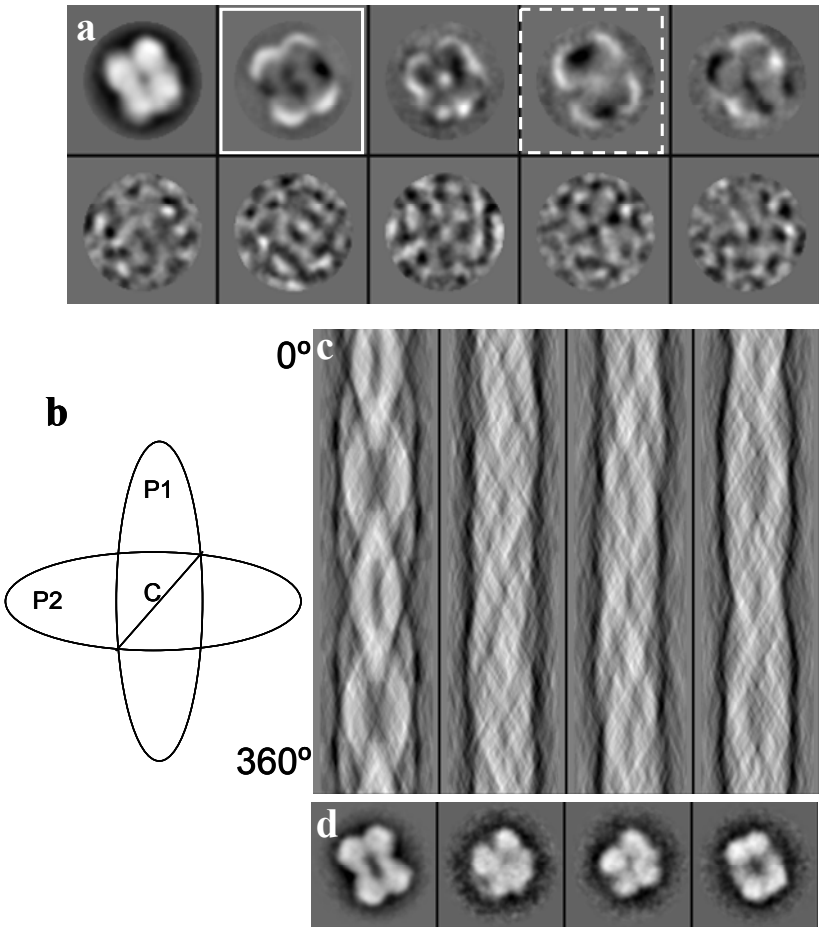

Figure 1.3 MSA classification and angular reconstitution
Eigneimages resulting from MSA and eigenvector analysis (a) in ascending order of significance (top to bottom, left to right). The primary density differences (white box) between projections of SecA (Chapter III) are shown and subsequent images describe smaller variations in the population. White dashed box indicates the 2-fold symmetry present in SecA. The common lines theorem (b). The two projections P1, P2 share a common line across their intersection. Examples of sinograms (c) generated by the corresponding SecA class averages (d). Each sinogram is a sum of 1D densities of the projection across 360º. Correlation between every line (horizontally) of each sinogram will find the common line between the corresponding projections. Note the periodic pattern of the first sinogram due to the strong 2-fold symmetry of the first projection.
two, the relative orientations of all three projections are known. Addition of

projections is done by comparing the new projections to pre-determined pairs (8). In real space this is achieved by converting projections to 1D lines that are the sum of the densities across one direction. In plane rotations of the projections from 0 to 360º will produce 360 lines that are used to compose a sinogram or a sum of 1D lines across every direction (21).

Individual sinograms from different projections are then compared to each other across each line (sinogram correlation) and the correlation maxima indicate a common line (Figure 1.3c, d). Once the Euler angles of all projections (in this case class averages) are determined they are combined to produce a 3D density map composed of all the projections in 3D space. The exact filtered back projection algorithm, which is commonly used in medical imaging, is used to create the 3D map (44). This method removes very low frequency information from the projections i.e. filtering, to reduce blurring due to an overlap at the central section of the projections.

The 3D map initially obtained may be noisy and inaccurate and must be refined. Different programs employ different refinement procedures. Another single particle analysis package, IMAGIC (40), reprojects the 3D map in 3D space to create 2D reprojections that can be used to translationally and rotationally re-align the raw particles to the reprojection that matches them best. This in turn improves the subsequent re-classification step as each particle is now rotationally invariant in 2D space and only its orientation is different with respect to other particles. Thus, more particles can be placed in the same class increasing the signal to noise ratio. EMAN (9) employs a similar approach to IMAGIC by creating a fixed number of class averages based on the number

of reprojections used (determined by the angular increment of reprojection) and assigning each raw particle to the class that matches it best. The convergence of the reconstruction after each refinement step, i.e. how well each new model agrees with the previous one, can be followed by observing the similarity of a 3D model to the next by a Fourier shell correlation (FSC) (8) and provides an indication of when no more improvement is seen (9). Finally the resolution of the 3D map is determined by dividing the data-set into two halves and creating two "independent" reconstructions. In reality the two reconstructions are not entirely independent since the Euler angles for each projection or class have been predetermined and are the same in the two halves. The FSC between these two maps measures the correlation of their 3D transforms and is essentially measuring the signal to noise ratio of the 3D reconstruction across different spatial frequencies or resolutions. At low resolution the two objects are very similar and have a high correlation. At higher spatial frequencies their correlation decreases and eventually falls to near zero. The correlation cut-off is normally set at 0.5 which means a signal to noise ratio of 1 (19). An alternative to the 0.5 criterion is the 3σ (three standard deviations over the random noise value) curve that is obtained by multiplying the correlation expected for noise by three (45). Using this criterion leads to a resolution estimation slightly higher than that of the 0.5 cutoff and some groups argue that the 0.5 criterion is more conservative or safe (19). In any case the FSC gives us a resolution cutoff of the reliable information within the 3D map so as not to over-interpret the data. This is not to say that higher resolution information is not present in the structure, but such information should be interpreted with caution.

During the reconstruction a choice of symmetry has to be made to impose to the 3D structure. Rotational symmetry (Cn) refers to a single axis of rotational symmetry. Dn symmetry refres to two axes of symmetry: one n-fold axis of rotational symmetry and a 2-fols axis passing perpendicular to the n-fols axis. Particle images obtained from electron micrographs can display n-fold rotational symmetry across one orientation. Determination of n can be achieved by rotational filtering. This method refers to the technique by which a particle displaying n-fold symmetry is rotated n times by $2\pi/n$ and an average of the rotated images is created. During the averaging, noise which does not display rotational symmetry will be reduced by the square root of n (7).

Analysis of the accuracy of n selection can be determined visually or using rotational power spectra. In the first case, the rotational average is inspected for the symmetry which was imposed. If averaging reduced structural features that were present in the raw particles or unsymmetrized class averages, then it is likely that the wrong symmetry was used. If an enhancement of structural features as well as preservation of symmetry is observed then the choice of n is likely to be correct.

Rotational power spectra display the angular harmonics of a particle over a range of n symmetries. The images have to be converted from Cartesian coordinates (x,y) to polar coordinates (r,φ) and the center of the rotational symmetry axis has to be chosen accurately. In these coordinates the particle will exhibit periodicity in the angle φ. If a particular symmetry m is present then the power of a harmonic will be maximum at n= m, 2m, 3m and so on (7).

Symmetry estimation can be challenging depending on how strong the rotational

symmetry is displayed, the contrast of the particle and the general preservation of the particle. Multimers with high rotational symmetry and small subunits can be more challenging than those will lower symmetry but large subunits as the spatial distance between subunits might reside beyond the resolution of the micrograph. Furthermore this distance might be close to a particular spatial frequency that is attenuated by the choice of defocus used to acquire the micrograph. When possible, biochemical data must be used to complement structural studies for the correct determination of symmetry.

1.2 INTRODUCTION TO THE BACTERIOPHAGE LAMBDA HOLIN

1.2.1 Mechanisms of phage release from the host cell

Bacteriophage infections end with the release of phage progeny from the host cell into the immediate environment, thus beginning a new cycle of infection. In order to achieve this, the host cell envelope has to be compromised for it to allow the large (25-200 nm) phage particles to exit. In the case of the filamentous phages, e.g. M13, the virions extrude from the host membrane while simultaneously acquiring their protein coat from the host membrane. This exit strategy does not result in host cell lysis or death (46). Some small, single-stranded DNA or RNA phages achieve host lysis by interfering with the cell envelope synthesis pathway. The result is similar to the action of antibiotics like penicillin that inhibit cell wall synthesis. The dividing cell cannot synthesize the cell wall, leading to the creation of gaps. Eventually the osmotic pressure causes cell lysis and release of the phage progeny. The single-stranded RNA bacteriophage Qβ

maturation protein A2, prevents cell wall synthesis by inhibiting MurA, which catalyzes the first commited step in murein synthesis from the sugar UDP-GlcNAc (47). The small microvirus φX174, inhibits a later step of cell wall biosynthesis (48). In either case, a single phage-encoded protein acts to cause host cell lysis and release of phage progeny.

Double-stranded DNA bacteriophages, including lambda and phage 21, achieve host lysis and release of phage progeny by employing a holin-endolysin system (49). This two-component system relies on a phage encoded, cytoplasmically localized endolysin, which gains access to its substrate via a "hole" or lesion that is formed in the inner membrane by the second component, the holin.

1.2.2 The lambda lysis genes

The lysis genes of λ are transcribed by the late promoter p'R that is activated approximately 8 minutes after infection (49). The lysis cassette encodes the holin (*S*), the endolysin (*R*) and two more proteins involved in host lysis, *Rz* and *Rz1* (Figure 1.4 a). Under standard laboratory conditions *Rz* and *Rz1* (an inner membrane protein and an outer membrane lipoprotein, J. Berry and R.Young, unpublished data, (50,51)) are not required for lysis unless millimolar amounts of divalent cations are present in the media (52). Thus, the holin and endolysin are the only two components required for saltatory lysis under standard conditions and expression of the lysis cassette from a thermally induced lysogen leads to a characteristic, abrupt decline in the optical density of the culture at ~50 minutes post induction.

1.2.3 Endolysins

Although expression of the holin leads to cessation of growth, ultimately the rapid lysis occurs due to the degradation by the endolysin of the cell wall that acts as an "exoskeleton", maintaining the shape of the cell and preventing osmotic shock. The activities of endolysins on the murein cell wall can be divided into five catalytic groups (53). Transglycosylases (i) and glycosidases (ii) attack the β-1,4-glycosidic bonds between N-acetylmuramic acid (MurNAc) and N-acetylglucosamine (GlcNAc) sugar units that compose the glycan part of the peptidoglycan cell wall. Glucosamidases (iii) and amidases (iv) hydrolyse the linkages within the pentapeptides, and endopeptidases (v) hydrolyze the inter-peptide links between adjacent pentapeptides attached to the neighbouring glycan strand. In the case of lambda, the endolysin R is a small 18 kDa muralytic transglycosylase. The crystal structure of λR reveales that it has a globular shape with dimensions of 40 x 32 x 32 Å (Figure 1.4 d) (54). The overall structure resembles other lysozymes in that two domains sandwich the active site and are connected by a long alpha-helix. The active site was found to contain Glu19 which was previously implicated as an essential catalytic residue (55).

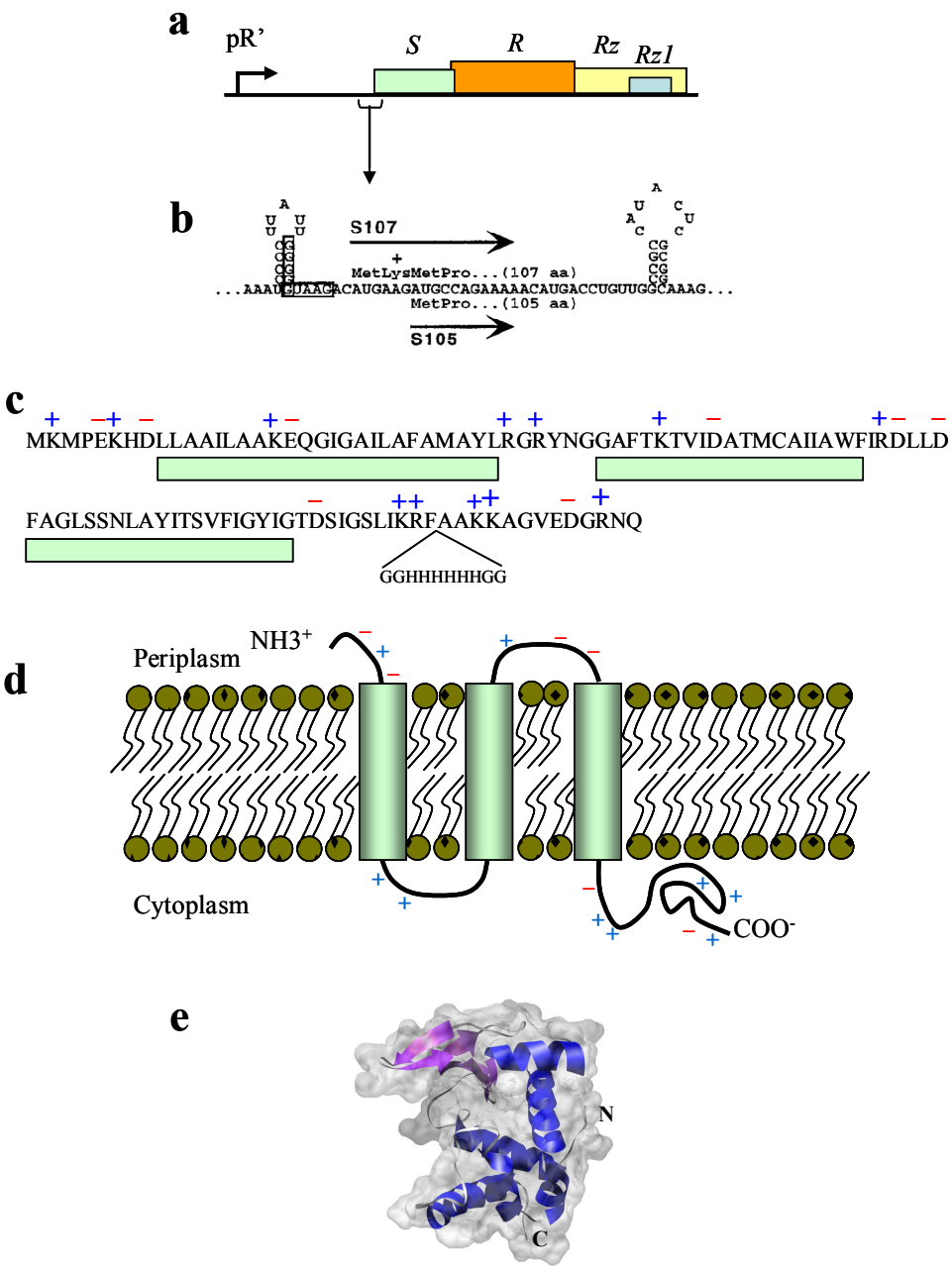

Figure 1.4 Lambda lysis genes and gene products
The lambda lysis cassette (a). Structure directed initiation (SDI) motif of S105 and S107 (b). Protein sequence of S (c). Topology diagram of S105 (d). Crystal structure of Lambda R endolysin (PDB ID 1AM7).

1.2.4 Holins

Holins are small, phage-encoded, integral membrane proteins that create a "hole" or lesion in the inner membrane of the host cell and can be divided into two main classes. Class I holins, which include the prototype lambda S105 or S holin, are generally larger than 90 amino acids and have three transmembrane domains (TMDs), while Class II holins, e.g. the $S^{21}68$ holin of phage 21, are smaller and contain only enough hydrophobic stretches to comprise two TMDs (49).

Apart from disrupting the host membrane, holins play an important role in lysis timing. Under standard laboratory conditions, the holin gene alone controls the onset of lysis (56). Holin triggering (lesion formation) coincides with the loss of the membrane potential of the cell and leads to growth arrest. Furthermore, premature triggering of the holin can be induced by addition of a respiratory poison such as the proton un-coupler 2,4-dinitrophenol (57). Lysis timing in the context of phage evolution is critical. A delay in lysis ensures that the host cell is not killed prior to the complete assembly of phage progeny as virion production would stop. However, late lysis would also confer an evolutionary disadvantage if the environment was populated with a large number of available hosts, as despite a larger burst size, the competition for new host cells would be compromised.

1.2.5 Localization and topology of the lambda holin S

Although it was earlier shown that S had a permeability effect on membranes (58,59), the localization of S was shown by isopynic sucrose gradients to be almost exclusively

in the inner membrane fraction (60). The topology of S has been experimentally determined by gene fusion, protease accessibility and other biochemical studies and is shown in Figure 1.4 d. The C terminus of S was shown to be cytoplasmicly localized by accessibility to proteinase K in inverted membrane vesicles (IMVs), but not spheroplasts (61). Graschopf *et al.* (1999) showed that when the N terminus of *P. aeruginosa* phage Pf3 coat protein was fused to the N terminus of S, the resulting Pf3ΦS105 product caused lysis of induced cultures in a signal peptidase-dependent manner (62). Protease shaving experiments on spheroplasts showed the loss of the α-Pf3 antibody epitope indicating the N terminus was residing on the periplasmic side. Later Gründling *et al.* (2000) showed that lysis by this fusion was *S* allele specific and not a result of a generalized insult to the membrane (63). The primary sequence of S indicates three hydrophobic stretches that were thought to comprise the three TMDs. Using the cysteine-specific, membrane-impermeable modification agent iodoacetamide (IASD), Gründling *et al.* (2000) showed that continuous residues within these three domains were not accessible to modification, where as the intervening sequences as well as the N terminus were easily labeled (63). This provided robust evidence for the three TMDs in S and verified the N "out" C "in" topology. More recently it has also been shown that purified S protein cannot be sequenced by N-terminal sequencing unless it is deformylated, indicating again the N terminus of S immediately localizes to the periplasmic side immediately after its synthesis, thereby avoiding deformylation by the cytoplasmic deformylase (64).

1.2.6 S encodes a holin and a holin inhibitor

A common paradigm found in many lambdoid holin sequences is the presence of two translation initiation sites. In *S* the two AUG codons at +1 and +7 give rise to two gene products, S107, and S105, such that S105 is shorter by a Met-Lys. Translation from the Met1 or Met3 codons is controlled by two stem loop structures in the mRNA, one ~7 nt. upstreamstream of the first start codon and another at ~ +30 (65) (Figure 1.4 b, c). The two polypeptides are produced *in vivo* at a ratio of 2-S105:1-S107 (66) and have been shown to interact directly forming heterodimers (67). S107 acts as an inhibitor of S105 and delays lysis, compared to expression of S105 alone. The extra positive charge at the N terminus of S107 is thought to prevent the periplasmic localization of its N terminus due to the membrane potential across the membrane (68). However, upon membrane depolarization the N terminus of S107 is no longer restricted to the cytoplasmic side and flips across the membrane, thereby converting S107 into a topology and function equivalent to S105. The presence of a holin inhibitor is thought to act as an additional regulation mechanism for fine tuning the timing of lysis. Further more the sequestering of S105 in S105:S107 heterodimers prior to triggering results in an instant availability of up to three times more S that is capable of hole formation once the membrane potential has collapsed. This sudden increase in active S also gives rise to the saltatory nature of lysis (67).

1.2.7 Structure of the S mediated lesion

The exact nature of the “hole” that is thought to be formed by S105 still remains

uncertain. Gründling *et al*. (2000) showed that S could be efficiently crosslinked in the inner membrane using the membrane permeable crosslinker dithiobis-succinimidyl-propionate (DSP) (63). Crosslinking provided evidence for the capacity of S to form higher oligomers that could be extracted with the detergent Triton X-100. Conversely, the majority of mutations in the three TMDs of S that abolished holin function were also shown to prevent oligomerization demonstrating the importance of a multimeric assembly for hole formation. One such mutation, A52V, in the second transmembrane domain, was shown to block S at the dimer stage. In contrast, the missence change R59C retained the ability to oligomerize despite being unable to trigger even after the collapse of the membrane potential by respiratory poisons.

Wang *et al.* (2003) showed that the lesions formed by the S holin were large enough to allow the passage of an R-β galactosidase fusion to the cell wall (69). In order to allow this 480 kDa tetrameric assembly exit, the hole would have to be larger than 12 nm in diameter and constitute the largest known membrane pore formed by an α-helical inner membrane protein. In light of these findings, it was reasoned that the lesion formed by S might not be a well delineated hole but rather a two-dimensional proteinaceous raft that upon triggering would form a non-specific lesion in the host membrane. The raft model is also supported by studies which showed that mutations in all three TMDs of S affect lysis timing, suggesting that all three TMDs are involved in protein-protein interactions as would be expected in a protein raft. This model also predicted that S mutants defective in oligomerization would fail to assemble into rafts and mutants blocked after oligomerization, were defective in propagating the wave of depolarization

accross the raft. Interestingly in the same study it was found that the *RφlacZ* gene product could not escape if S105 was triggered at 20 minutes by addition of respiratory poisons. Conversely an early lysis mutant, $S105_{a52g}$, that triggers ~20 minutes post-induction allowed passage of the chimeric endolysin. The above results suggest that the lesion created by premature triggering of S would be smaller than the mature self-triggering one and that $S105_{a52g}$ can assemble into a mature, triggerable complex, at lower concentrations than those required for wild-type S.

The lack of intermediate or high resolution structural information for the holin family has hindered a complete molecular understanding of lesion formation. A breakthrough in the structural determination pathway of S came with the successful over-expression of the holin in a T7-based system to produce the milligram amounts of protein required for biochemical and structural characterization (70). Initial attempts to place an oligohistidine tag (τ) at the N or C-terminus of S resulted in either a lysis defective phenotype or one in which early triggering was accompanied by a slow release of R, for the N and C tags respectively (71). After an exhaustive search for a site in which to insert a hexahistidine tag, that would aid purification without altering the phenotype of *S*, Smith *et al.* (1998) found that when the sequence GGH_6GG was placed internally at positions 88 or 94, the tagged S gene supported lysis with timing similar to un-tagged *S* (71). Furthermore the phenotypes of $S105\tau94_{a52g}$ and $S105\tau94_{a52v}$ were similar to the untagged alleles. Because *S105τ88* accumulated reduced amounts of S105 the construct *S105τ94* was cloned in pET11a and used in later biochemical characterizations of S105, including this study. S105τ94 was initially purified in the

non-ionic detergent n-Octyl-β-D-Glucopyranoside (OG) (72). When purified protein was subjected to circular dichroism spectroscopy, approximately 40% of the protein or 46 residues corresponded to α-helix, consistent with S having two TMDs (72). However, S105 was particularly unstable in OG and was later exchanged for 6M guanidine-HCl, a chaotropic salt that usually unfolds proteins. However, even in this chaotrope S105 retained the same alpha-helical content as in OG. Using the guanidine-HCl solubilized S, Smith *et al.* also performed *in vitro* hole forming assays on liposomes containing a self-quenching fluorescent dye. Upon addition of S105t94 to the liposomes, dye release was observed indicating a hole or lesion was being formed in the bilayer. In contrast, under the same conditions, no dye was released when $S105_{a52v}$ was added to the liposomes.

Deaton *et al.* (2004) further improved the solubilization and purification of S105 and found that of many detergents screened, the zwitterionic detergent Empigen BB (EBB) could very efficiently solubilize and keep S105 stable once purified. Furthermore, he showed that stably solubilized S displayed enough alpha-helix for 3 TMDs or about 60 residues as determined by CD (64).

Further *in vitro* characterization of S105 mediated lesions were performed using the GroEL chaperone as a substitute for the detergent. Deaton *et al.* (2004) showed that in the presence of GroEL, S105 could be efficiently solubilized and delivered to liposomes in the absence of detergents (73). Furthermore the release of a self-quenching fluorescent dye from the liposomes was *S* allele specific. $S105_{a52v}$ could not effect dye release and the temperature sensitive $S105_{a55t}$ allowed release only at the permissive temperature. Moreover it was found that GroEL was capable of keeping S105 soluble at

ratios of up to 800 S105 molecules to one GroEL tetradecamer or about 100 times more than the capacity of the GroEL cavity allows. In contrast, S107 loading could not exceed 7 molecules per GroEL molecule. Electron microscopy showed that S105 was forming protein shells around GroEL (73) and single particle analysis combined with 3D reconstructions indicated protein densities inside only one of the two GroEL cavities, consistent with GroEL's asymmetric loading of unfolded proteins (74).

1.3 INTRODUCTION TO THE SEC TRANSLOCATION SYSTEM

1.3.1 The Sec system of protein translocation

In all organisms the transport of materials across membranes is essential. Trafficking of molecules and macromolecules across the plasma membrane or across compartmentalized organelles ensures that essential nutrients and gene products arrive at, or are targeted to their final destination (75). The transport of proteins across membranes, or protein translocation, is a universal process. In bacteria, the Sec system constitutes the generalized pathway for secretion of newly synthesized proteins across the plasma membrane and also the path for integration of some membrane proteins into the lipid bilayer (75) (Figure 1.5a). This system has been studied extensively in the Gram-negative bacterium *Escherichia coli* and to a lesser degree in Gram-positive bacteria and archaea.

The core components of the Sec system are the cytoplasmic ATPase SecA and the membrane embedded heterotrimeric complex SecYEG, which forms the channel for

translocation of pre-proteins. SecA mechanically drives translocation by ATP hydrolysis and is present in all bacteria but absent in the archaea (76). The SecYEG complex, besides being present in eubacteria, has homologues both in the archaea (the SecY complex) (77) and in eukaryotes (the Sec61 complex) (78) with the latter two being more closely related to each other. SecYEG along with a membrane embedded protein called YidC is the point of integration of a subset of membrane proteins into the lipid bilayer (79). Additionally, YidC is required for the Sec-independent membrane insertion

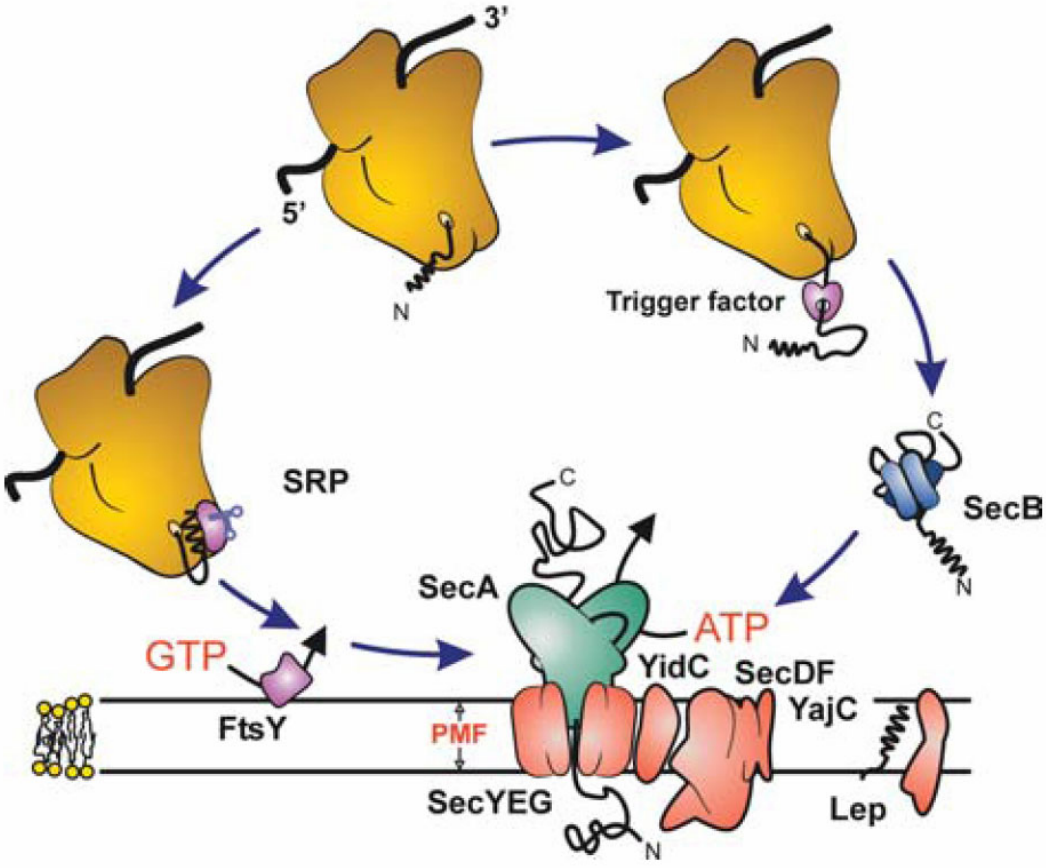

Figure 1.5 Schematic representation of protein translocation in *E. coli*
The post-translational SecA dependent secretion and co-translational SRP dependent membrane integration pathways. Taken from Keyzer *et al.* (2003).

of certain membrane proteins including the filamentous bacteriophage M13 procoat and Pf3 coat proteins (80). Also present in Gram-negative bacteria is a chaperone called SecB. SecB interacts with both the pre-protein and SecA, thereby preventing the pre-protein from folding prematurely, prior to secretion (75).

Pre- proteins targeted for secretion through the Sec pathway have an N- terminal signal sequence that is recognized by the SRP (Signal Recognition Particle) (81), SecB and SecA (75). This sequence is later removed by leader peptidase as the N-terminus of the pre-protein emerges from the non-cytoplasmic side of the membrane. SecA undergoes cycles of ATP hydrolysis, mechanically coupling ATP hydrolysis to insertion of the pre-protein through the SecYEG channel (75).

The components of the Sec system were identified by genetic methods in the early 1980's. Emr *et al.* (1981) isolated suppressor mutants that allowed the outer membrane λ protein LamB with a defective signal sequence to be exported. These suppressors were mapped to three loci called *prlA*, *prlB* and *prlC* (prl for protein localization) (82). *prlA* was later mapped downstream from the L17 ribosomal subunit *rplQ*, and the gene was named *secY* (83). In the same year Oliver *et al.* isolated temperature sensitive *E.coli* mutants that showed pleiotropic defects in the secretion of certain exported proteins (84). They created a *malE-lacZ* fusion, which when induced by maltose caused cell lethality, presumably by blocking the secretion machinery. Mutants isolated with signal-sequence mutations remained cytoplasmic and allowed for growth on maltose and high β-galactosidase activity. Some of the derivatives of the fusion with unaffected signal sequences and elevated β-galactosidase activity were screened at

higher temperatures and mutants defective in the transport of MalE and MBP were isolated. These studies identified a new gene, *secA*, involved in secretion. Using a similar approach, Kumamoto *et al.* discovered *secB* (85), leading to the identification of other genes involved in protein secretion (86-88).

1.3.2 Pre-protein targeting

The first step in protein translocation is the proper targeting of the pre-protein. As a newly synthesized protein emerges from the ribosome, competition for the signal sequence and the destination of the nascent protein begins. Both SecB and the SRP, the latter of which is associated with the ribosome, recognize and bind the ribosome nascent chain (RNC) (75). Another protein called trigger factor, a cis-trans proline isomerase, is proposed to compete with SRP for the signal-sequence and may increase the requirement for SecB binding and targeting (89). Recognition and binding of SecB and SRP to the nascent chain is determined by various factors like the length of the nascent protein and the hydrophobicity of the signal sequence (89). For example, increased hydrophobicity (as in the case of inner membrane proteins) enhances SRP binding and decreases the requirement for SecB targeting. Also, SecB prefers to bind longer nascent chains of >200 residues (90), whereas SRP seems to bind to the signal sequences of shorter nascent proteins (89). Collectively, these factors provide a quality control mechanism for the proper targeting of membrane-destined and secreted proteins. Proper targeting of the pre-protein is followed by the secretion or membrane insertion of the pre-protein.

1.3.3 Pre-protein secretion

In Gram-negative bacteria SecB binds to the pre-protein and subsequently interacts with SecYEG-bound SecA. Interaction between SecB and SecA occurs at the positively charged C-terminus of SecA, which is rich in basic residues and contains 3 cysteines and a histidine that coordinate a zinc ion (91). Once bound to SecB, the pre-protein is transferred to SecA, resulting in an increase in the affinity of SecA for ATP. A large conformational change that accompanies ATP binding results in dissociation of SecB from SecA, allowing the translocation process to begin (75). Hydrolysis of ATP leads to dissociation of the pre-protein from SecA. SecA could either dissociate from the pre-protein: SecYEG complex or interact once again with the pre-protein: SecYEG complex, thereby continuing ATP-dependent translocation (92). It has been shown that SecA is required only for initiation of translocation and the translocation can be further driven by the proton motive force (PMF) (93). In fact increasing the concentration of SecA leads to an ATP-dependency for translocation to occur (93).

1.3.4 Structural studies of the Sec translocon

The SecYEG channel has been studied structurally using electron microscopy and X-ray crystallography. A 9Å electron crystallographic map from 2-dimensional crystals of *E.coli* SecYEG shows the arrangement of the 13 α-helices, although their allocation to a specific TMD cannot be established at this resolution (94). The authors proposed that SecYEG is active as a dimer since a two-fold symmetry axis was present in the unit cell and no obvious channel was visible in a SecYEG monomer. However, this hypothesis

was challenged with the solution of the crystal-structure of an archeal SecY complex at 3.2 Å resolution (77). The structure from *Methanococcus jannaschii* indicates that SecY may assemble as an oligomer but that only one heterotrimeric complex would act as the protein conducting channel. The heterotrimeric complex has an hourglass shape and a central pore with a diameter of 20-25 Å on the cytoplasmic side and tapers to a smaller pore of 5-8Å on the periplasmic side. This pore is lined with hydrophobic (isoleucine, leucine, valine and methionine) residues that may form the seal of the pore, excluding entry of hydrophilic molecules. Although the periplasmic pore is narrow, a small widening could accommodate an α-helix (diameter of ~10 Å) through the pore. The cytoplasmic side of the pore contains a short α-helix proposed to form a plug that can become displaced to allow polypeptide entry. In summary, the crystal structure of the SecY complex provides the first high resolution information about the workings of this channel. The oligomeric state might depend on SecA binding as previous low resolution EM studies indicate that SecYEG assembles into a tetramer in the presence of SecA (95). However the exact oligomeric state of the active channel is yet to be resolved.

Understanding the functional mechanism of the Sec proteins is important since the Sec system is ubiquitous, and at least some components of this system are essential for cell survival. The components of the Sec system in certain intracellular pathogens like *Mycobacterium tuberculosis,* differ from those of the host and this makes the study of the Sec system in mycobacteria particularly interesting.

1.3.5 The Sec pathway in mycobacteria

Mycobacteria possess most of the major components of the Sec system, namely *secA*, *secY*, *secE*, and *secG*. However, the sec genes are not located close to each other and are thus not directly co-expressed in a transcriptional unit. Interestingly, no *secB* homologue is present in the *M. tuberculosis* H37RV genome, and *tb* SecA lacks the C-terminal motif required for SecB binding (96).

SecY, SecE and SecG from *M. tuberculosis* are predicted to have 10, 1 and 3 transmembrane spanning regions respectively with molecular masses of 47, 17 and 8 kDa. Examination of a sequence alignment of *tb* SecY with *M. jannaschii* SecYα shows approximately 20% sequence identity (Figure 1.6). Even with little sequence identity, an obvious structural similarity in the positioning of the TM domain can be observed. The exact positions of the TM helices of the SecYα structure show considerable overlap with the predicted TM positions in *tb* SecY. The position of TM2a in *M. jannaschii,* which corresponds to the plug region in the crystal structure lies outside the predicted TMD2 for *M. tuberculosis.* Although secondary structure predictions suggest that this region in *M. tuberculosis* could form a single helix, it is possible that TM2 in *M. tuberculosis* could be composed of two smaller helices as seen in *M. jannaschii*.

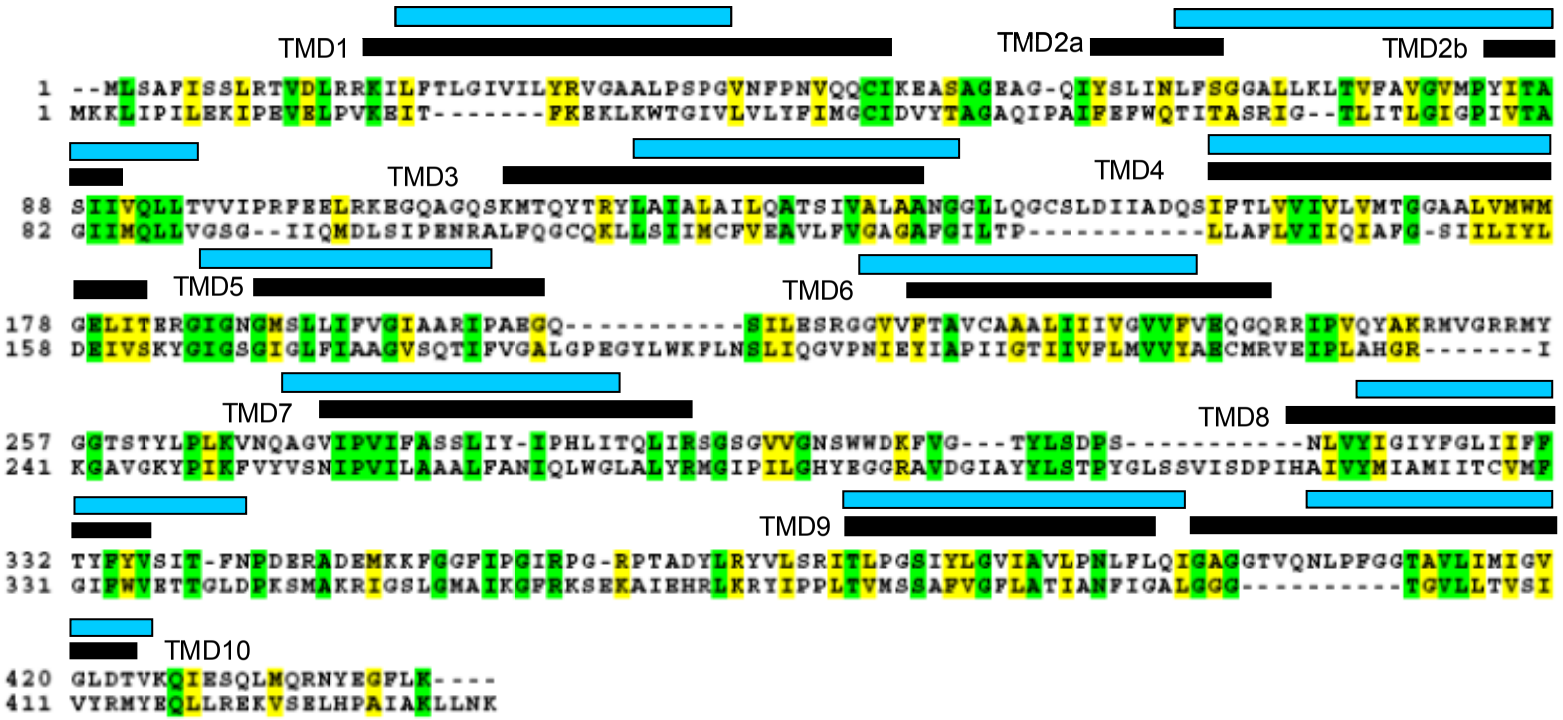

Figure 1.6 Sequence alignment of tb SecY and *M. jannaschii* SecYα

The positions of the predicted TMDs of tb SecY (top sequence) are shown by blue lines and the positions of the TMDs from the crystal structure of Mj SecYα (bottom sequence, PDB ID 1RH5) are shown by black lines. Alignment created by CLUSTALW.

The positions of the residues that line the pore-ring in *M. jannaschii* SecYα correspond to hydrophobic residues in the aligned *tb* SecY sequence. The *M. jannaschii* residues Ile 75, Val 79, Ile 170, Ile 174, Ile 260 and Leu 406 correspond to the *tb* SecY residues Val 80, Ile 86, Met 190, Ile 174 Ile 276 and Ile 414. Since these residues are largely conserved in the two proteins, it is likely that the pore surface will be similar in the two proteins. The periplasmic loop connecting TM6 and TM7 as well as the loop connecting TM8 and TM9 in the *M. jannaschii* structure potentially represent binding sites for partner proteins. The sequence alignment shows a gap of about 7 residues between TMD6 and TMD7 of the *M. jannaschii* sequence. The loop region between TM8 and TM9 also contains two highly conserved arginine residues (R360 and R372 in *M. jannaschii*) that are present in *tb* SecY and implicated in SecA as well as ribosome binding (77).

Interestingly, mycobacteria also carry a second *secA* gene, *secA2,* whose product shares 40% amino acid identity to SecA1 (97). Other pathogenic Gram-positive bacteria have also been found to carry a second *secA* gene. These include *Listeria monocytogenes*, *Bacillus anthracis*, *Staphylococcus aureu*s and *Streptococcus pneumoniae* (98). Braunstein *et al*. showed that *secA2* deletions of *M. smegmatis* are not lethal but produce growth defects on rich media and enhance the sensitivity to azide by two-fold (97). In contrast, a *secA1* deletion *c*ould not be rescued by constitutive expression of *secA2* indicating that these two gene products do not have redundant roles. However the high amino acid sequence similarity between the two proteins and the hypersensitivity of *ΔsecA2* mutants to azide suggest that SecA2 may also function not

only in secretion of a subset of Sec dependent proteins but also a specific group of proteins. Analysis of the secreted products in culture filtrates of *M. tuberculosis* and their comparison in the presence and absence of SecA2 has helped in the identification of a subset of proteins that depend on SecA2 for their secretion (99). Among these were catalase peroxidase (KatG) and superoxide dismutase (SodA). Both these proteins are involved in the protective response that is generated in *Mycobacterium tuberculosis* to counteract the oxidative burst within the macrophage. This response includes the breakdown of reactive superoxide radicals (by SodA) and hydrogen peroxide (by KatG). Interestingly, both proteins lack a classical signal sequence raising the question of how they are targeted for SecA2 dependent secretion.

CHAPTER II
IN VITRO STRUCTURAL AND BIOCHEMICAL STUDIES OF THE BACTERIOPHAGE LAMBDA HOLIN

2.1 INTRODUCTION

Bacteriophage lambda uses a two component system to lyse the host cell. The holin S, multimerizes in the host membrane and at a specific time, creates a lesion that allows the endolysin R to access the cell wall for degradation. The *S* gene produces two products, the holin and the anti-holin, by alternative translation. S105 is a small membrane protein with three TMDs and is the prototype class I holin which adopts an N "out" C "in" topology. The longer gene product, S107 is a holin inhibitor by virtue of an extra positive charge at its N-terminus. This charge prevents the N terminus from localizing to the periplasm while the cell maintains an active membrane potential. Mutations in S can delay, accelerate or abolish lesion formation. The exact nature of this lesion is unclear, but previous studies indicate that S oligomerizes in the membrane and might form a proteinaceous raft that destabilizes the membrane, thereby causing cell lysis. This chapter describes the *in vitro* structural and biochemical characterization of the S holin.

2.2 RESULTS

2.2.1 S105 forms multimers in non-ionic detergent

Previous studies have shown that S forms multimeric complexes *in vivo* that permit it to

disrupt the membrane (63). To study the nature of this multimer, S105 was purified in the zwitterionic detergent Empigen BB using Ni-NTA chromatography and subjected to gel permeation chromatography (Figure 2.1b). In this detergent, S105 eluted as one peak below 50 kDa, possibly corresponding to S105 dimers in micelles, leading to the conclusion that EBB must be disrupting the higher order oligomers that are known to exist in the membrane. In an effort to preserve S105 complexes in a native state, the non-ionic detergent n-dodecyl-β-D-maltopyranoside (DDM) was used to extract and purify S105. When purified protein was subjected to gel-permeation chromatography, three peaks were obtained (Figure 2.1 a). The first peak corresponded to a large > 4×10^7 Da species that eluted in the void volume of the column. The second peak contained > 670 kDa S105 complexes and the last peak corresponded to S dimers. To verify that the larger species in the first two peaks where not the result of the detergent used but rather an intrinsic property of S105, the lysis-defective mutant $S105_{a52v}$, which has been shown previously to be blocked at the dimer stage *in vivo* (63), was extracted and purified in DDM. Subjecting this preparation to gel permeation chromatography yielded a single lower molecular mass peak similar to the one observed with S105 in EBB (Figure 2.1c). Finally a non-lytic mutant of S that lacks the first TMD, $S_{\Delta TM1}$, was also purified and subjected to gel-permeation chromatography. $S_{\Delta TM1}$ elution profiles closely matched those of $S105_{a52v}$ (data not shown).

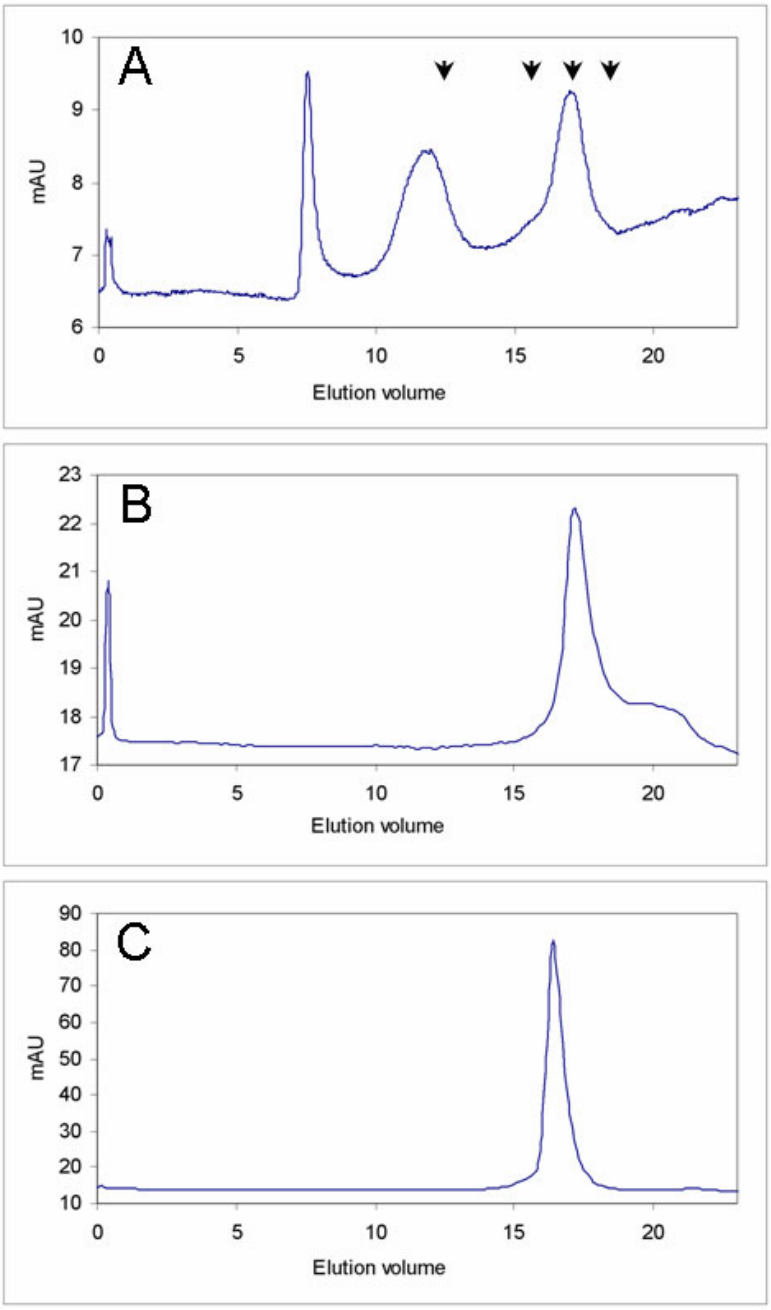

Figure 2.1. Gel permeation chromatography of S
Superose 6 profiles of S105 in DDM (A), S105 in EBB (B) and of $S105_{a52v}$ in DDM (C). Arrows indicate the position of molecular mass standards. Left to right: 670, 158, 44 and 17 kDa. The void volume for this column is approximately 7.5 ml. Samples (100-200 µl of 0.5-1.0 mg/ml) were separated at a flow rate of 0.5 ml/min on an AKTA FPLC.

2.2.2 S forms multimeric rings

To characterize the large oligomers obtained with S105 solubilized in DDM, purified S105 samples were prepared and examined by electron microscopy. Under these conditions, S105 was found to form large, well dispersed ring-shaped structures (Figure 2.2). These rings were also able to stack on top of each other forming filaments of different lengths, with the shortest stacks consisting of just two rings as judged from side-on projections. Fractions from the first peak (void volume) consisted mostly of longer stacks (Figure 2.2a) while the second peak contained shorter stacks and stacks of just two rings, termed ring-dimers (Figure 2.2b). While ring-dimers were very abundant, only a small number of single rings were observed. Also present in this peak were small stacks of up to five rings. This is not surprising since gel permeation separates according to size and therefore the shape of the S105 rings allows for stacks with a height up to the diameter of a ring to be eluted along with ring dimers. S105 was also extracted and purified using the non-ionic detergent n-octyl-β-D-glucopyranoside (OG) and n-decyl-β-D-maltoside (DM). Electron microscopic analysis of OG- purified S105 also revealed rings and ring-stacks but these samples were not as homogeneous as the DDM-purified protein and showed evidence of small aggregates and precipitation (Figure 2.3a). S purified in DM however, showed good dispersion of particles in a manner similar to DDM (Figure 2.3b). Taken together, the above results confirmed that S105 forms large multimeric complexes when purified in a non-ionic detergent.

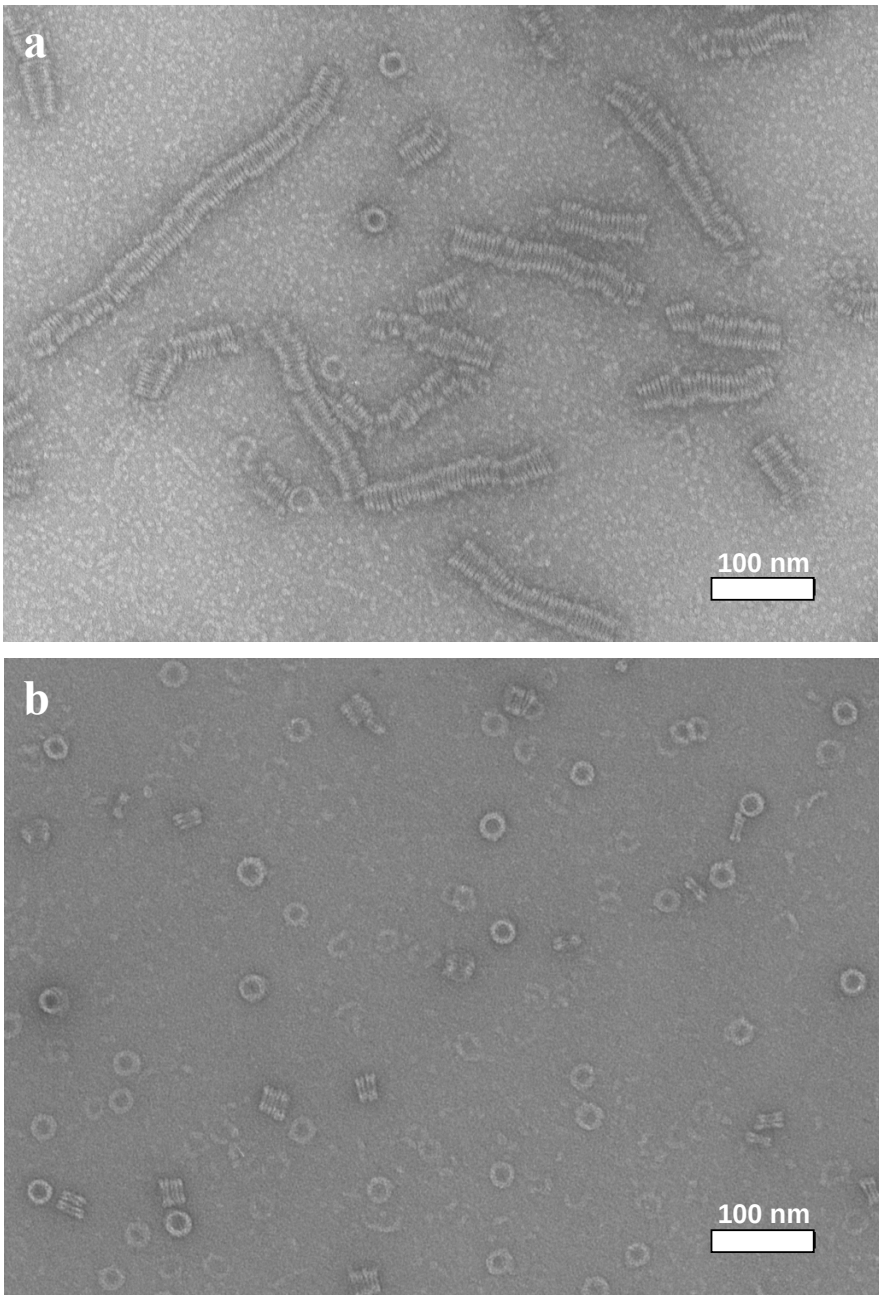

Figure 2.2 Electron microscopy of S105 gel permeation fractions
Fractions from the first and second peaks were pooled and prepared for negative stain EM. Peak 1 (a) consisted primarily of long stacks of rings while peak 2 (b) contained mostly single rings, ring dimers and stacks measuring approximately up to the diameter of a single ring.

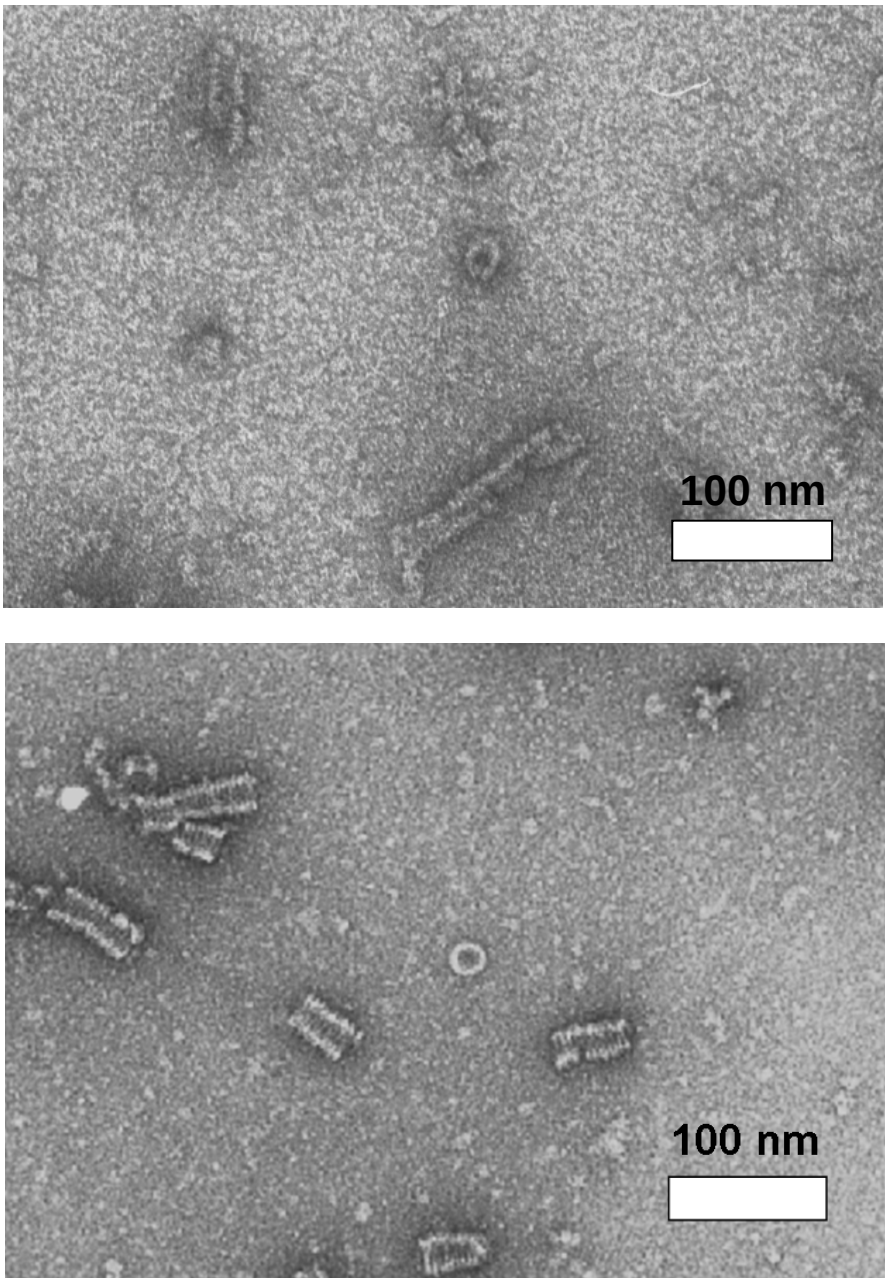

Figure 2.3 S105 purification in other non-ionic detergents
Electron micrographs of S105 extracted and purified in either OG (a) or DM (b). OG-purified samples showed evidence of aggregation as compared to DDM or DM-purified S105.

The *S* mutant *S105$_{r59c}$* has been described previously (63) and was shown to be lysis-defective and untriggerable, even if the membrane has depolarized. However, unlike other lysis defective mutants which were blocked at the dimer stage, S105$_{r59c}$ exhibited oligomerization indistinguishable with S105, when crosslinked with DSP (63). It was therefore reasoned that this variant of S105 was blocked at a later stage of hole formation. To observe the oligomerization properties of this mutant protein, pET30b *S105 $_{r59c}$* was constructed by quickchange site-directed mutagenesis, overexpressed and purified in DDM as before. Interestingly, the mutant protein eluted as a cloudy but homogeneous suspension. Observation of this sample with EM revealed that S105$_{r59c}$ formed rings that looked similar to S105 rings without further image analysis being performed. S105$_{r59c}$ however, displayed increased stacking properties and most stacks were associated with each other along the length of the stack. This gave rise to very electron dense clusters of stacks that were probably several layers thick (Figure 2.4). This property of S105$_{r59c}$ indicates a surface change of the periphery of the rings that allows for the sideways association. The stack associations also explain the cloudy nature of the eluted protein due to light scattering by these very large (>1 μm) assemblies.

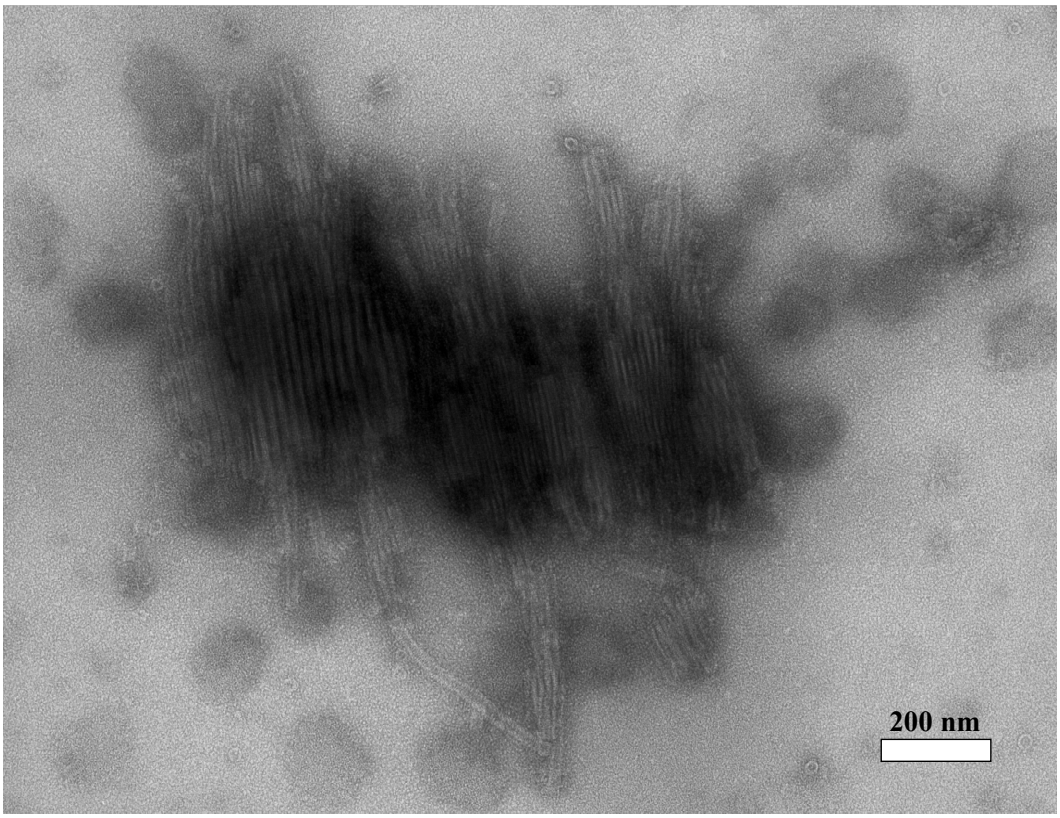

Figure 2.4 Oligomerization properties of $S105_{r59c}$
Electron micrograph of $S105_{r59c}$ purified in 0.1% DDM. The stacking properties of this mutant are clearly evident as well as the sideway association of the stacks that measure > 1μm.

2.2.3 Single particle analysis of S105 rings

To further investigate the structure and possible symmetry of the S105 rings, negatively stained data were subjected to single particle analysis. A visual inspection of the classes and eigenimages indicated a slight heterogeneity in the diameter of the rings (Figure 2.5a). In addition, no subunit separations could be discerned within the ring. This heterogeneity may be due to a variation in the number of subunits that make up a ring. Such variability in subunit number has previously been reported for the cholesterol-dependent cytolysins, that also form multimeric rings of large diameter (26). The data sets were therefore analyzed using an approach similar to the one performed on heterogeneous pneumolysin rings (26,42). Reference-free classification was carried out using only the first 5 eigenvectors in order to allow the data to be separated on the basis of low-resolution features such as particle diameter. Inspection of 15 class averages indicated two clusters based on the outer diameter: 24.5 nm (8 classes, 65% of particles) and 26-28 nm (7 classes, 35% of particles). Particles from a single class within the 24.5 nm-cluster were extracted and re-classified using a larger number of eigenvectors leading to averages showing more structural detail compared to the initial class averages and the eigenimages no longer revealed any evidence of size heterogeneity (Figure 2.5b). The best class average, based on intra-class variance, was chosen as a reference to rotationally align the raw particles followed by classification. The final class average indicated a possible 18-fold rotational symmetry (C18) when inspected vissualy. Imposing several rotational symmetries to this class average (C16 to C20) resulted in an enhancement of features for C18 but not for other symmetries (Figure 2.5c). These

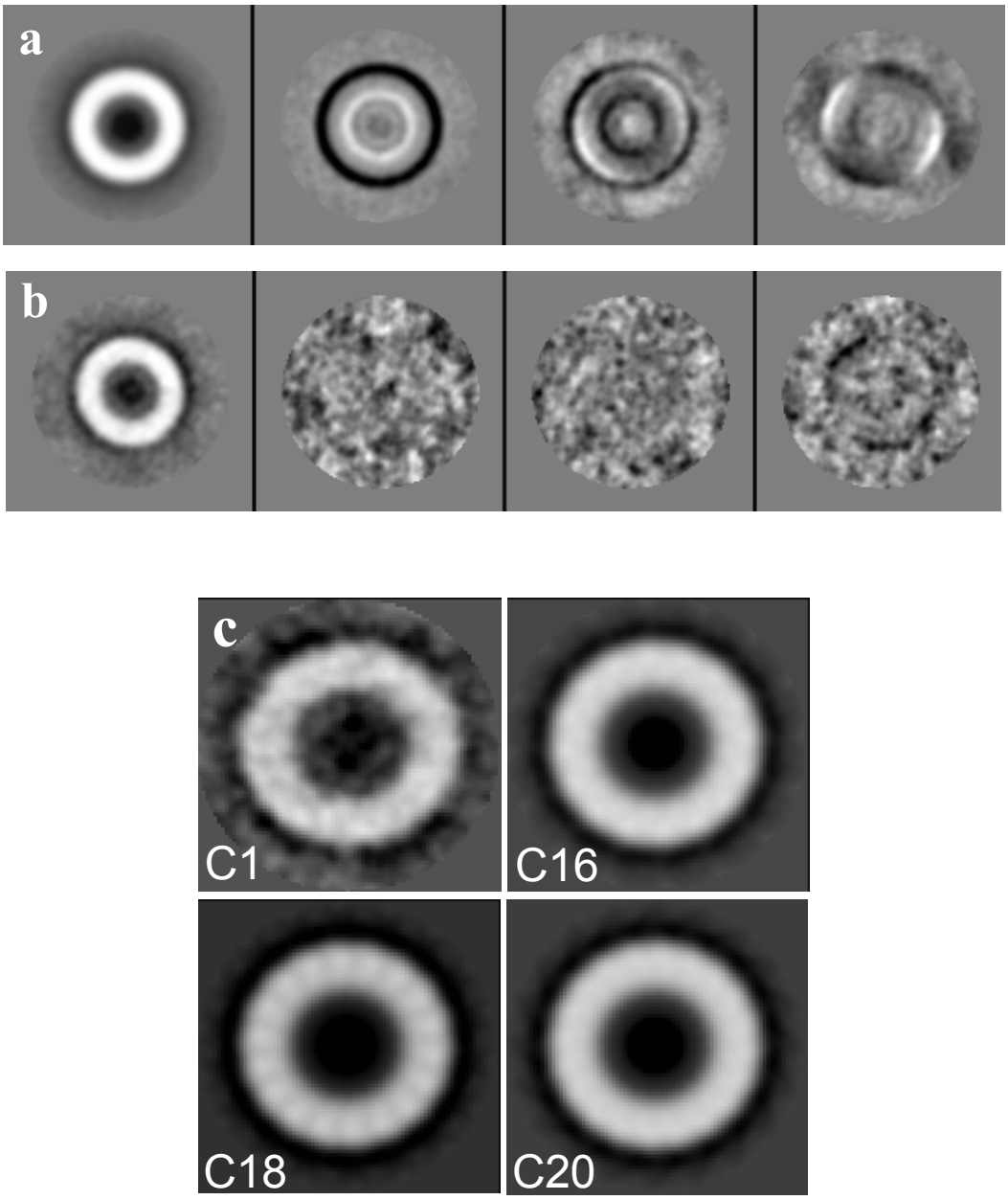

Figure 2.5 Single particle analysis of negatively stained S105
Eigenimages of the first 4 factors used during classification of (a) complete data set and (b) after particle extraction from a single class. The second and third eigenimages in (a) indicate the size variability in the data set as indicated by the concentric circles. These circles are absent in (b). Class averages obtained from analysis of (b) by imposing different symmetries (C1= no symmetry imposed).

observations suggest that the 24.5-nm ring consists of 18 repeating units or protomers. Given the dimensions of the ring and the molecular mass of S105 (~12kDa), it is not possible that one ring could be comprised of 18 S105 molecules. It is therefore likely that S needs to oligomerize in order to form a protomer.

2.2.4 Cryo electron microscopy of S rings

Negative staining provides good contrast for biological molecules but the density observed may not be an accurate representation of the true protein density. Cryo-electron microscopy, albeit lower contrast, provides true protein density retrieval without the distorting effects associated with staining and dehydration. A cryo-electron microscopy dataset of S105 in DDM was collected and the data ware separated based on particle diameters as before (Figure 2.6). Particles from the same defocus micrographs were manually selected and classified using only the first 5 eigenvectors. From the initial 1836 particles, 1079 (~60%) fell into classes that had the same outer diameter. The 3-D reconstruction of the S-rings (ring-dimers) assuming C18 rotational symmetry from the most frequent size group indicates that each ring measures 23 nm in diameter with an inner diameter of 9 nm for the upper ring and 8 nm for the lower ring (Figure 2.7a). The difference in the inner diameters of the lower vs. upper ring may arise from conformational changes required for the interaction. Each ring in the dimer has a flat lower and a tapered upper surface. This gives rise to a slight radial asymmetry with each ring measuring 4 nm in height towards the outer perimeter and tapering towards the central hole. This side-on feature of the rings does not indicate a 2-fold symmetry

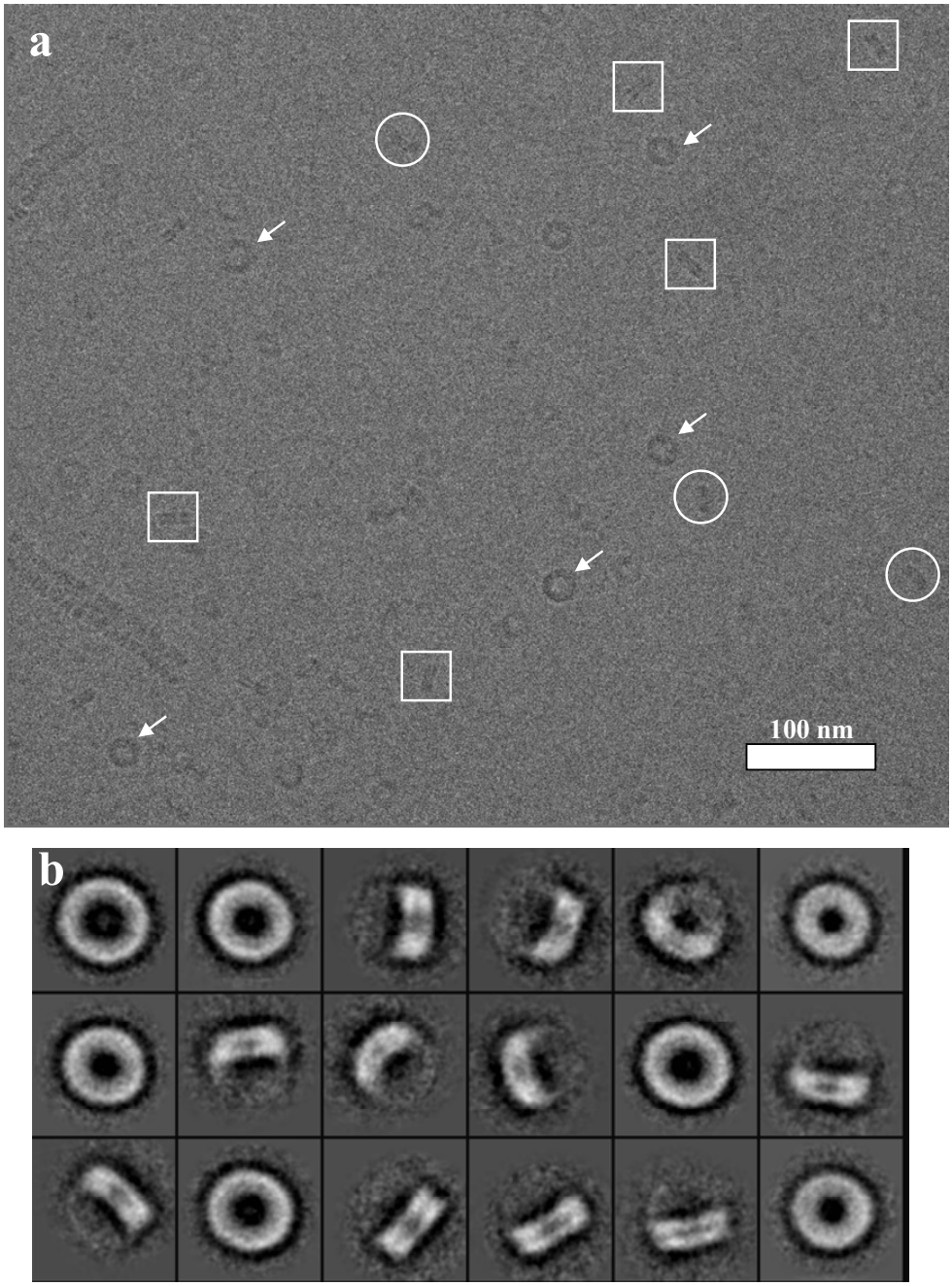

Figure 2.6 Cryo-EM and particle classification of S105
Cryo-electron micrograph of S105 in TBS, 0.1% DDM (a). Boxes indicate side-on views, circles indicate intermediate views and arrows indicate face-on views. Initial classification of particles to separate data according to diameter (b).

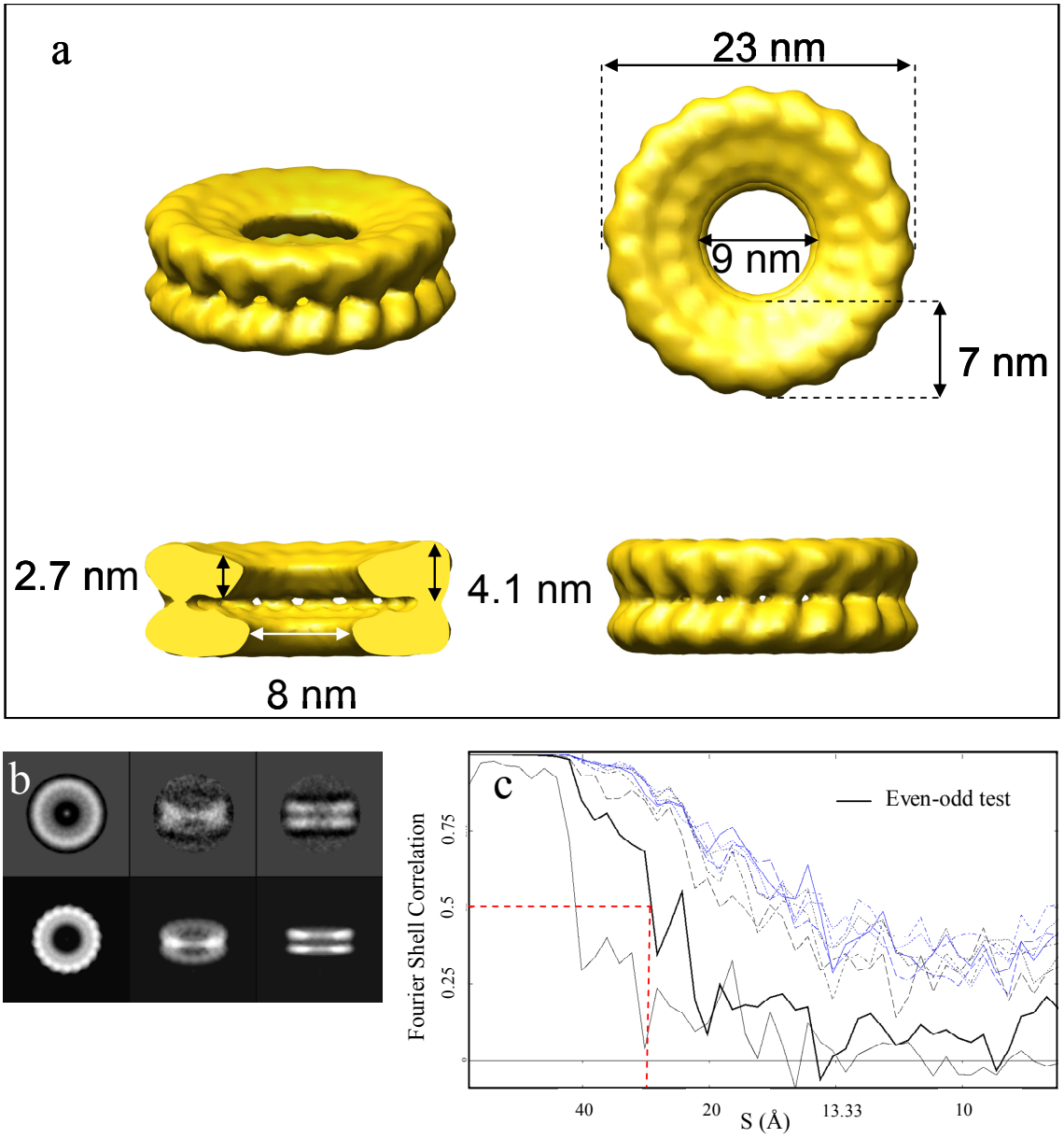

Figure 2.7 Three-dimensional reconstruction of S105 ring dimer
The cryo-EM structure of S105 ring dimer (a) was calculated from 1079 particles extracted from the initial 1836 particle data set. Main class averages (b, top row) and corresponding reprojections of the 3D volume (b, lower row) show good agreement and confirm the accuracy for the rendering threshold The final resolution was 2.6 nm as calculated by the Fourier shell correlation with a 0.5 cut-off (c).

suggesting a head-to-tail arrangement of the upper vs. the lower ring rather than a tail-to-tail arrangement. The final resolution of the 3D map was 2.6 nm as judged by the Fourier shell correlation using the 0.5 criterion (Figure 2.7c).

2.2.5 S105 remains soluble in the absence of detergent

DDM is one of the preferred detergents for maintaining membrane protein integrity but is characterized by a low critical micelle concentration (CMC) (0.008% w/v), the concentration at which half the detergent molecules are in micelles. DDM also has a high aggregation number, making its removal by dialysis very ineffective. To investigate the minimum concentration of DDM required to keep S105 in solution for liposome reconstitution, the initial DDM concentration of 0.1 % (w/v) was reduced by stepwise addition of buffer. Interestingly, at ~0.004% (w/v) DDM, S105 began to form ring-shaped oligomers with larger dimensions than observed at the higher DDM concentrations (mean diameter of 40 nm). These ring-shaped structures were very heterogeneous with regards to diameter, shape and ring thickness around the periphery (Figure 2.8a). To ensure that sub-CMC concentrations of DDM were not responsible for these structures, S105 was dialyzed in the presence of Calbiosorb^{TM} biobeads which bind any residual detergent. Surprisingly, the large ring-like structures were still observed, and even after further dilution with distilled water these “soluble rings” were stable in solution. The ability to form soluble rings was not observed with $S105_{a52v}$, which aggregated and precipitated once detergent was removed.

Dialysis from EBB-purified S105 into TBS also yielded soluble rings but only if

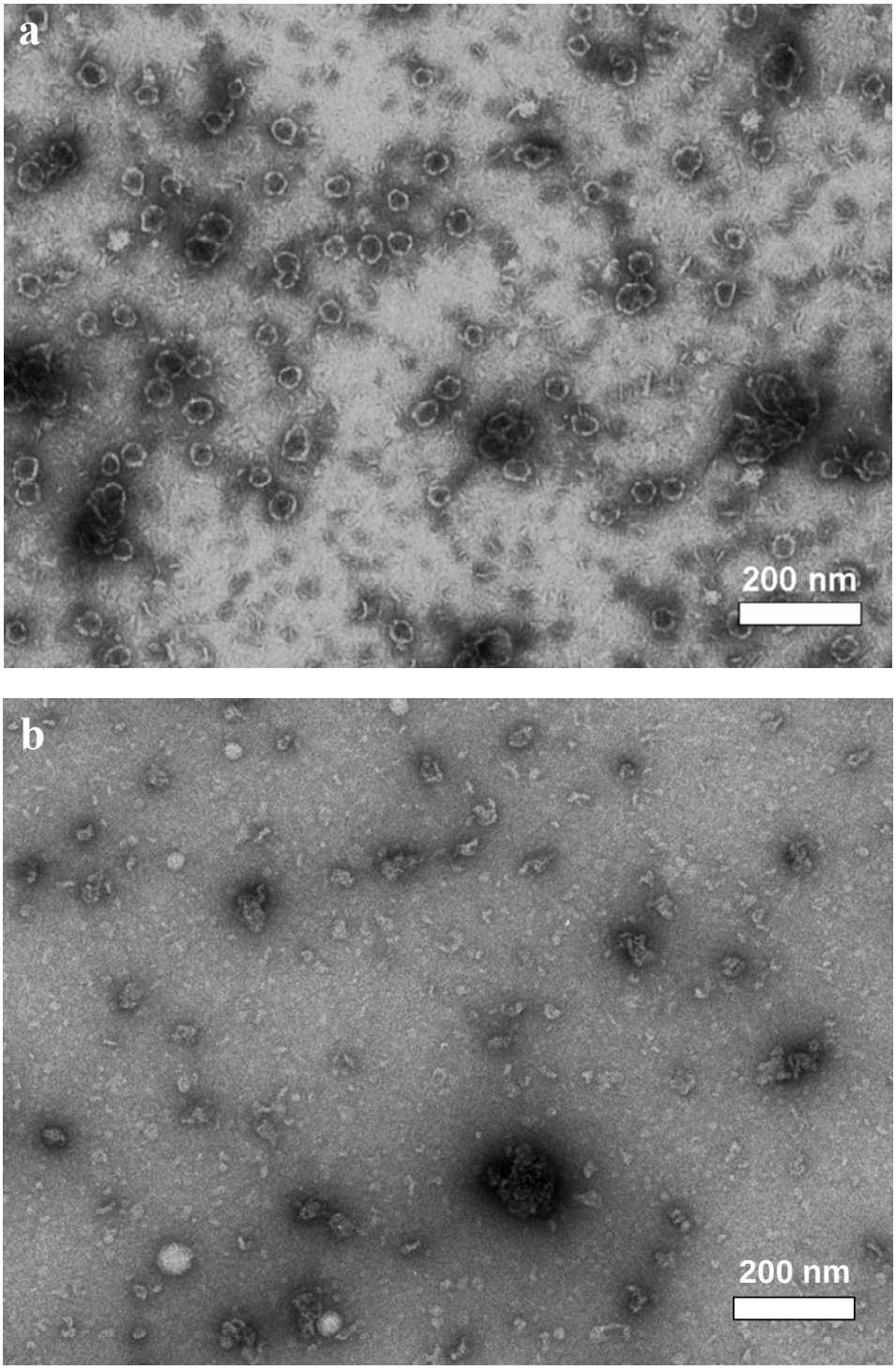

Figure 2.8 S105 remains soluble in the absence of detergent
S105 in 1% (v/v) EBB was dialyzed against TBS in the presence (a) or absence of (b) 1 mM DTT.

1mM DTT was included in the dialysis buffer. The requirement of S105 for DTT in order to remain water soluble suggested that the single cysteine, C51, was being modified and possibly leading to disulphide bonded dimers of S that could not oligomerize to form the large soluble rings. To investigate this hypothesis, the cysteineless mutant $S105_{c51s}$ was purified in 1% EBB (v/v) and dialyzed against TBS with and without 1mM DTT. $S105_{c51s}$ was found to form the soluble rings irrespective of the presence of DTT, indicating that C51 in the parental S105 was being oxidized to a state that prevented soluble ring formation (Figure 2.9).

Interestingly, soluble rings could revert to the smaller homogeneous rings by addition of 0.1% (w/v) DDM or 1% (w/v) DM. This conversion was very rapid and occured within 5 min of addition of DDM, roughly the time required to add DDM and prepare grids for EM analysis.

Finally, when samples were removed from dialysis prior to complete detergent removal, structures that resembled “honeycombs” were observed (Figure 2.10). These assemblies of rings that measured ~60 nm across seemed to be forming by the association of the rings observed in DDM. However in the same samples, soluble rings were also present. Thus at these sub-cmc detergent concentrations, intermediates in the transition from DDM-rings to soluble rings, were captured. The rings, which are probably still hydrophobic at their outer surface, are likely associating together to shield these surfaces from water. The outer surface of these “honeycomb” structures seems continuous and may have already switched to expose hydrophilic surfaces. In summary when S105 is solubilized in a non-ionic detergent, relatively, homogeneous rings are

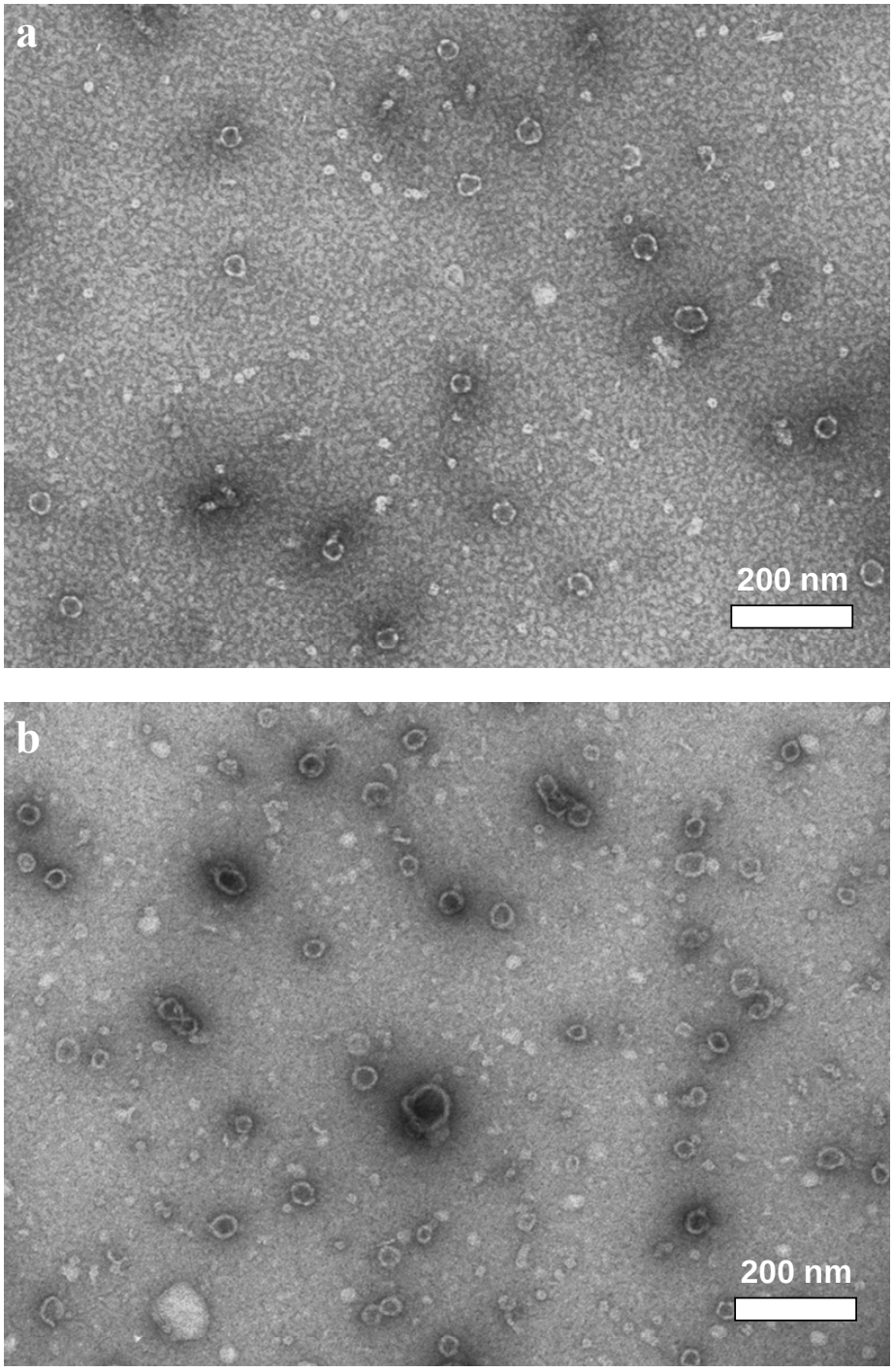

Figure 2.9 $S105_{c51s}$ does not require DTT to form soluble rings
$S105_{c51s}$ in 1% (v/v) EBB was dialyzed against TBS in the presence (a) or absence of (b) 1 mM DTT.

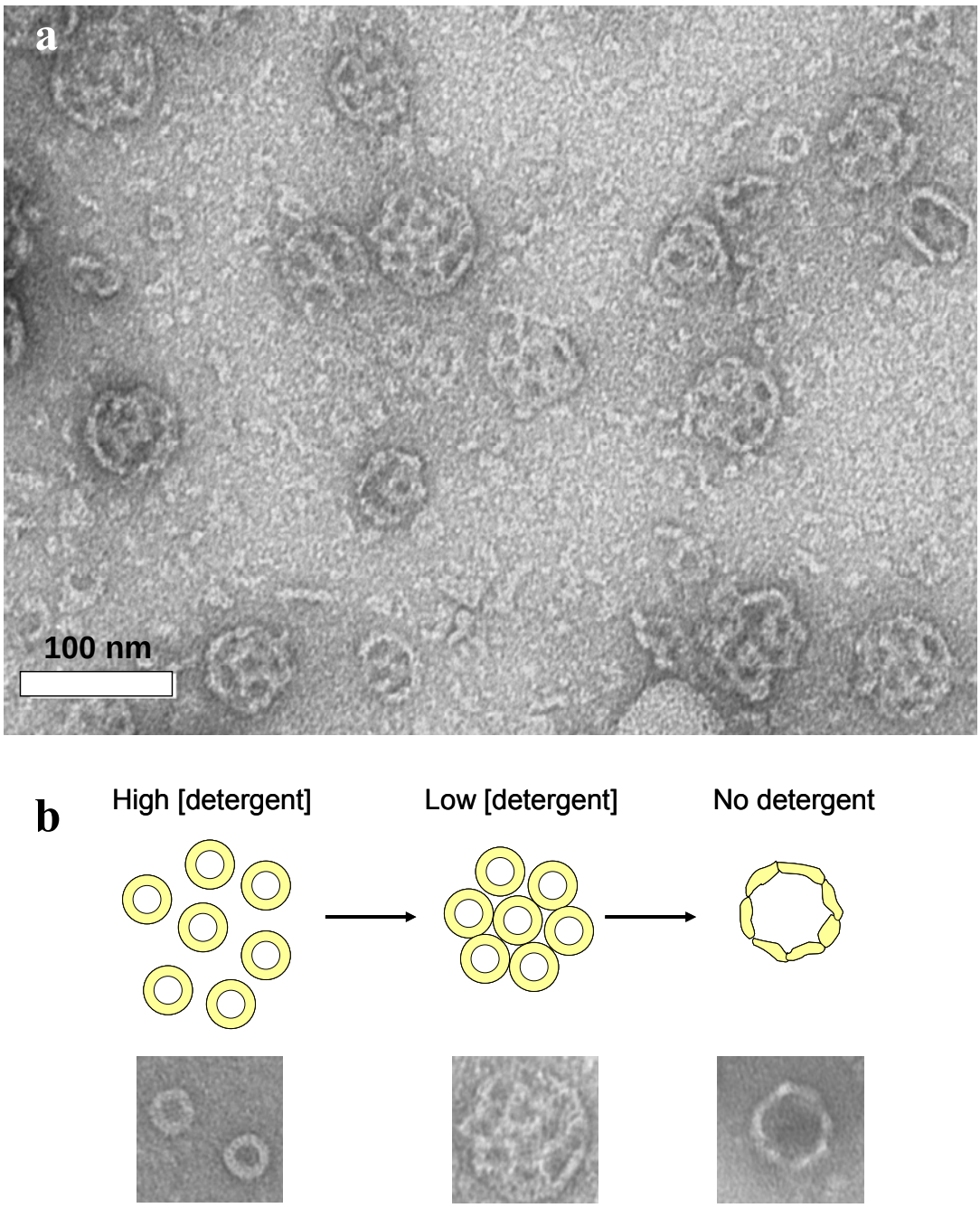

Figure 2.10 S105 forms rafts of rings at low detergent concentration
S105 was dialysed agianst TBS, 1 mM DTT to remove EBB and stopped prior to complete removal of EBB (a). Summary of structures formed by S105 at different detergent concentrations.

formed. When the detergent concentration reaches a critical minimum, these rings will associate to form rafts of rings that resemble “honeycombs”. Finally, upon complete detergent removal, S105 remains water soluble by forming the larger, heterogeneous soluble rings.

2.2.6 Circular dichroism spectroscopy

The unexpected finding that S105, an integral membrane protein, can remain soluble in the absence of detergents prompted further investigation into the nature of the soluble rings. Using circular dichroism spectroscopy, we compared the secondary structure content of S105 in PBS (soluble rings) and S105 in PBS after addition of EBB or DDM. All three buffer conditions gave almost identical overlapping spectra and an α-helical content of 48% or approximately 55 residues (Figure 2.11). As a control, $S105_{a52v}$ and $S_{\Delta TM1}$ purified directly in EBB were measured for their α-helical content. $S105_{a52v}$ also contained 48% α-helix and as expected, $S_{\Delta TM1}$ only contained 38% or 36 residues α-helix. These data closely matched previous CD measurements (64) and showed that soluble S105 retains its α-helical character and enough α-helix to comprise three TMDs.

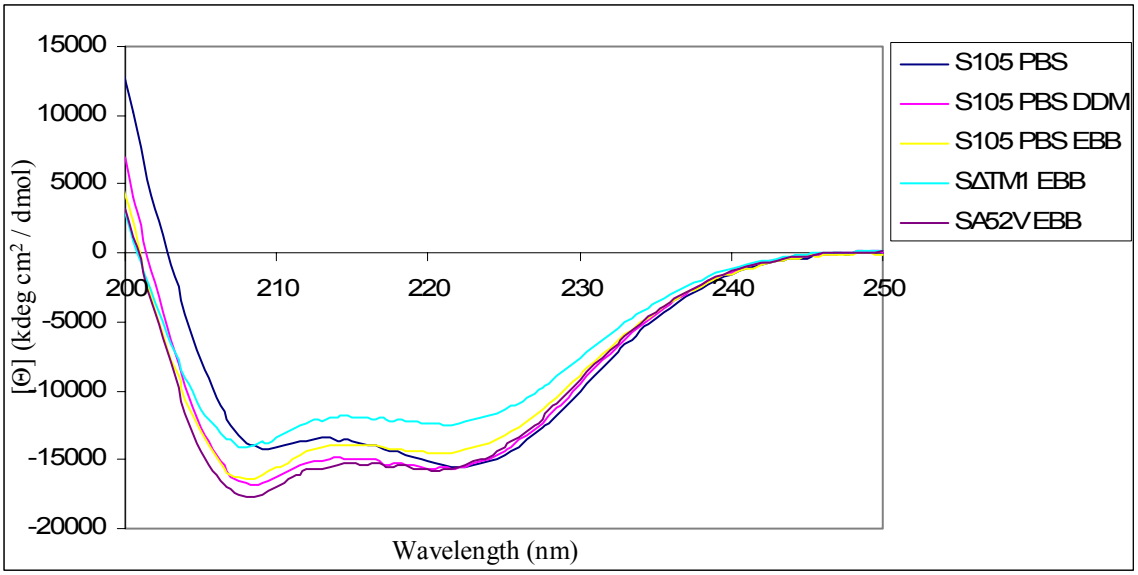

Figure 2.11 Circular dichroism spectroscopy of S alleles
CD spectra of S105 soluble rings as well as after the addition of detergent are shown. The percentage alpha-helix contents are given in the results section.

2.2.7 Crosslinking S105 in *E. coli* membrane*s*

The multimeric rings observed after extraction in non-ionic detergents may represent a form that that is present in the membrane at a stage after S105 oligomerization. As previously reported, oligomers can be captured in the membrane using the membrane permeable crosslinker DSP (Dithiobis-succinimidylpropionate). DSP has 2 reactive N-hydroxysuccinimide groups that react with primary amines such as the N-terminal methionine or lysines and has a 12Å spacer arm. Furthermore crosslinking is reversible by addition of reducing agents that break the disulfide group. To study the nature of S oligomers in the membrane, S105 was overexpressed in *E. coli* C43(DE3) and IMVs were isolated by sedimentation centrifugation. Trial studies using increasing amounts of DSP (50 μM-30 mM) indicated that 1 mM DSP was sufficient and no increase in crosslinking was observed beyond that concentration. S105 IMVs were subjected to 1 mM DSP and extracted using DDM or the zwitterionic detergents EBB or Zwittergent (Figure 2.12). Crosslinked S105 was then purified and gel-permeation profiles were obtained. Due to the fact that very large complexes might arise from such an experiment, the second ultracentrifugation step after detergent extraction was omitted to avoid sedimentation of such a complex. Instead, total solubilized membranes were filtered through a 0.22 μm filter.

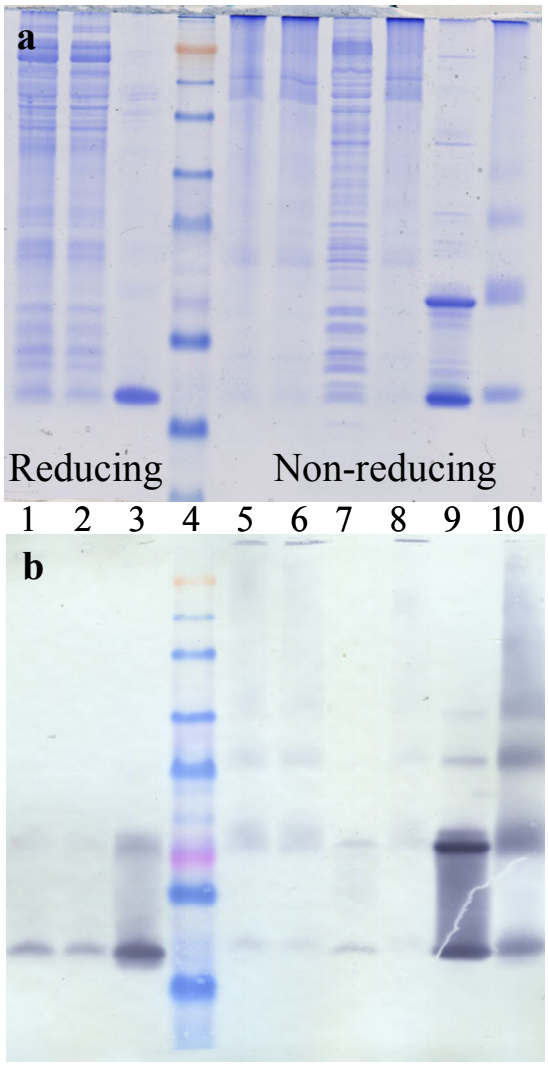

Figure 2.12 DSP crosslinking of S in *E.coli* membranes
Coomassie blue (a) and α-S105 (b) SDS-PAGE of S105. (1) 1mM DSP membranes before filtering (b) after filtering (3) 1 mM DSP TALON eluate (4) MW marker (5) same as 1 (6) same as 2 (7) 0 mM DSP TALON flow-through (8) 1 mM DSP TALON flow-through (9) 0 mM DSP TALON eluate (10) 1 mM DSP TALON eluate.

In the absence of crosslinker, the gel-permeation profiles of S105 remained largely unchanged (Figure 2.13a). S105 in DDM gave three characteristic peaks corresponding to stacks, rings and dimers. S105 in EBB eluted in the low molecular mass region with two peaks corresponding to monomers and dimers. When membranes were subjected to 1 mM DSP prior to detergent extraction, both detergent profiles showed changes. The DDM profile indicated a large decrease in the void volume (stacks) and a broadening of the second peak corresponding to rings. Interestingly the third peak was also significantly decreased. Overall DSP crosslinking prior to DDM extraction resulted in a shift of the protein population to the second peak, which corresponds to rings and small stacks.

DSP crosslinking prior to EBB extraction resulted in two observations. Firstly, protein eluted immediately after the void volume with a steady increase towards low molecular weight species indicating that high molecular mass species had been crosslinked. Secondly, the ratio of dimer to monomer increased as compaired to uncrosslinked samples. This almost linear increase is probably due to the increase of crosslinking efficiency towards smaller complexes, which is also indicated by the western blot and the ladder effect of crosslinking observed (Figure 2.12).

Electron microscopic examination of the fractions from this experiment showed that S105 extracted in DDM after DSP treatment formed rings but no stacks were observed (Figure 2.13c). In many cases the rings were incomplete, thus forming arc shaped structures. Unfortunately, examination of the higher molecular weight fractions from EBB or Zwittergent-extracted DSP samples revealed protein aggregates. This is not

surprising since zwitterionic detergents break protein-protein interactions, and these complexes are now held together only by the crosslinker itself, instead of the intrinsic protein interactions.

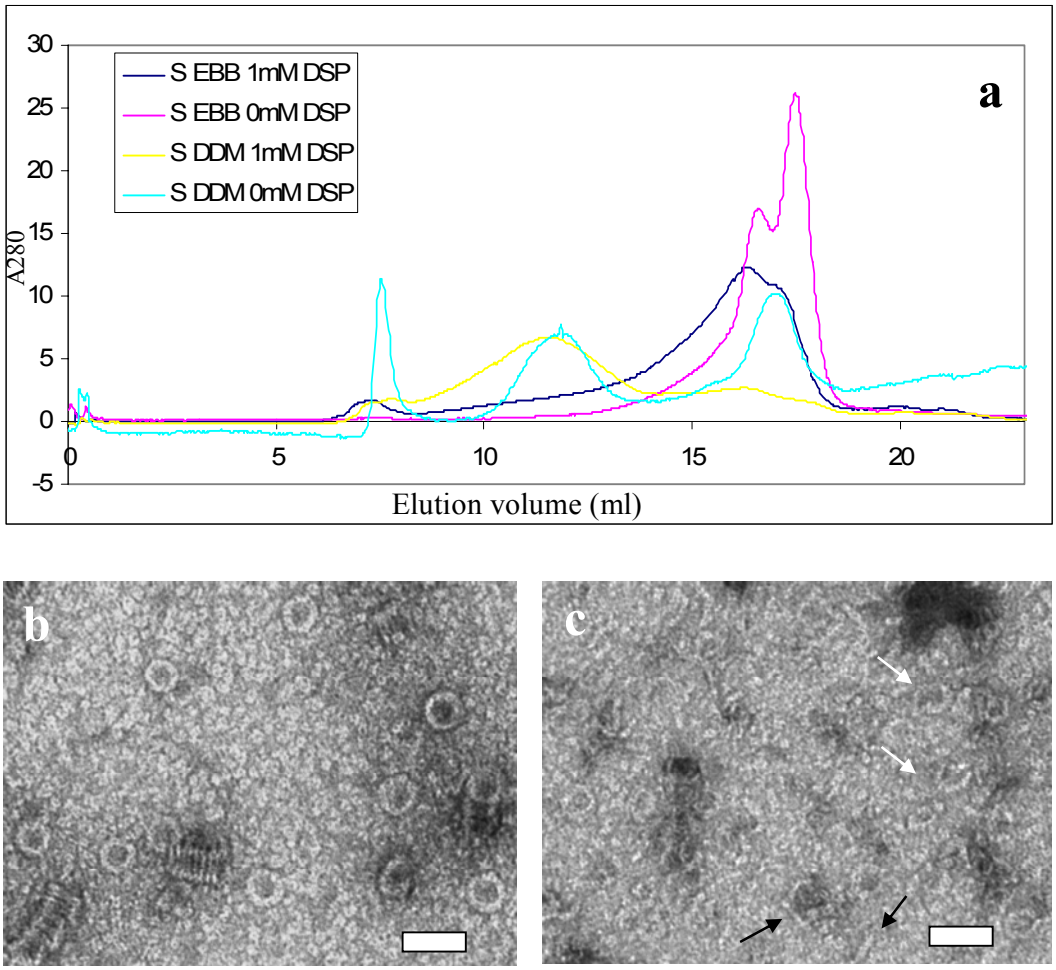

Figure 2.13 DSP crosslinked S105 gel-permeation results and EM
Superose-6 elution profiles of DSP crosslinked S105 in DDM and EBB. Samples from fractions corresponding to the DDM 12th elution fraction were examined by EM. (b) 0 mM DSP (b) 1 mM DSP (white arrows indicate arcs, black arrows indicate linear fragments). Scale bars correspond to 50 nm.

2.2.8 Observing S in *E. coli* membranes

The ring structures observed in non-ionic detergent may represent the multimeric form that is present in the lipid bilayer as detergents mimic the lipid environment of membrane proteins by protecting the hydrophobic TMD cores from water. However the exact nature of the lipids (head-group charge, size and distribution) and their ability to form bilayers as opposed to micelles may affect the multimeric assembly formed by S. To this extent, an effort to directly visualize overexpressed S105 in native *E. coli* membranes was carried out. *E. coli* C43 (DE3) cells overexpressing *S105*, $S105_{a52v}$ or $S_{\Delta TM1}$ were disrupted by French-press and IMVs were harvested. Since the membrane vesicles are primarily inside-out (100), the His6-tag proximal to the C terminus of S105 should be accessible for binding to a metal-affinity resin. IMVs were resuspended and passed over a TALON column to enrich for vesicles containing S105. Estimation of the total yield showed that only 5-10% of S105 could be immobilized and eluted using this procedure, probably because the larger vesicles containing S105 were unable to enter the pores of the matrix. Eluted fractions were negatively stained and examined by EM. IMVs containing S105 were observed to be generally smaller than 100 nm. Ring-shaped structures were observed that resembled purified soluble S105 rings both in size and morphology, particularly in the segmented pattern of these rings (Figure 2.14a, d).

$S105_{a52v}$ fractions displayed an entirely different pattern. Large vesicles with diameters up to 150 nm were abundant and no ring structures were observed (Figure 2.14e). Furthermore these vesicles were not observed in the vector-only control (data not shown). Since the amount of $S105_{a52v}$ expressed is almost 5 times greater than

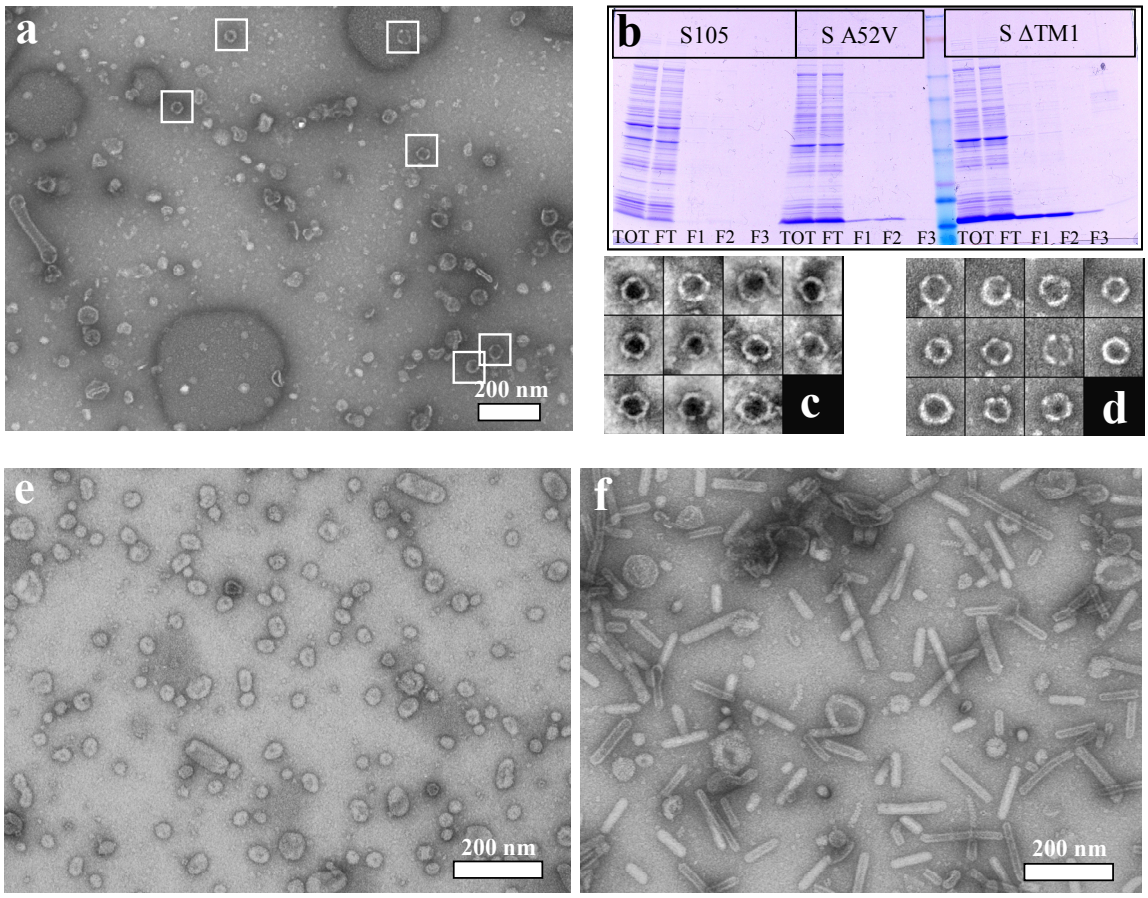

Figure 2.14 Enrichment of S IMVs from *E.coli* C43(DE3)
Cells overexpressing S alleles were French pressed and IMVs were enriched for S by passing over metal affinity resin. The S105 elution (a) showed rings (white boxes and (d)) similar in morphology and size to purified soluble S105 (c). SDS-PAGE (b) of affinity purification. $S105_{a52v}$ (e) and $S_{\Delta TM1}$ (f) fractions.

S105, and SDS-PAGE indicates that $S105_{a52v}$ is the major protein in these fractions, these vesicles are likely to be densely packed with $S105_{a52v}$. A fast Fourier transform (FFT) of the vesicle images did not reveal any reflections, indicating that the S105 molecules are not packed in a regular array.

The fractions from $S_{\Delta TM1}$ also displayed densely packed protein vesicles but these were elongated and measured up to 200 nm in length (Figure 2.14f). These tubular shaped vesicles also showed no indication of regular protein packing.

The observation that S105 enriched IMV fractions contained soluble S105 rings raised the question of how much S105 may be soluble in the cell under over-expression conditions. Since the large soluble rings showed a sedimentation behavior similar to the vesicles, sedimentation centrifugation could not be used to separate membrane vesicles from the soluble protein. To achieve this separation, flotation gradients were used. A preformed gradient of iodixanol was prepared as shown (Figure 2.15). French-press lysates and purified soluble S105 rings were loaded on the bottom of the gradient at a final iodixanol concentration of 30% (1.21 g/ml) and overlaid with buffer containing 44% iodixanol (1.16 g/ml). Finally buffer was overlaid on top of the 44% layer and the samples were spun at 165,000 x g for 3 hours at 4ºC. Fractions were removed from the top of the gradient and analyzed by TEM and Western blotting.

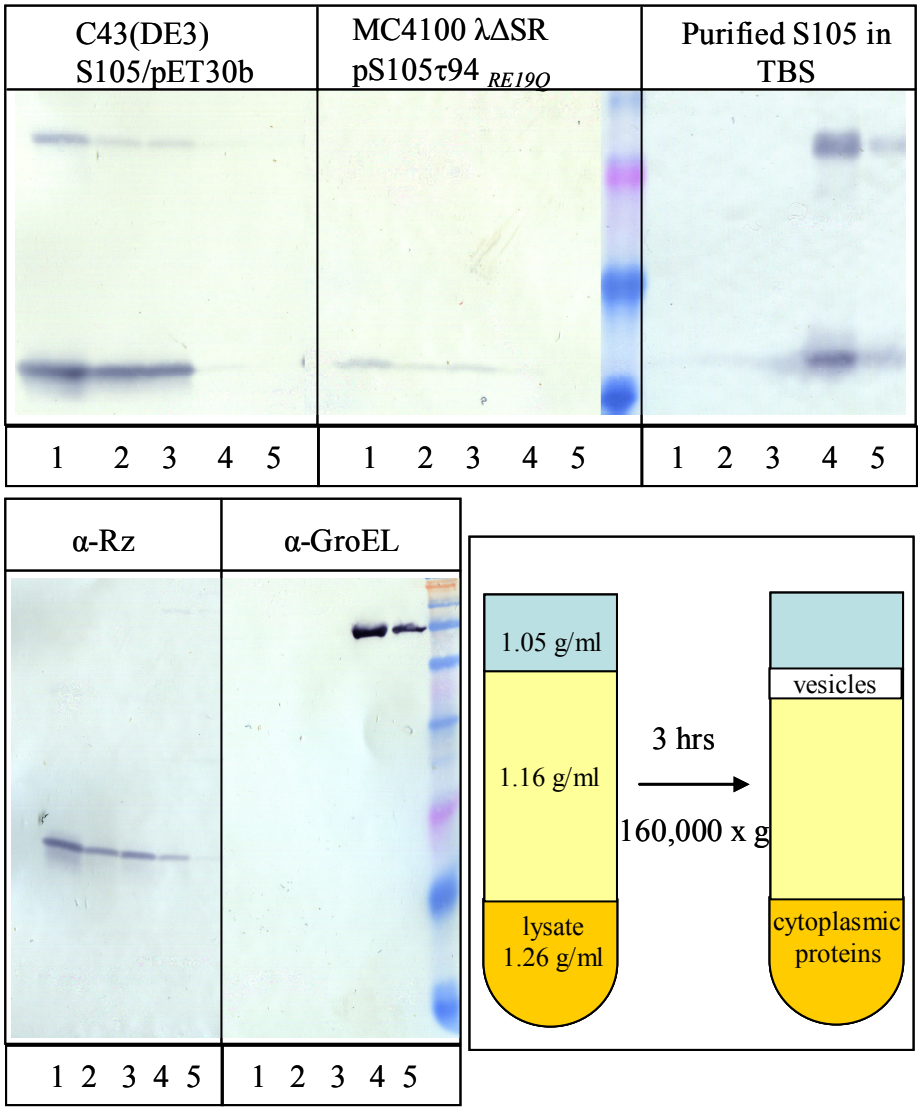

Figure 2.15 Flotation gradient centrifugation of S IMVs
Lysates from cell expressing *S105* and purified S105 soluble rings were subjected to flotation gradient centrifugation. Fractions were collected from the top of the gradient and subjected to Western blot analysis. For the localization of Rz and GroEL, samples were obtained from the MC4100 λΔSR pS105τ94 $_{R\,E19Q}$, fractions and developed using their respective antibodies.

Western blots of the resulting flotation fractions are shown in Figure 2.15. Controls for the correct localization of the inner membrane proteins and cytoplasmic proteins were used from MC4100 λΔSR pS105τ94 $_{R\ E19Q}$ samples. Rz, an inner membrane protein, is encoded on pS105 and the cytoplasmic chaperonin GroEL is endogenous to the expression host. In order to avoid lysis of the cells prior to their harvesting the catalytic glutamate, E19, of R (55) was mutated to a glutamine by site-directed mutagenesis. Westen blots indicated that GroEL was present in the two highest density fractions (fractions 4 and 5) and Rz was primarily present in the three lower density fractions (fractions 1-3). S105 expressed from MC4100 λΔSR pS105τ94 $_{R\ E19Q}$, localized almost exclusively to the membrane vesicle fractions with less than 5% present in the cytoplasmic fractions. S105 expressed from C43(DE3) S105/pET30b also localized to the first three membrane fractions with approximately 1% in fraction 4. Purified S105 soluble rings failed to float and remained in the high density fractions (4 and 5). The small amounts of S105 present in the cytoplasmic fractions may be a result of contamination during fractionation, as a small amount of Rz was also found in fraction 4. Collectively, the flotation results confirm previous isopycnic gradient results that S is localized to the membrane (60). Furthermore overexpression does not result in accumulation of cytoplasmicly localized S105. Finally, purified S soluble rings do not float to lower densities, consistent with them being protein complexes of high density.

2.2.9 Liposome reconstitution of S

To further investigate the nature of S105 multimers in a lipid environment, liposome

reconstitution trials were carried out. S105 purified in either DDM or EBB was mixed with dried lipid at protein to lipid ratios of 1:500-1:1000. After extensive dialysis to remove detergent in the presence of biobeads, samples were floated on a sucrose step gradient and fractions were subsequently blotted for protein localization. The flotation results shown in Figure 2.16 indicate that when dialyzed from EBB, S105 and S105 $_{a52v}$ associate with the liposomes as seen by the protein localization in the low (0-20% sucrose) density fractions. In contrast, dialysis from DDM resulted in most of the protein remaining in the higher density fractions but with some amount floating with the liposomes. As a control, soluble S105 rings (S105 in TBS) were added directly to pre-formed liposomes and subjected to the same flotation procedure. In this case, all the protein remained in the high density region and did not float along with the liposomes. Examination of the liposome reconstituted samples with TEM resulted in different results for samples dialyzed from DDM and EBB. Dialysis from EBB solubilized S105 (Figure 2.17-d) did not reveal any S mediated lesions in the liposomes. In fact control liposomes (with no protein) resembled both liposomes reconstituted with $S105_{a52v}$ and those reconstituted with S105. S105 reconstituted from DDM however, could be visualized as large rings that resemble the S105 soluble rings (Figure 2.18). The rings were sometimes associated with the liposomes but could also be seen separated from them.

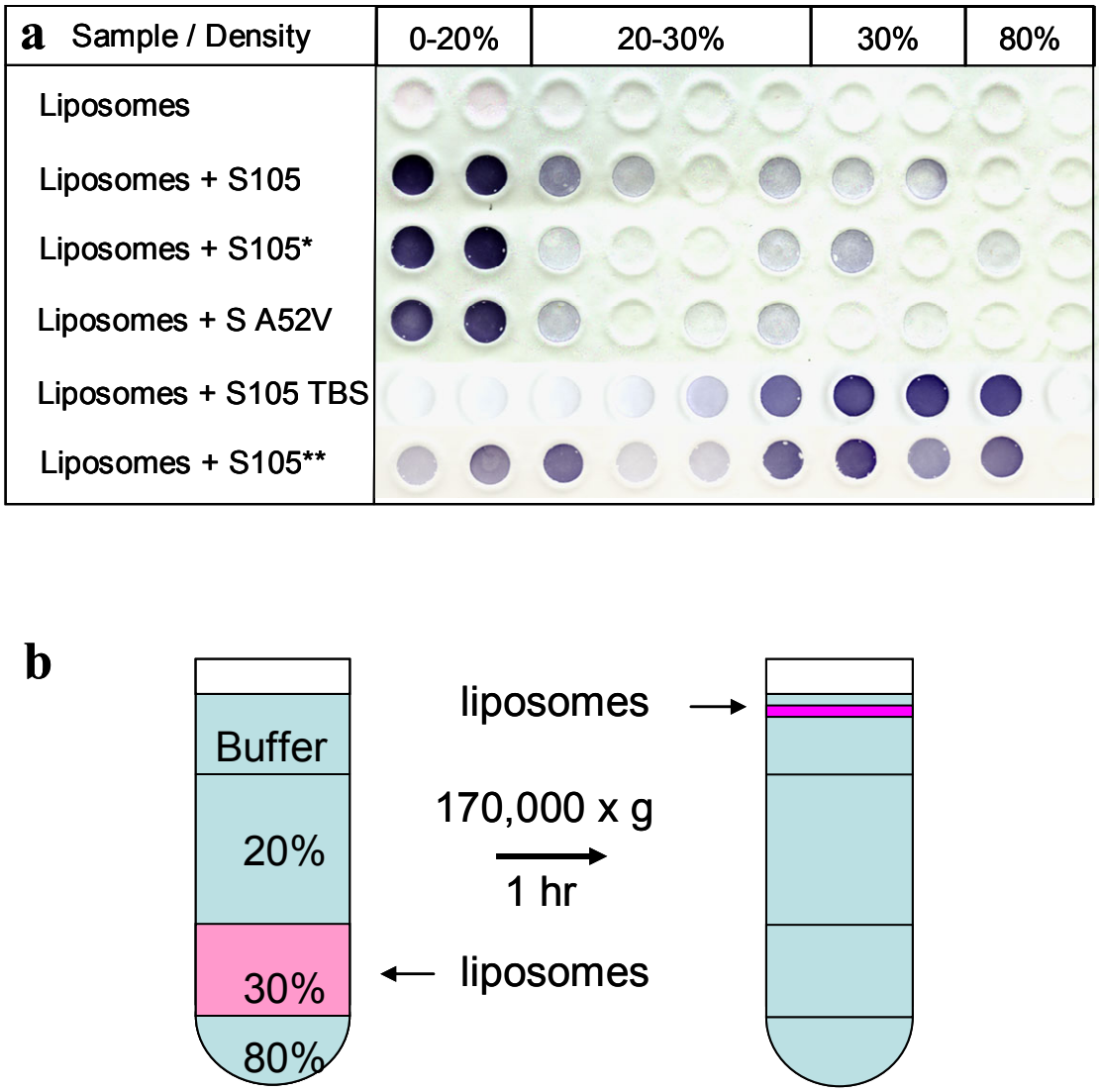

Figure 2.16 Liposome reconstitution and sucrose gradient flotation assay
S105 in EBB was reconstituted into PC:PG liposomes by dialysis (a). The samples were then floated on a sucrose gradient and fractionated from the top (low density). Fractions were blotted and developed using α-S105 antibody against the N terminus of S105. * indicates S105 that had been reduced with DTT prior to reconstitution. ** indicates S105 that was reconstituted from DDM instead of EBB. A schematic representation of the flotation gradient is shown (b)

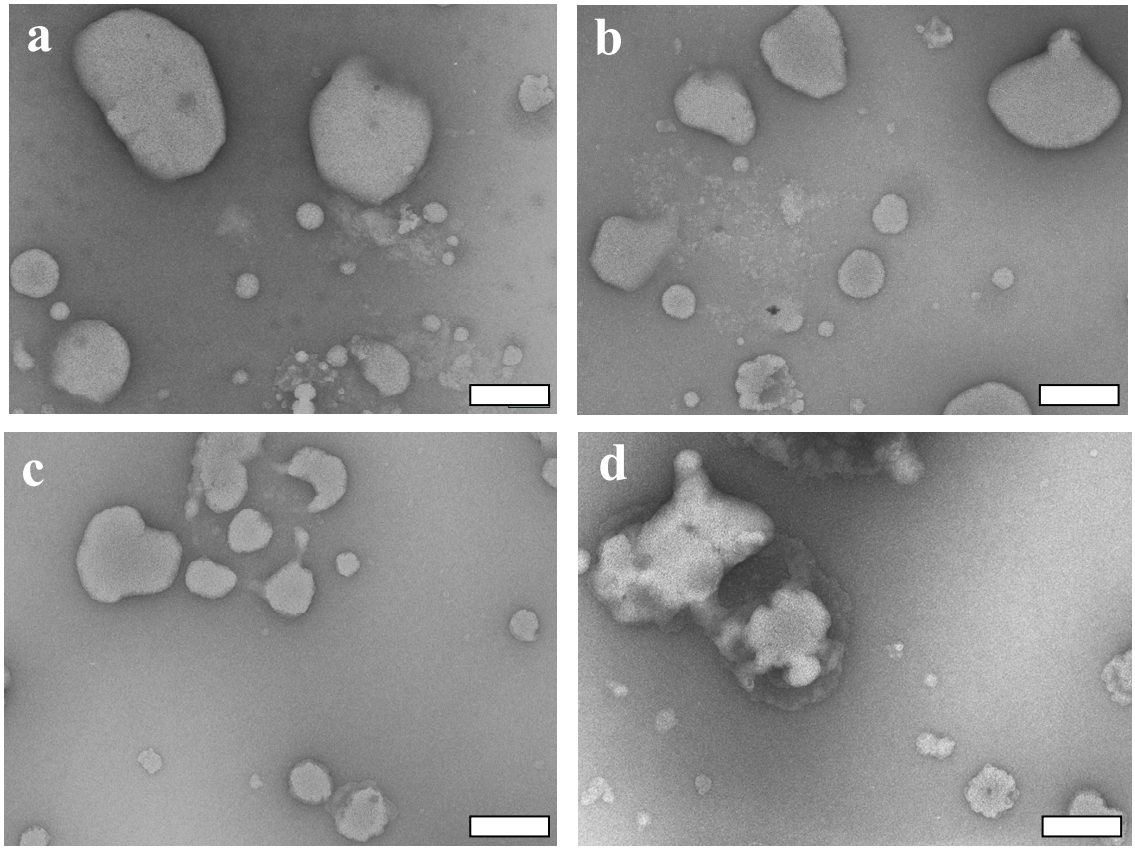

Figure 2.17 Liposome reconstitution of S105 from EBB
Electron micrographs of flotation fractions (0-20%) after S reconstitution from EBB. Shown are liposomes only (a) $S105_{a52v}$ plus liposomes (b) S105 plus liposomes (c, d). The protein to lipid molar ratio was 1:1000. Scale bar corresponds to 400 nm.

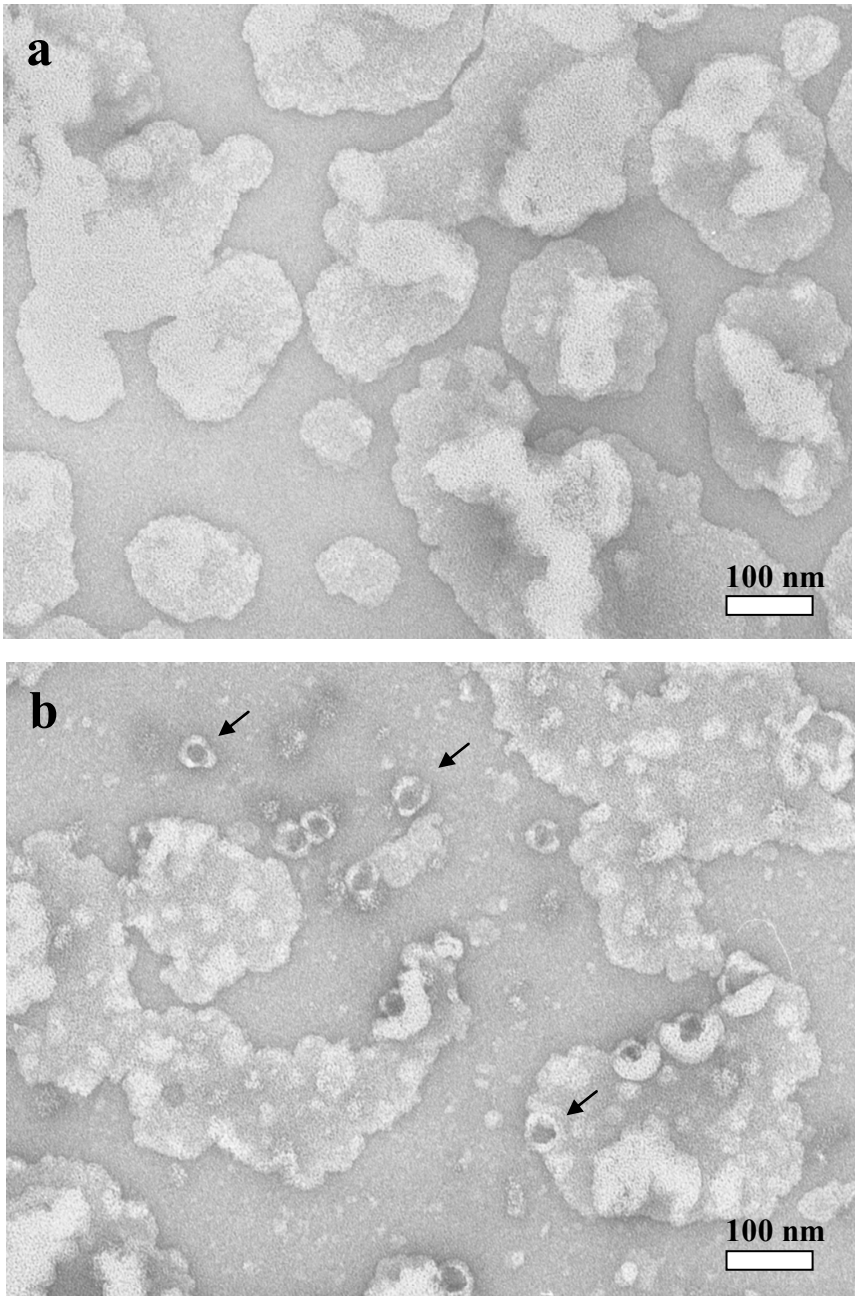

Figure 2.18 Liposome reconstitution of S105 from DDM
Electron micrographs of total dialysis samples after S105 reconstitution from DDM. Shown are liposomes only (a) and S105 plus liposomes at a molar ratio of 1:500. Arrows indicate structures that resemble the water soluble form of S105, sometimes present around the liposomes and sometimes associated with the liposomes.

2.2.10 Cysteine accessibility and nanogold labeling of DDM solubilized S

Previous studies on the accessibility of cysteine reactive agents on residues in the S105 polypeptide were used to map the three transmembrane domains of S105 (100). Using SCAM (substituted cysteine accessibility method) Gründling *et al*. showed that stretches of residues in the predicted transmembrane regions of S105 are resistant to modification by the cysteine modifying agent IASD (iodoacetamide) in IMVs but are accessible after treatment with 2% SDS. Moreover, the intervening loop regions are easily modified as expected. These experiments showed that the 3 regions are inaccessible to the membrane impermeable reagent as they were embedded in the lipid bilayer. Interestingly, the C terminus of S105, which is rich in basic residues and cytoplasmically located, is resistant to IASD labeling. Only after boiling at 100 ºC or treatment with high salt, did the cysteines at positions 94 and 95 react to IASD. This suggested that the C terminal cytoplasmic region is interacting with the negatively charged inner leaflet of the bilayer.

To investigate the accessibility of cysteines in the DDM-solubilized S105 rings, the reagent PEG-maleimide (molecular mass of 10 kDa) was used to modify two cysteine mutants of S105 ($S105_{c51s\ h7c}$ and $S105_{c51s\ a99c}$) and the single cysteine (C51) in the middle of the second transmembrane domain of wild type S105 (Figure 2.19a). Pegylation was monitored by a 10 kDa size shift on SDS-PAGE.

The pegylation results for DDM rings are shown in Figure 2.19b. In S105, cysteine 51 was not modidied, however, pegylation of S was obsrved once the rings were solubilized by SDS, which breaks down the rings to monomers and dimers. In $S105_{c51s\ h7c}$, the cysteine was accessible in the absence of SDS indicating that C7 must be

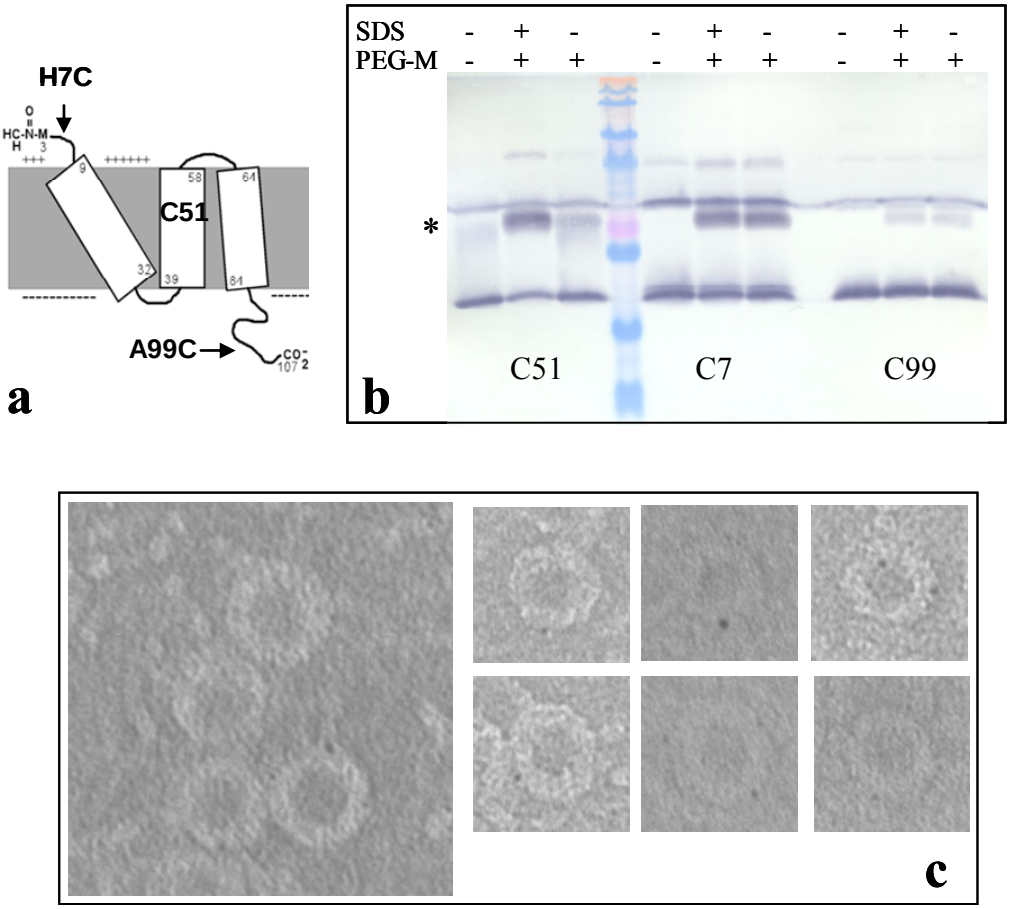

Figure 2.19 Cysteine accessibility and nanogold labelling of S105 DDM rings
Cysteines at the N-terminus (C7), 2nd TMD (C51) and C-terminus of S105 (C99) (a) were screened for their accessibility to PEG-maleimide in the context of DDM rings (b). The asterix (*) indicates the position of the pegylated product. The accessibility of C7 was used to localize the N terminus of S by monomaleimido nanogold labeling and TEM (c).

exposed in the DDM rings. Finally when the cysteine was placed at the far C-terminus (A99C) it became inaccessible to pegylation even in the presence of SDS. Incomplete modification at positions 7 and 51 is attributed to oxidation that occurs during the metal affinity purification and it has been shown that protecting groups attached to the cysteine prior to purification prevent oxidation of the cysteine and allow 100% labeling (J. Dewey and R. Young, unpublished data)

The accessibility of the cysteine at position in $S105_{c51s\ h7c}$ was used to label and localize the N-terminus of S105 in the DDM rings. Using monomaleimido nanogold which carries a gold atom cluster 1.4 nm in diameter and reacts with cysteines, we were able to localize the N terminus of S105 in the DDM rings. Although the labeling efficiency seemed low, the majority of labeled rings had the nanogold cluster at the outer periphery of the rings (Figure 2.19c).

2.2.11 Protease protection of S DDM rings and S in IMVs

The C terminus of membrane embedded S105 has been shown to reside in the cytoplasm as determined by its accessibility to proteinase K in IMVs (61). In order to investigate what regions of S are accessible to protease in the S DDM rings, "protease-shaving" experiments using proteinase K (PK) and trypsin were performed. These two proteases have distinguishable specificities, with the former cleaving at the carboxy end of aliphatic and aromatic residues and the latter cleaving at the carboxy end of lysine and arginine residues. The protease digestions of S proteins in IMVs and purified in DDM were analyzed by immunoblot, using antibodies that recognize either the N-terminus or

the C-terminus of S.

As shown in Figure 2.20, S105 in IMVs was cleaved by noth proteases at the C-terminus but also at the N-terminus by PK. Similarly, S105 in DDM was also cleaved by both proteases at the C-terminus but was protected from PK cleavage at the N-terminus. The dimer-blocked $S105_{a52v}$ in IMVs exhibited the same pattern as S105 in IMVs but produced a second, trypsin-derived fragment when solubilized in DDM. Interestigly, N-terminal cleavage by PK was observed in both IMVs and in DDM. $S105_{r59c}$ in IMVs was cleaved by both proteases at the C-terminus but to a lesser degree and PK cleaved the N-terminus similar to S105. Finally, $S105_{r59c}$ in DDM was cleaved by both proteases at the C-terminus and produced a second-trypsin derived fragment similar to $S105_{a52v}$ in DDM. Once again the N-terminus was protected from PK digestion in $S105_{r59c}$ in DDM.

Next, the effect of PK and trypsin digestion on the oligomerization properties of S105 in DDM, were examined. Digested products were prepared for EM by negative staining without further purification. EM of digested products of either protease indicated that S105 was still forming rings and were still able to stack (data not shown).

Overall, these results indicate that the C-terminus of S105 is susceptible to protease cleavage in both IMVs and DDM rings, suggesting a similarity in tertiary structure in the two environments. Furthermore EM indicates that the C-terminus is not required for oligomerization into rings, consistent with previous findings that truncations of the C-terminus of S105 do not prevent lysis but rather affect timing of lysis. Moreover

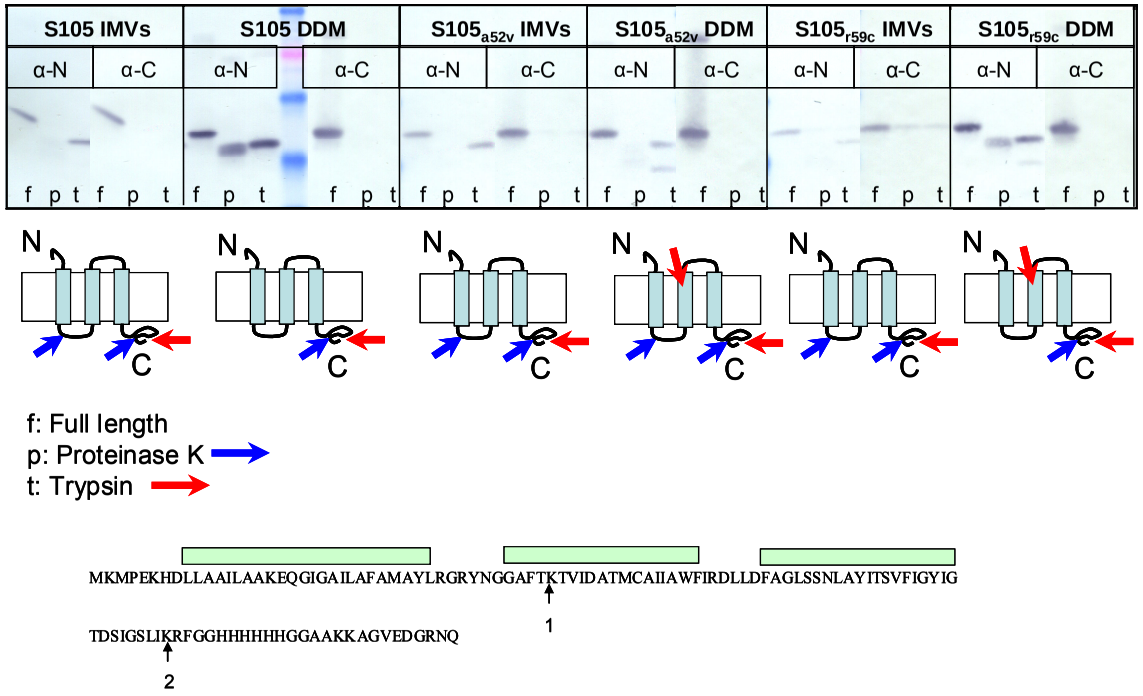

Figure 2.20 Protease digestion of S alleles
S in IMVs or purified in DDM was subjected to protease digestion. Predicted positions of the trypsin cleavage points are shown on the S sequence (see text).

these results indicate that the C-terminus of S105 does not participate in ring stacking in into rings, consistent with previous findings that truncations of the C-terminus of S105 do not prevent lysis but rather affect timing of lysis. Moreover these results indicate that the C-terminus of S105 does not participate in ring stacking in detergent solution. Finally although $S105_{r59c}$ rings appear identical to those formed by S105, the protease digestion experiments indicate some difference in helcal arrangement in the rings.

2.3. DISCUSSION

2.3.1 Purification of S in non-ionic detergent

The lambda S105 holin is the prototype class I holin. It is a small, integral membrane protein with three TMDs and an N-out, C-in topology. The biological function of S105 is to permeabilize the inner membrane of the bacterial host in order to allow the endolysin, R, access to the cell wall for degradation. This would require a hole of at least 4 nm in diameter and not surprisingly, crosslinking studies have shown that S assembles into higher oligomers *in vivo* in order to form such a hole. In the current study we have used a mild, non-ionic detergent to solubilize, purify and structurally and biochemically characterize the lambda holin S105.

Previous purification strategies for S involved the non-ionic detergent OG and the zwitterionic detergent EBB (64,72). S105 was found to unstable when solubilized by OG, shown by its rapid precipitation shortly after purification and by its reduced α-helical content. This raised difficulties for the biochemical analysis of S that was

subsequently transferred into chaotropic agents such as 6M guanidine-HCl. An improvement in the stability of S was achieved when EBB was used for solubilization and purification of S. EBB however, is a harsher detergent that is known to break protein-protein interactions. This property may account for the lack of precipitation and the stability of S in EBB. In this study, S purified in EBB was shown to form monomers and dimers but no higher oligomers could be seen by gel permeation chromatography.

An attempt to preserve the oligomeric state of S after solubilization and purification was made by using another widely used non-ionic detergent, DDM. Like OG, DDM, is a member of the glycoside detergent family, and has a 12 carbon acyl chain and a maltose head-group. In this study we observed that DDM solubilized S formed large macromolecular complexes. The gel-permeation elution profiles combined with TEM of the individual peaks showed that S105 assembled into multimeric ring assemblies and that these assemblies were allele-specific. The finding that the non-lytic mutant $S105_{a52v}$ does not oligomerize into rings under identical conditions seems to suggest that the oligomerization in the presence of detergent micelles is analogous to oligomerization in the presence of lipids and may therefore not be a phenomenon induced by detergent. DDM has been extensively used for the purification and structural characterization of other membrane proteins and has been found to preserve membrane protein oligomers in a functional and physiologically relevant form. Examples include the metal transporter YiiP, which was found to be a dimer both in DDM and when reconstituted in membranes (101), and the multidrug transporter EmrE, which forms a functional dimer in DDM (102).

2.3.2 Single particle analysis of S rings

Single particle analysis of the DDM rings indicated limited heterogeneity in ring diameter. Eigenimages showed characteristic concentric circles indicative of size variation amongst the particles. This heterogeneity proved to be a limiting factor in the resolution of subsequent 3D reconstructions but could be partially overcome by separating the raw data according to diameter and processing only particles of similar size. Although this method improved the resolution of 2D and 3D maps, it is likely that some heterogeneity still exists. Improvements to future reconstructions will require larger data sets that can be separated into more size groups without compromising the signal to noise ratio.

Due to the heterogeneity in ring diameter it is likely that rings with slightly smaller or lightly larger diameters will consist of fewer or more protomers respectively. In this study it was found that rings from the most abundant size group are composed of approximately 18 protomers as shown by imposing different rotational symmetry to end-on class averages of the DDM rings.

The 3D reconstruction of the dimeric ring indicates an assembly with an outer diameter of 23 nm and an average inner diameter of 8.5 nm. Ring stacking raises the question of whether the rings associate in a head to head or in a head to tail manner. Evidense that favors a head to tail association arises from several points. First, imposing C18 instead of D18 (an 18-fold rotational axis perpendicular to a 2-fold symmetry axis) symmetry allows for any asymmetry perpendicular to the 18-fold axis to be revealed and suggest head-to-tail stacking. Infact when a D18 reconstruction was calculated (which

would support a head to head association with 2-fold symmetry between the two rings), the resolution of the map decreased and structural features were reduced indicating that an incorrect symmetry had been imposed to the 3D reconstruction (data not shown). Secondly it is often observed that odd numbers of rings are seen stacking. Thirdly, head to head or tail to tail associations would require two types of interaction surfaces; one within the interface of two single rings in a ring dimer, and one between the individual ring dimers forming the longer stacks. Finally the distribution of charges in S105 is polar in character as S105 is overall negatively charged on the periplasmic side and positively charged on the cytoplasmic side. This would suggest that the rings are associating in a head to tail arrangement mediated by the electrostatic interactions from either side of S105. The relevance of the formation of ring dimers and higher order stacks remains to be elucidated but is unlikely to be relevant to the biological function of the complex.

Comparing the data from negatively stained samples with the cryo-EM data revealed a difference in the inner and outer ring diameters: the outer diameter decreased from 24.5 (negative stain) to 23 nm (cryo) and the inner diameter decreased from 11 (negative stain) to 9 nm (cryo). This type of size discrepancy has been previously reported for detergent- solubilized membrane protein complexes and is thought to be due to the stain exclusion by detergent surrounding the protein thus over-emphasizing the size of the complex in projection (103,104). The increase in the inner diameter can be explained by an accumulation of stain in the gap between the lower and the upper ring. Since this ring-ring interface is thought to be hydrophilic, stain may also bind to these parts of the structure, thus further exaggerating the size of the hole.

2.3.3 Structural features of the S ring

The height of the outer edge of both rings is approximately 4 nm and agrees well with the width of the inner membrane and the average length of a trans-membrane α-helix (105). The upper surface of each ring decreases in height towards the central hole, while the lower surface in each ring remains somewhat flat across its entire width. This slight asymmetry across the width of the ring suggests that the helices of S molecules lining the lumen of the ring might be orientated differently compared to the helices of S molecules on the periphery of the ring or that there may be different helices at different positions altogether. This may be due to the fact that helices facing the central hole would be in an aqueous environment whereas helices at the periphery would be in intimate contact with lipids. Furthermore this asymmetry is likely to be important for the intrinsic curvature that is present in the rings. For curvature to arise, the molecules of S lining the cavity might be more spatially constricted than the molecules on the periphery thus inducing this curvature. Alternatively there may be fewer molecules lining the cavity than there are lining the periphery of the ring.The assignment of tertiary structure features however, and the arrangement of S105 molecules within one protomer must await high-resolution studies.

There is good agreement between the molecular volume and the molecular mass of an 18mer if one protomer was to consist of four S monomers. Accordingly, each ring would consist of 72 S molecules with a molecular mass of 860 kDa. The estimated molecular mass is also in agreement with the gel permeation chromatography results, which indicated a complex significantly larger than 670 kDa. When the surface

rendering threshold is adjusted to account for a mass of 860 kDa per ring, the resulting 3D map is free of noise and the corresponding re-projections agree well with the class averages (Figure 2.7c). Assuming a diameter of 1 nm for an α-helix and an average shortest Cα-Cα inter-helix distance of 5.5 Å (as calculated from 16 known membrane protein crystal structures) (105), the cross-sectional surface area of the ring can accommodate approximately 62 molecules of S, which closely matches the calculation from the molecular volume.

The average inner diameter of the rings (8.5nm) is large enough to allow R to transverse the inner membrane but is not large enough to allow passage of the previously reported R-β galactosidase fusion. It has been previously estimated that during phage infection approximately 1000-3000 S molecules accumulate in the membrane (66,70). In terms of number of rings, this number would correspond to 13-40 rings, if all the S molecules were to be assembled in this manner. The finding that at low, sub-cmc detergent concentrations, S forms large assemblies of rings that resemble two-dimensional rafts suggests that this type of aggregation may occur in the membrane. These rafts of rings may coordinate triggering within the S population and explain why such a large number of S molecules is required, since an 8.5 nm hole can form from just ~70 S molecules.The lesion formed by such rafts would be in excess of 60 nm in diameter.

2.3.4 Rings formed by S107 and $S105_{r59c}$

The *S* gene is transcribed from the late promoter pR' and starts accumulating in the inner

membrane 8 minutes post infection. It continues to accumulate without any deleterious physiological effects on the host cell until it triggers leading to "hole" formation at a specific time determined by the primary sequence of the holin. The triggering is instantaneous and allows the anti-holin S107 to be converted into an effector holin. We have also purified the S107 anti-holin in DDM and confirmed that it forms rings identical to S105 (data not shown). This is to be expected since in the absence of a membrane potential, the extra positive charge of S107 should have no effect on the tertiary structure or helix topology. Interestingly the S mutant $S105_{r59c}$, which is deficient in lysis but has been shown to oligomerize, also formed rings when purified in DDM. This finding could suggest that the rings observed in DDM do not represent the hole that is formed *in vivo.* However, without high-resolution information on the molecular arrangement of molecules in the S105 versus the $S105_{r59c}$ rings a conclusion cannot be made. Furthermore protease protection experiments discussed below suggest that differences do exist between the rings formed by the two alleles.

2.3.5 Liposome reconstitution and the water soluble properties of S

The proteoliposome reconstitution results indicated that when S is reconstituted from EBB most of the protein is associated with the liposomes. Examination of these samples with EM however did not show any signs of S105 derived lesions, and no noticeable difference between the liposome only and $S105_{a52v}$ controls. This observation suggests that S105 failed to multimerize and create a lesion in the membrane. This may be related to the starting state of S105 in EBB which consists of only monomers and dimers that

may not assemble in the membrane under the conditions used. In contrast, reconstitution from DDM resulted in most of the protein not associating with the liposomes. Furthermore rings remarkably similar to the soluble S105 were seen with EM. This result is probably due to the ability of S105 to remain soluble in the absence of detergent. This ability reflects the intrinsic nature of the holin to form aqueous channels and the amphipathic nature of its TMDs. In order for S105 to remain soluble, this would require that the lipid contacting surfaces of S be shielded in these structures, whilst exposing the hydrophilic faces of the helices. The segmented morphology of the soluble rings suggests that they may form from the smaller, DDM solubilized-like rings that may have undergone a helical packing rearrangement in order to tolerate the hydrophilic environment. This further illustrates the dynamic flexibility required for S105 to proceed from a harmless membrane embedded state to one of membrane permibilization.

The finding that S105 purified directly in EBB requires reducing agent in order to form soluble rings whereas the cysteineless mutant $S105_{c51s}$ does not indicates that S is being oxidized during purification, probably during the metal affinity step. The resin bound cobalt or nickel is likely oxidizing the cysteines of S105, thus forming disulphide linked dimers that fail to assemble into higher oligomers.

2.3.6 Observing S in native membranes

To study the nature of the S105 complex in the membrane of *E. coli* an attempt to cross-link S in IMVs was performed. The chemical crosslinker DSP was used to crosslink S prior to detergent extraction and the purified crosslinked products were analyzed by gel

permeation and TEM. Gel permeation chromatography of the purified products indicated that in the presence of crosslinker, S105 could be captured forming higher oligomers when extracted in EBB and that the size of these higher oligomers was in the size-range of the rings. Furthermore the ratio of dimer to monomer increased compaired to un-crosslinked samples. As mentioned previously, the efficiency of crosslinking seemed to decrease for larger complexes as also indicated by Western blot analysis. Interestingly, no protein eluted in the void volume, indicating that larger complexes do not exist or are very inefficiently crosslinked. When S105 was crosslinked prior to extraction with DDM, the elution profile closely matched that of DDM solubilized, un-crossliked S105. A broadening of the ring peak was observed along with an overall shift of the protein population towards the ring peak. TEM of these fractions indicated that S105 was still forming rings but many rings were open or incomplete. These species may account for the broadening of the ring peak since open rings increases their Stokes radius, and incomplete rings are smaller. The lack of stacks in crosslinked samples may be a result of the loss of flexibility from crosslinking thus hindering the docking of individual rings. Furthermore, the many open or incomplete rings present would also hinder stack formation. The presence of opened or incomplete rings may suggest that not all the S105 had oligomerized completely in the membrane prior to crosslinking and extraction. Once crosslinked these incomplete rings may have lost their ability to add on new protomers due to intermolecular restraints caused by the crosslinker, i.e. the subunits are not in the correct conformation and act as terminators for further elongation. It is also possible that complete rings may have fallen apart, again from introducing new bonds that add stress

to the rings. Overall, the DSP crosslinking results show that even in the presence of a harsh zwitterionic detergent, complexes of similar size to the S105 DDM rings can be captured in the membrane prior to detergent extraction. Furthermore even when crosslinked prior to DDM extraction, S105 retains its ability to form rings and structures that resemble rings, despite the fact that the monomers have been crosslinked with high efficiency into dimers and possibly tetramers (as shown by the EBB extracted crosslinked products).

Enrichment of IMVs from *E. coli* cells overexpressing *S* alleles using a metal affinity resin was also carried out to try and observe S in a native environment. EM of these samples showed strikingly different results for S105, $S105_{a52v}$ and $S_{\Delta TM1}$. S105 enriched samples contained structures that resembled S105 soluble rings, both in shape and size. This finding was unexpected and subsequent flotation gradient centrifugation of S105 IMVs indicated that S105 does not localize to the soluble protein fraction but floats with the membrane vesicles under both near-native expression levels (transcription from pS105 with Q provided from the λΔSR lysogen) and overexpression conditions. On the other hand purified soluble S105 remained in the high density soluble fraction in the same experiment. This results confirmed earlier isopycnic sucrose gradients that showed that most of S105 was located in the inner membrane fraction (60). Interestingly, it was also mentioned in the same study that sonication of the IMVs cause some of the S to partition in the soluble protein fraction, although no further studies were performed. It is still unclear how these soluble ring-like structures were present in the enriched IMV samples. One possibility is that S105 in these rings is associated with lipid, thus reducing

the buoyant density of the complex and allowing it to float in the gradients. Finaly because S is expected to be in a triggered form under the current experimental conditions, the soluble rings observed may suggest that part of the triggered S population might form such complexes. The exact role of these structures however, requires further investigation.

2.3.7 Biochemical characterization of S in membranes and detergent

The 3D structure of the S105 DDM rings at 2.6 nm does not provide information as to the arrangement of the S molecules within the ring complex. Higher resolution cryo-EM studies will need to be performed in order to pass the 1.0 nm barrier where α-helices can be discerned. In order to investigate the possible arrangement S was conjugated to cysteine-reactive nanogold particles of 1.4nm size. An initial accessibility study for reactive cysteines that are exposed in the DDM rings indicated that only the N terminus could be used. $S105_{h7c}$ could be modified with PEG-maleimide, but C51 in S105 as well as a C terminal cysteine placed at position 99 (A99C) could not be labeled. These results suggest consistency between S in IMVs (100) and purified ring structures. The N-terminus which is exposed *in vivo* in the periplasmic space is also accessible in the DDM rings, whilst C51 is either buried inside the helical contacts or protected from pegylation, due to the detergent molecules surrounding the hydrophobic core of S105. The former is most likely since the SDS molecules allow this cysteine to be labeled by breaking the intermolecular interactions. This result suggests that C51 is shielded by intermolecular helix packing between individual S105 molecules in the DDM rings. This would agree

with the previous finding that C51 cbe oxidized in the membrane to form disulphide linked dimers (63). Moreover the efficiency of this oxidative crosslinking was specific to a certain face of TMD 2 that is thought to form the S-S dimer interface *in vivo*. Surprisingly C99, near the C-terminus, was inaccessible even after SDS addition. It was speculated earlier that the basic residues in this region might be interacting with the membrane and shielding the region from modifying agents. However with the experiments reported here, this region seems resistant to modification even in the absence of the lipid bilayer. It is likely that strong salt bridges between the basic and acidic residues at the C-terminus may be creating a rigid, SDS resistant fold.

The accessibility of position 7 to pegylation allowed us to probe its localization within the rings using monomaleimido nanogold reagent. From the data collected from top views of labeled rings the nanogold was predominantly found on the outer surface of the rings. This indicates that the N-terminus is exposed on the periphery of the rings. However it cannot be ruled out that the N termini of S also reside at other positions that may be inaccessible to the nanogold. The quartenary arrangement of S molecules in the protomers might only expose the N-termini of half of the molecules, while the other half might be buried in protein contacts within the protomer.

The protease accessibility experiments performed on S alleles in IMVs and on purified S in DDM have provided more information on the subunit organization in the rings. It was found that the C-terminus was accessible to both PK and trypsin in S105, $S105_{a52v}$ and $S105_{r59c}$ IMVs. Cleavage is estimated to be at position K92 or R93. These results confirm previous studies that localized the C-terminus of S in the cytoplasm,

which corresponds to the exposed surface of IMVs. Similarly, the C-terminus of all three alleles was also accessible in the DDM purified proteins indicating that it is not buried in the detergent micelle or by the intermolecular interactions. EM examination of DDM S105 digestion products revealed that the rings persist and even form stacks despite the lack of the C-terminus, and thus it must not be required for oligomerization of S. This finding supports earlier data that C-terminal truncations of S still confer lysis *in vivo* and thus can still oligomerize to form the hole. The C-terminus is thought to regulate lysis timing as truncations or residue alterations to this domain can slow down or accelerate lysis.

Interestingly the N-terminus was also digested by protease K in IMVs for all three alleles. In inverted membrane vesicles the N-terminus should reside on the opposite side of the bilayer. The result that the N-termini of $S105_{a52v}$ and $S105_{r59c}$ in IMVs are also cleaved rules out access of the protease through an S derived hole. Thus it is most likely that PK is cleaving the N terminus at the first loop that would be exposed in the IMVs. This fragment ultimately runs off the gel. The fact that the N-terminus of $S105_{a52v}$ in DDM is also accessible to PK but the N-termini of S105 and $S105_{r59c}$ in DDM are protected, suggests that this protection may arise from ring stacking that shields the N-terminus.

Finally, because a second N-terminal trypsin fragment is obtained for $S105_{r59c}$ but not for S105, a difference in subunit arrangement that is not visible in raw electron micrographs exists. The same fragment is obtainable in $S105_{a52v}$ in DDM. Both $S105_{a52v}$ and $S105_{r59c}$ fail to trigger *in vivo*. $S105_{a52v}$ displays an oligomerization defect but

$S105_{r59c}$ oligomerizes but fails to trigger even after inducing a membrane potential collapse (63). Since again this second fragment is not due to an inability to oligomerize it might indicate a conformation that is different in the two mutants but is absent in the wild type protein. The position of cleavage in this case is not as clear. The calculated position is near R59 but this residue is absent in R59C. Although calibration of molecular weight standards gives an accurate linear trendline (R^2=0.98) S, like other membrane proteins, has irregular electrophoretic mobility and estimation of fragments sizes might not be accurate. The other possible cleavage position would be at K41 that resides inside the 2nd TMD. An accurate assessment of fragment size will require other methods like mass spectrometry.

The results above show similarities of S in the membrane and micelle environments. This can be also viewed as surprising considering the difference of these two environments. Although both lipids and detergents act to shield the hydrophobic TMDs, the micelle should theoretically offer a larger surface of accessibility to the proteases. To some extent this is seen by the 2nd N terminal fragment of $S105_{a52v}$ and $S105_{r59c}$ in DDM. One could imagine that the 2nd loop would also be cleaved or that residues inside the rings might also be accessible. However, no other extra proteolysis fragments were observed from detergent solubilized S.

Our study of S in the detergent DDM has also opened a new path towards achieving high resolution structural data. During this doctorate study we have observed that buffer conditions that produce monodisperse samples as observed with EM favor crystallization conditions and vice versa (SecA, chapter III). The stability and good

dispersion seen for S in DDM suggested that these may represent favorable conditions for crystallographic studies. In collaboration with A. Ramesh in the laboratory of Dr. J. Sacchettini we have attempted crystallization of S alleles in DDM. Although the heterogeneity of the wild type protein with respect to ring diameter has hindered crystal growth, we have been able to grow crystals of DDM-solubilized $S_{\Delta TM1}$ and $S105_{a52v}$. Crystals of the former protein are reproducible and have diffracted to ~7Å. Optimization trials for obtaining higher resolution diffraction quality crystals are currently in progress. The crystallization results are quite promising, considering the difficulty and statistics governing membrane protein structural genomics. An atomic resolution structure of S will provide a wealth of information on the molecular mechanism of lysis and will set new foundations for further biochemical and biophysical characterization.

2.3.8 A model for S induced lysis

Based on the results obtained in this study and previous work a model for S mediated lysis can be hypothesized (Figure 2.21). In the wild-type lambda life cycle, the products of the *S* gene start to accumulate for 40 minutes before triggering. As the concentration of S increases in the membrane, dimers and tetramers will accumulate giving rise to the tetrameric protomers of the S derived hole. S107 molecules may function by interacting with S105 monomers to prevent S105 dimer and protomer formation. The asymmetric shape of the protomers gives rise to a curved multimer. The further assembly of protomers into an arc-shaped polymer continues in a time and concentration dependent manner until the arc closes into a ring. Prior to ring completion the surfaces surrounding

the arc shaped multimers should still be hydrophobic and interacting with lipids. Evidence for this comes from EM studies of $S105_{r59c}$. Premature triggering by energy poisons prior to complete assembly could result in an instant availability of S107:S105 heterodimers to participate in accelerated ring completion. Approximately 13-40 rings could form in the membrane which may then associate to form rafts of rings as observed by the "honeycomb" structures formed by S105 at low detergent concentrations. This raft of rings may act to coordinate hole formation from each ring and may be important in timing the onset of triggering. A collapse of the membrane potential provides the molecular switch for the amphipathic helices to rearrange within the ring and to expose the hydrophilic surfaces to the lipid thus causing a thermodynamically unfavorable intermediate that escalates to the dissociation of the protein lipid interaction and formation of a hole. The hydrophilic exposure could occur in the inner or outer surfaces, or alternatively in both surfaces. In one model lipid is extruded from the cavity of each ring leaving a hole that is now stabilized by hydrophilic lining of protein, similar to the holes induced by the pore forming toxins (26). In another scenario, some of the rings in the raft may dissociate completely from the raft leaving a much larger hole which would again be lined by S that originated from the outer-most molecules in the raft. The rings that dissociate from the raft could still be associated with the lipids they enclosed. Two points of support for the second model arise from the observation that purified S105 can remain water soluble *in vitro* and that S105 enriched IMV fractions contained soluble ring-like structures that may be associated with some membrane lipid. These structures may be indicative of the final triggered form of S since cells harvested for

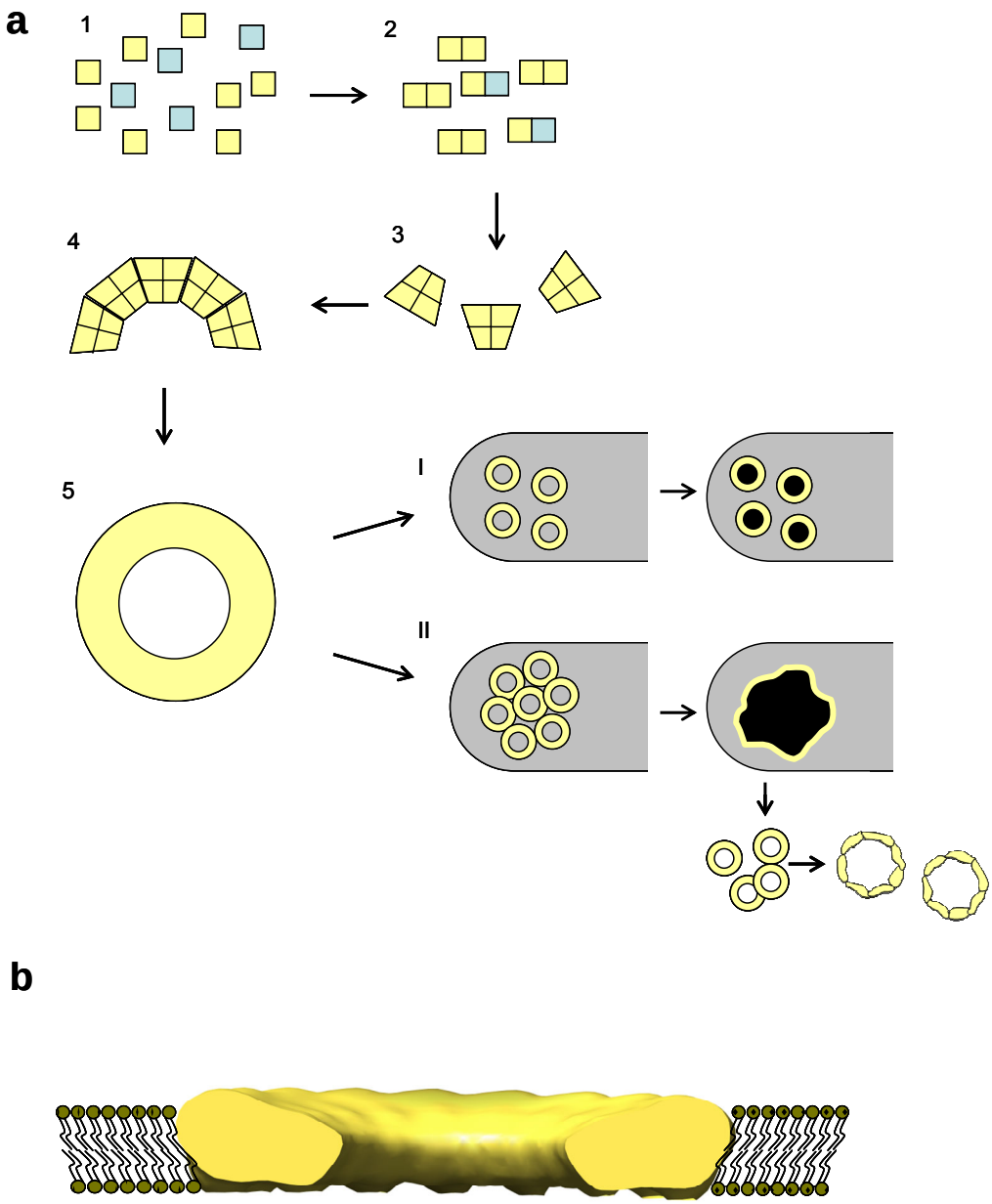

Figure 2.21 Model for S mediated lysis
Possible steps in S mediated lysis (a). Yellow and blue squares denote S105 and S107 respectively. Putative location of the S ring assembly in the lipid bilayer (b).

characterization no longer have an active membrane potential.

The switch for membrane depolarization may be S derived, and arise only once a ring has been completed and all the protomers are in the correct conformation to participate simultaneously in the switch. $S105_{r59c}$ rings might be blocked at coordinating this switch. An example of subunit coordination in ring structures are the cholesterol dependent cytolysins that oligomerize on the surface of the membrane in a pre-pore state prior to the insertion of their β-hairpins into the bilayer (106). It is suggested that pre-formation of the β-barrel from the individual β-hairpins prior to insertion would be energetically favorable over individual subunit insertion due to the hydrogen bonding fulfillment of each strand. A more plausible switch which depends on the coordination of all the rings may arise from raft formation. One the raft of rings has expanded to a certain size spontaneous holes, possibly in the space between rings, may form causing a local collapse of the membrane potential and triggering. The fast lysing mutant $S105_{a52g,}$ could be proficient at forming rafts of rings that can destabilize when the rafts are sufficiently smaller than those formed by S105. This would explain how the R-LacZ fusion cannot escape from holes formed by premature triggering of S105 but can exit from holes made by $S105_{a52g.}$ The holed formed by premature triggering would be from individual rings not assembled in rafts and would lead to holes only large enough to allow R exit.

2.4 EXPERIMENTAL PROCEDURES

2.4.1 S105 overexpression

S105τ94 which has the sequence for $G_2H_6G_2$ embedded between amino acid positions 94-95 (72) was cloned into the pET30b vector (Novagen) and overexpressed in *E.coli* C43(DE3). All *S* allele products purified for this study contained the $G_2H_6G_2$ sequence at position 94. From a fresh transformation plate, several colonies were selected for a 60 ml starter culture at that was grown at 37 ºC. At an OD_{550} of 0.6, 10 ml of culture was used to inoculate 1 liter of LB supplemented with 50µg/ml kanamycin. Typically 6 liter cultures were grown to OD_{550}= 0.7 and induced with 1 mM IPTG for a further 50 minutes at 37 ºC. Cells were then harvested by centrifugation in a Sorvall JA-10 rotor for 15 minutes at 8,000 rpm and at 4 ºC. Cell pellets were then frozen at -20 °C.

2.4.2 S purification

Cell pellets were resuspended in lysis buffer (10 mM Tris pH 7.9, 150 mM NaCl, 1mM DTT, 1mM PMSF) and lysed by French press. Lysates were cleared of un-lysed cells by centrifugation at 10,000 rcf in a Sorvall SS-34 rotor. The supernatants were then spun for 90 minutes at 100,000 x g and at 4 ºC in a Ti 50.2 Beckman ultracentrifuge rotor to harvest inner and outer membranes. Total membrane pellets were then homogenized in extraction buffer (10 mM Tris pH 7.9, 150 mM NaCl, 1% DDM or EBB) and extracted for 2 hours with gentle shaking at 4°C. Unsolubilized material was removed by ultracentrifugation for 1 hour at 100,000 x g and at 4 ºC. Typically solubilized

membrane material from a 2 liter culture was applied to 250µl of a TALON metal affinity resin in a 10 ml disposable column that had been equilibrated with 4 bed volumes of buffer A (10 mM Tris pH 7.0, 150 mM NaCl, 0.1% (w/v) DDM or 1% (v/v) EBB). The column was then washed with 20 bed volumes of buffer A and non-specifically bound proteins were removed by addition of 4 bed volumes of buffer A plus 20mM imidazole followed by 4 bed volumes of buffer A plus 40Mm imidazole. Finally S was eluted in 4 bed volumes of buffer A plus 500mM imidazole. No adjustment to the pH of buffer A was performed after addiditon of imidazole.

2.4.3 S purification in EBB

Cell pellets were re-suspended in lysis buffer (100 mM Na_2HPO_4, 100mM KCl, 1mM DTT, 1mM PMSF) and lysed by French-press. Lysates where cleared of unlysed cells by centrifugation for 15 minutes at 10,000 x g and at 4 ºC in a Sorvall SS-34 rotor. The supernatants were then spun for 90 minutes at 100,000 x g and at 4 ºC in a Ti 50.2 Beckman ultracentrifuge rotor to harvest inner and outer membranes. Membrane pellets were then homogenized in extraction buffer (20 mM BES pH 7.9, 500 mM NaCl, 10 mM $MgCl_2$, 1% (v/v) EBB, 10% (v/v) glycerol and 1mM PMSF) in a 40ml capacity glass homogenizer and left overnight at 4°C with gentle shaking. Unsolubilized material was removed by ultracentrifugation. The supernatant was then loaded onto a 5 ml His-trap HP column (Amersham) and further purification was performed on an AKTA FPLC. Briefly after sample loading, the column was washed with 20 bed volumes of buffer A (20 mM BES pH 7.9, 500 mM NaCl, 1% (v/v) EBB, 10% (v/v) glycerol) to

remove non-specifically bound proteins. The column was then equilibrated with a low salt buffer B (20 mM BES pH 7.9, 1% (v/v) EBB) and bound protein was eluted using 2 bed volumes of buffer C (20 mM BES pH 7.9, 500 mM imidazole, 1% (v/v) EBB). The eluted material was pooled and loaded onto a 5 ml anion-exchange column (Q-trap HP, Amersham) equilibrated with buffer B. Because S is positively charged at pH 7.9 (theoretical PI=9.2) it elutes in the flow through in a pure form.

2.4.4 Gel permeation chromatography

Gel permeation chromatography was carried out using a Superose-6 column (fractionation range of 5×10^3-5×10^6 Da, Pharmacia). Experiments were carried out on the AKTA FPLC. Purified S samples (100-200µl of 0.5-1 mg/ml) were injected and separated at a flow rate of 0.5 ml/min. Calibration standards were purchased from Biorad and used according to the manufacturer.

2.4.5 Negative stain electron microscopy

In most cases negatively stained specimens were prepared according to Valentine *et al.* (107). Carbon coated mica was floated onto 60-100 µl sample droplets on parafilm for 60 seconds followed by floating onto a 150 µl aqueous solution of aqueous uranyl acetate (2% w/v) for 10 seconds. In cases where specimens contained detergent the samples were placed in the inverted caps of PCR tubes to avoid sample spreading on the parafilm. The carbon pieces were then picked up with 400 square mesh copper grids (Electron microscopy sciences, G-400-Cu) and blotted dry with Watman No1 filter

paper. Electron micrographs were recorded in a JEOL 1200-EX Transmission electron microscope (TEM) operating at an acceleration voltage of 100 kV at a calibrated magnification of 24,038 x. Micrographs were digitized using the Leafscan 45 microdensitometer using a 10 μm scan step corresponding to 4.16 Å/pixel on the specimen level.

2.4.6 Cryo-electron microscopy

S purified in DDM at a concentration of 0.5 mg/ml was applied to C-Flat holey carbon grids that had been freshly glow-discharged for 60 seconds. Vitrification in liquid ethane was carried out in an FEI Vitrobot at 100% humidity with 3 second blotting. Specimens were observed in a JEOL 2200 FS microscope using a Gatan 626 cryo-specimen holder. Electron micrographs were acquired using a Tietz TVIPS F224HD CCD camera under lose doses (< 10 e-/$Å^2$). The final pixel size was 4.2 Å/pixel.

2.4.7 Single particle analysis

Particles from electron micrographs of negatively stained and cryo-specimens were selected interactively using the EMAN boxer software. Datasets were then processed using the IMAGIC-5 software package (40) to produce reference free classes using only the first 5 eigenvectors. For the cryo data, 1836 ring dimers in end-on, side-on and intermediate projections were selected. After reference-free classification 1079 particles from classes with a 23 nm outer diameter were extracted and processed separately. Further processing was performed using the EMAN single particle analysis package (9)

and a 3D reconstruction assuming C18 rotational symmetry was obtained. An initial 3D starting model was generated using the "startcsym" command. The model was subsequently used for refinement through projection matching. The resolution of the map after 8 iterations of refinement was 2.6 nm using the Fourier Shell Correlation (FSC) 0.5 criterion (8). The final map was low-pass filtered to 2.6 nm and surface rendered using UCSF Chimera (108). The rendering threshold was adjusted to enclose protein densities of 1.72 MDa assuming a protein density of 0.81 Da/Å^3 (109).

2.4.8 Preparation of water-soluble S and circular dichroism spectroscopy

For comparison of the secondary structure of water-soluble, DDM and EBB-solubilized S, water-soluble S was prepared by extensive dialysis of 1 ml of EBB purified S (50-70μM) against 1 liter of PBS dialysis buffer (20 mM sodium phosphate pH 7.8, 150 mM NaCl, 1 mM DTT) in the presence of 1ml of Calbiosorb beads for a minimum of 5 days with daily buffer exchanges at 4ºC. DDM or EBB was added to S in PBS (10-20μM) to a final detergent concentration of 0.1% (w/v) and 1% (v/v) respectively and left to incubate for 30 min at room temperature. Samples were subsequently spun for 10 minutes at 16,000 x g and at 4 ºC to remove any precipitated material. Protein concentrations were estimated using the absorbance at 280 nm. CD measurements were carried out in an Aviv 62DS CD spectrometer using a 0.1cm pathlength cuvette with wavelength scans from 250-200nm.

2.4.9 Enrichment of inverted membrane vesicles containing S

C43(DE3) cells over-expressing *S* alleles were grown, induced and harvested as before. Cells were French-pressed and cell debris was removed by centrifugation at 10,000 xg for 15 minutes in a Sorval SS-34 type rotor. Membranes from such preparations, which have been shown to be predominantly inside-out (110), were collected by centrifugation at 100,000 x g for 90 minutes and at 4 ºC in a Beckman Ti 50.2 rotor, resuspended in TBS buffer (10 mM Tris 7.9, 150 mM NaCl) and homogenized in a glass homogenizer. Typically membranes from a 100 ml culture of an OD_{550} 1.2, were then passed over 100 µl of a TALON metal affinity resin equilibrated with TBS, followed by washing with 1-2 ml TBS, and eluted (500 µl of 10 mM Tris pH 7.9, 150 mM NaCl, 150 mM imidazole) in 100 µl fractions.

2.4.10 Membrane flotation

Cultures (200 ml) of MC4100 λΔSR pS105 and C43(DE3) S105/pET30b were induced at OD_{550}=0.3 and grown for a further 50 minutes. Cells were harvested and re-suspended in 1 ml of resuspension buffer (10 mM Tris pH 7.8, 0.75 M sucrose) and supplemented with 30 µl of 20 mg/ml hen egg white lysozyme. After 5 minutes incubation on ice, 2 ml of 1.5 mM EDTA were added sequentially at a rate of 1ml/min under constant stirring/mixing at 4 ºC. The resulting spheroplasts were then sonicated 3 times (15 seconds at a time), and lysates were cleared of cell debris by a low spin centrifugation. Flotations were essentially carried out according to (111) with modification. Iodixanol solutions were buffered with TBS-EDTA (10 mM Tris 7.9, 150 mM NaCl, 5 mM

EDTA). Lysates (40 µl) were mixed with 160 µl of 55% (w/v) iodixanol resulting in a 44% (w/v) solution that was added to the bottom of a TLS-55 type rotor centrifuge tube. This was overlaid with 800 µl of 30% (w/v) iodixanol and finally 300 µl of buffer (TBS-EDTA) were added onto the top. Samples were centrifuged for 3 hours at 160,000 x g at 4 ºC. Fractions (186 µl) were collected from the top and analyzed by SDS-PAGE.

2.4.11 Liposome reconstitution

S was purified either in EBB or in DDM as described. Ethanolic solutions of DOPG (10 mg/ml) and egg PC (20 mg/ml) were mixed at a final weight to weight ratio of 3:7. A total of 3.85 mg of lipid was dried under nitrogen and left in a desiccator for one hour to evaporate residual ethanol. Dried lipid was then resuspended and vortexed in TBS buffer (10 mM tris pH 7.9, 150 mM NaCl). Once resuspended, EBB or DDM was added to final concentrations of 1% and 0.1% respectively. Detergent solubilized lipids were mixed with protein at the appropriate molar ratio and left overnight at 4 ºC. The next day samples (100-500 µl) were injected into slide-a-lyzer dialysis bags (MWCO 3.5 kDa) and dialyzed against TBS supplemented with 1mM DTT and 1ml of Calbiosorb beads. Dialysis was carried out for 7-10 days at 4ºC with daily buffer exchange.

2.4.12 Liposome flotation assays

Proteoliposomes were floated according to (112). Dialyzed liposome samples (200 µl) were mixed 1:1 with TBS, 60% (w/v) sucrose and overlaid in a 2.2 ml TLS-55 tube over a 300 µl TBS, 80% (w/v) sucrose cushion. The liposome-TBS, 30% (w/v) sucrose

mixture was subsequently overlaid with 800 µl TBS, 20% (w/v) sucrose and 400 µl of TBS buffer. Samples were spun for 1 hour at 49,000 rpm and at 4 ºC in a TLS-55 rotor. Fractions (200 µl) were collected from the top of the gradient with a pipette.

2.4.13 Cysteine pegylation and nanogold labeling

S105, $S105_{h7c}$ and S_{a99c} were purified using a TALON metal affinity resin as before. To 10 µl of protein (0.026 mg/ml in 500 mM Tris pH 7.0, 1 mM EDTA, 0.1% w/v DDM), 1 µl of 10 % (w/v) SDS was added or 1 µl of H_20 in the samples without SDS. PEG-maleimide (600 µM) was added (1 µl) to a final concentration of 50 µM and left for 30 minutes at 25 ºC. Samples were then supplemented with 1 µl of 10 mM DTT to quench the reaction and mixed with 7 µl of 2X reducing sample loading buffer. Samples were analyzed by SDS-PAGE on a 16.5% Tris-Tricine gel.

For nanogold labeling, purified $S105_{c51s\ h7c}$ was reduced using a final concentration of 1 mM DTT and left to incubate for 1 hour on ice. Reduced $S105_{c51s\ h7c}$ was then passed over a superose 6 column equilibrated with 20 mM sodium phosphate (pH 6.5) and 150 mM NaCl to remove the DTT. The second peak (rings and small stacks) was collected for nanogold labeling. For the reaction, 6 nmoles of lyophilized mono-maleimido nanogold was resuspended in 1 ml H_2O and 500 µl (3 nmoles) were mixed with 500 µl of $S105_{c51s\ h7c}$ (2.8 µM, 1.4 nmoles total) and left overnight to incubate with gentle shaking at 4 ºC. Excess, unreacted nanogold was removed by ultrafiltration using an Amicon Centriprep concentrator with a molecular mass cutoff of 30 kDa. The reaction mixture was concentrated to 100 µl and then diluted to 2 ml with

gel permeation buffer. This process was repeated 5 times. Samples were prepared for TEM using the method from Valentine *et al.* and negatively stained with either 2% (w/v) ammonium molybdate or methylamine tungstate (Nanovan) stain (107).

2.4.14 Chemical crosslinking of S in native *E. coli* membrane vesicles

S was over-expressed in C43(DE3) cells and membranes from 1 liter of culture were collected as before. Total membranes were resuspended in 20 ml of HBS buffer (50 mM HEPES pH 7.4, 150 mM NaCl) and homogenized using a glass homogenizer. The membranes were then split into 2 x 10 ml samples and 2 ml of 6 mM DSP in 100% (v/v) DMSO or 2 ml of 100% (v/v) DMSO were added to each sample and mixed. The crosslinking reaction was left for 30 minutes at room temperature while gently shaking. The reaction was stopped by addition of 1.2 ml of 1M glycine. DDM was subsequently added to a final concentration of 1% (w/v). The extraction was carried out for 2 hours at 4 ºC. Solubilized membranes were cleared from large membrane fragments by filtration through a 0.45µm followed by a 0.22 µm nitrocellulose syringe filter to avoid clogging of the filters of the metal affinity resin. Filtered material was then TALON-purified as before on 250 µl of resin equilibrated with HBS-DDM buffer (50 mM HEPES pH 7.4, 150 mM NaCl, 0.1% (w/v) DDM) and eluted in HBS-DDM supplemented with 500 mM imidazole. Purified samples were then subjected to gel permeation chromatography using a superose 6 column equilibrated with HBS-DDM buffer at a flow rate of 0.5 ml/min.

2.4.15 Protease shaving

S IMVs and purified S samples were prepared as before. For purified samples, 40 μg of protein in protease-DDM buffer (20 mM tris pH 7.9, 150 mM NaCl, 5 mM $CaCl_2$ and 0.1% (w/v) DDM) was incubated with 2 μg of protease. For IMVs, a total membrane protein amount of 40 μg in protease buffer (20 mM Tris pH 7.9, 150 mM NaCl, 5 mM $CaCl_2$) were incubated with 2 μg of protease. Digestions were carried out for 30 minutes at 37 ºC and then terminated by addition of PMSF to a final concentration of 1 μM.

CHAPTER III

STRUCTURAL STUDIES OF *M. tuberculosis* SecA TRANSLOCASE

3.1 INTRODUCTION

Protein translocation is an essential process in all organisms. This process allows the export of secreted proteins across the membrane and the integration of membrane proteins into the lipid bilayer. Newly synthesized proteins are translocated through the ubiquitous Sec translocation system. In eubacteria the core components of this system are the cytoplasmic ATPase SecA and the membrane protein conducting channel SecYEG. Structural characterization of these components could improve our understanding of the translocation process. This chapter describes structural studies of the mycobacterial SecA translocase by electron microscopy and single particle analysis and also describes the preliminary studies of the membrane embedded Sec components.

3.2 RESULTS

3.2.1 Single particle EM of *tb* SecA

To study the structure and oligomeric state of *tb* SecA in solution, purified protein was negatively stained and observed in the transmission electron microscope. Initial screening of SecA in the buffer used for purification resulted in small aggregates with no

single molecules of SecA observed. The lack of monodispersity prompted the screening of other conditions that might favor single particles. Variations in the pH and salt concentration as well as the addition of other ions including Mg^{2+} and Ca^{2+} were performed but did not improve the dispersity of SecA on the carbon support film. Collaborative efforts to obtain 3D crystals of *tb* SecA helped identify conditions that favored crystallization. The growth of 3D crystals requires that the molecules arrange in an organized manner which in turn requires the solution to be monodisperse, since random aggregates cannot contribute but rather hinder crystal formation. One such condition contained 3M sodium formate. Replication of this buffer condition for EM of *tb* SecA resulted in significant improvement of particle dispersity (Figure 3.1a). Further improvement was obtained by increasing the concentration of sodium formate to 4M (Figure 3.1b) and this condition was subsequently used for obtaining the *tb* SecA EM dataset.

Previous studies had shown that *E. coli* SecA exists as a dimer in solution (113,114). This symmetry was initially used to perform a 3D reconstruction of *tb* SecA from approximately 4,400 single particles acquired in the above condition. However, reconstructions with C2 rotational symmetry (one axis of 2-fold symmetry) resulted in maps, which upon closer inspection and back-projection did not resemble the class averages, indicative of an incorrect model. Furthermore the surface rendering threshold for a 200 kDa volume (a dimer of two 100 kDa molecules) resulted in a discontinuity of densities indicating that the volume enclosed a larger protein mass. Thus C4 (4-fold rotational symmetry) and D2 (dihedral symmetry) symmetries were imposed on the 3D

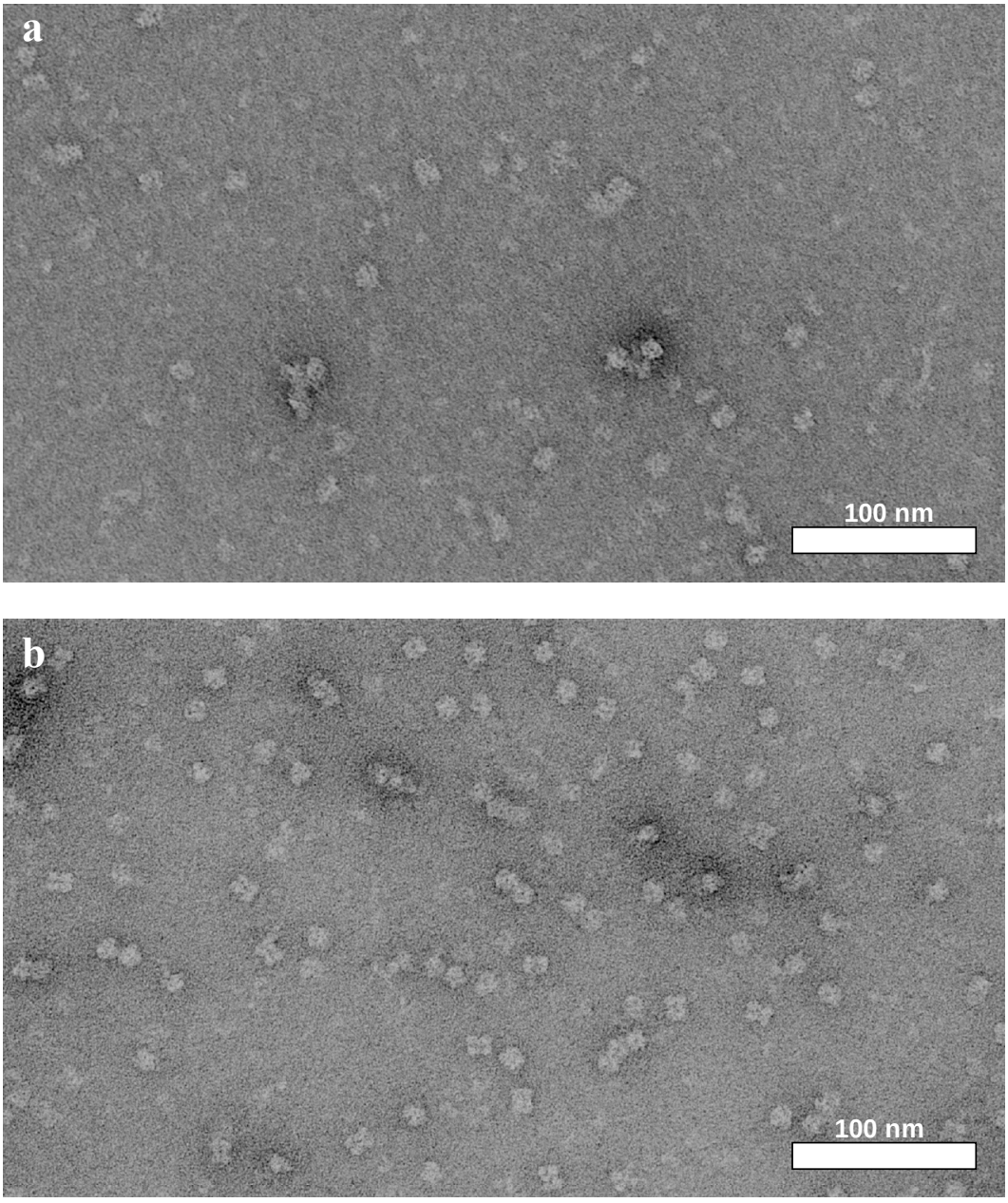

Figure 3.1 Electron micrographs of *tb* SecA
SecA at a concentration of 30 μg/ml in buffer supplemented with 3M (a) or 4M (b) sodium formate. The latter condition gave rise to improved monodispersity on the carbon support film and a good distribution of orientations.

reconstruction. Imposing D2 symmetry resulted in a 3D map that not only agreed with the class averages but also contained protein densities corresponding to a mass of approximately 450 kDa, which is in excellent agreement with the estimated 400 kDa of a SecA tetramer. The D2 symmetry was subsequently used for refinement of the 3D reconstruction of SecA to a final resolution of 23 Å as judged by the FSC 0.5 (Fourier Shell Correlation) criterion.

The 3D reconstruction of the tetrameric *tb* SecA shows a slightly elongated structure measuring 140 x 110 x 90 Å (Figure 3.2). Although the resolution of the reconstruction does not allow for determination of individual monomer boundaries, the symmetry restrictions allowed for some interpretation of subunit arrangement. A D2 dihedral symmetry dictates that a total of three dimeric interfaces exist (115). Two of these interfaces are formed by the monomers in each dimer and the third interface is formed by the association of the two dimers. In the 3D map of SecA two large cavities exist, which coincide with the centers of the 2-fold axes. A smaller cavity that was observed in the class averages but obscured in the 3D reconstruction lies at the center of the third 2-fold axis. Two of the three axes of 2-fold symmetry should pass through the cavities that are created between the monomers of SecA. A third cavity is created as a result of the association between the two dimers. Without knowledge of the monomer boundaries however, the exact allocation of these interfaces cannot be made. The only strong indication for identifying one of the 2 dimers was the one with the small cavity since most of the surface was buried in protein-protein interactions.

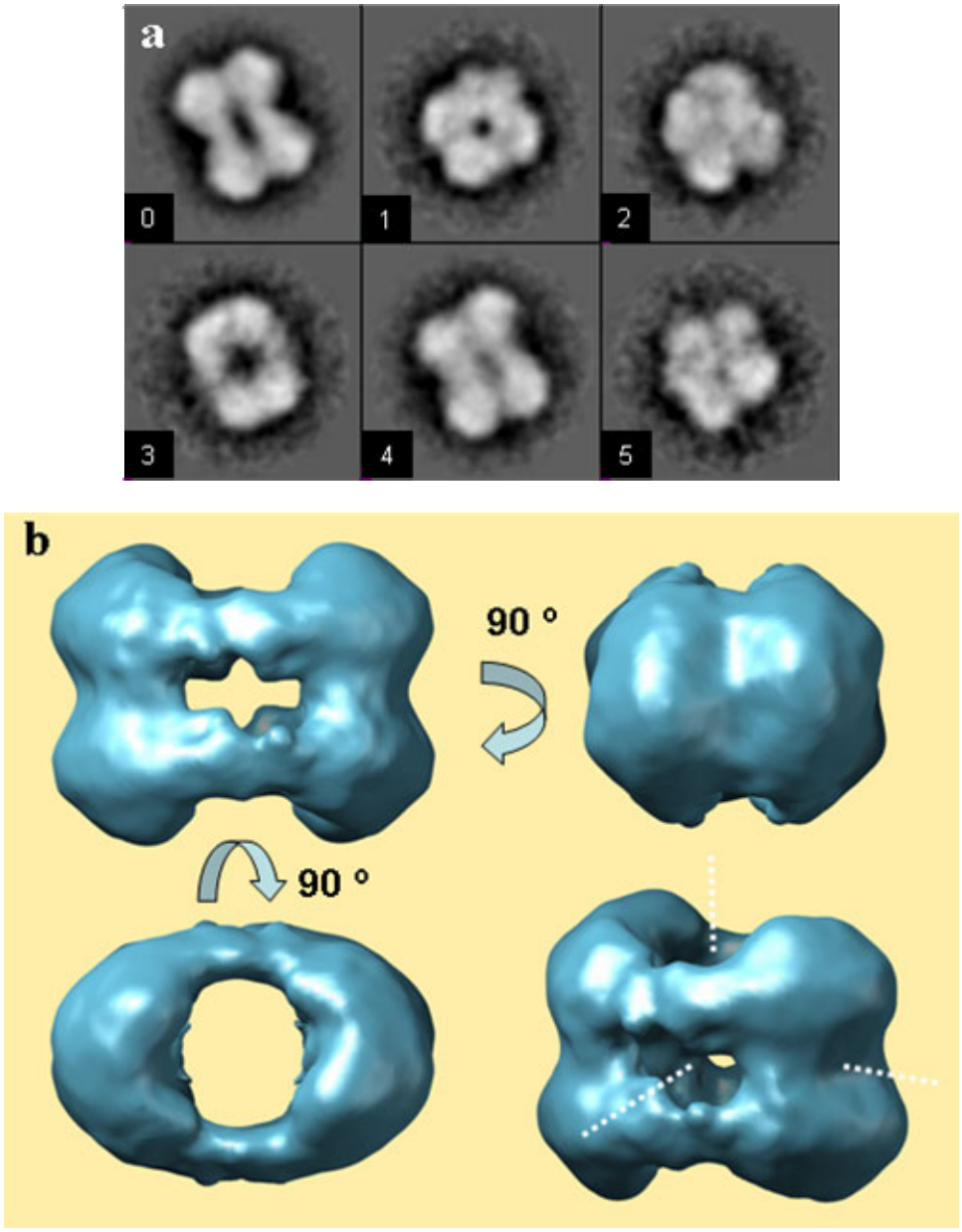

Figure 3.2 Single particle classification and 3D reconstruction of *tb* SecA
Classification of 4,500 single particles resulted in characteristic class averages shown (a). The most predominant orientation is shown by class 0. Class 1 displays a small cavity that was obscured in the final reconstruction but is supported by the crystal structure of *tb* SecA. Box size corresponds to 24 nm.
Surface rendering of the SecA 3D map (b). The three orthogonal views are shown with the angular relationship indicated. The lower right intermediate view showing the three axes of 2-fold rotational symmetry (white dashed lines). The upper right view corresponds to class 1 but the small cavity is not visible in the surface rendered map.

The validity of the EM reconstruction was confirmed with the completion of the atomic resolution structure of *tb* SecA that was solved in parallel to, and published with this EM study (116). The crystal lattice of SecA indicated that the same tetramer was being formed in the lattice that had a $P6_222$ space group. Docking of the EM density map to the crystal structure of *tb* SecA showed good agreement between the low and high resolution data (Figure 3.3). The crystal structure could be tightly enclosed by the EM density apart from the two cavities that appeared larger in the EM map. This is possibly a result of the negative stain accumulating and overemphasizing these features and has been reported previously for proteins subjected to negative staining (117,118). The small twist observed between the 2 dimers in the EM map was also confirmed by the crystal structure thus emphasizing the synergy that exists between different structural approaches.

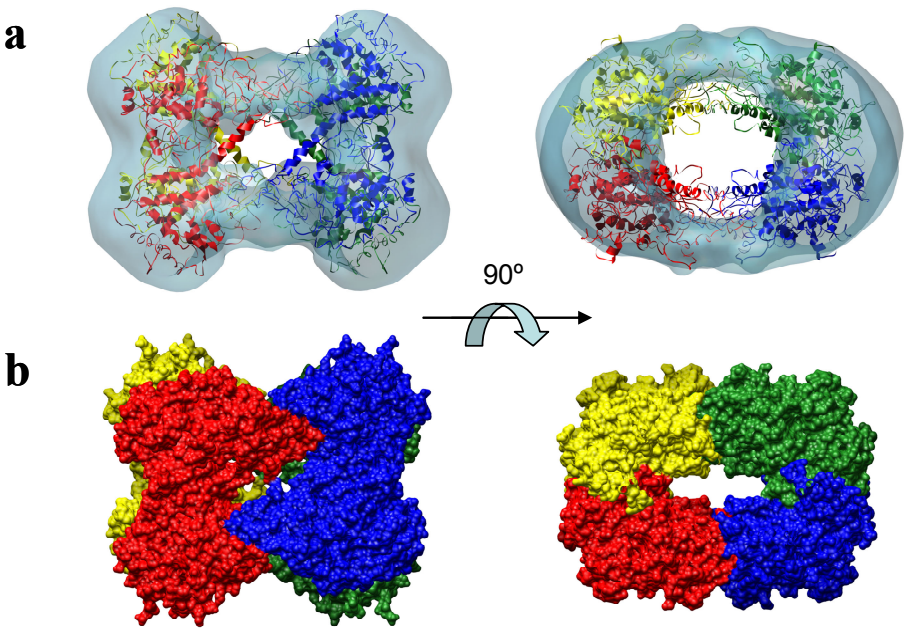

Figure 3.3 Comparison of EM map to *tb* SecA crystal structure
The crystal structure of apo *tb* SecA (PDB ID: 1NL3) was docked to the *tb* SecA map (a). Surface representation of the crystal structure (b) shown in the same orientations as in (a).

3.2.2 SecYEG overexpression study

At the time this study was conducted the oligomeric state of the SecA-SecYEG complex had not yet been established. The crystal structure of *M. jannaschii* SecY indicated that the functional channel might consist of a monomer of SecYEG, but the restrictions imposed by this narrow pore raised the question of whether higher oligomers might contribute to this channel. Furthermore the only high resolution structure for a protein conducting channel was that from the archaea which lack a *secA* gene. Therefore studies on the eubacterial homologues could provide more insight into the SecA-SecYEG interaction. Single particle EM offers an advantage in the study of large multiprotein complexes. With this in mind we attempted to overexpress and purify the mycobacterial SecYEG components for reconstitution with SecA, followed by structural studies by EM and X-ray crystallography.

The susceptibility of the *E. coli* SecY when not associated with SecE, by FtsH prompted us to co-express the two genes (119). The *secY* and *secE* genes were cloned in the petDuet vector, which allows co-expression of two genes, each under the control of a T7 promoter and each with its own Shine Dalgarno sequence. Due to the low levels of expression associated with membrane protein overexpression, *secY* carried a His_6 tag sequence at the 5' end of its reading frame to aid in the detection by α-His_6 antibodies and in the subsequent purification of the complex. In parallel, a second construct contained a His_6 tag sequence at the 5'of *secE*. The two constructs (Sec_{Nhis}YE and $SecY_{Nhis}$E) were used for small scale expression studies using the expression hosts BL21(DE3) and BL21(DE3) pLysS. Expression in these strains did not reveal any

accumulation of SecY as judged by Western blot analysis, but *secE* expression could be detected. C43(DE3), a derivative strain of BL21(DE3) that is resistant to toxic protein overexpression was also used in expression trials (120). This strain carries uncharacterized mutations that allow growth and high levels of overexpression of *uncF,* which encodes the b subunit of the F(1)F(o) complex. Overexpression of *uncF* induces the production of intracellular membranes that are speculated to help in the accumulation of target membrane proteins (121). We performed electron microscopy of ultra-thin sections of C43(DE3) expressing *uncF* from the vector pAVD10 and confirmed the production of the intracellular membranes (Figure 3.4a, b). The *uncF* gene was subsequently cloned into the pETDuet vector along with its promoter and Shine Dalgarno for co-expression with *secY* and *secE* but unfortunately no SecY could be detected.

The lack of accumulation of SecY prompted the consideration of if the problem was at the transcriptional, translational or post-translational level. The high GC content of mycobacteria can influence the heterologous expression of mycobacterial proteins in *E. coli* (122). The commercially available *E. coli* strain Rosetta (EMD Biosciences) is a derivative of BL21(DE3) that carries the plasmid pRARE. The genes for six tRNAs that recognize codons underrepresented in *E. coli* are expressed constitutively from pRARE. This aids in the expression of mammalian genes or genes from high GC content organisms. Inspection of the codon usage showed that five of these six codons (AGG, AGA, CUA, CCC, CGA) were present in *secY*. Expression tests using the Rosetta pLysS strain successfully allowed the accumulation of α-His detectable levels of SecY. The

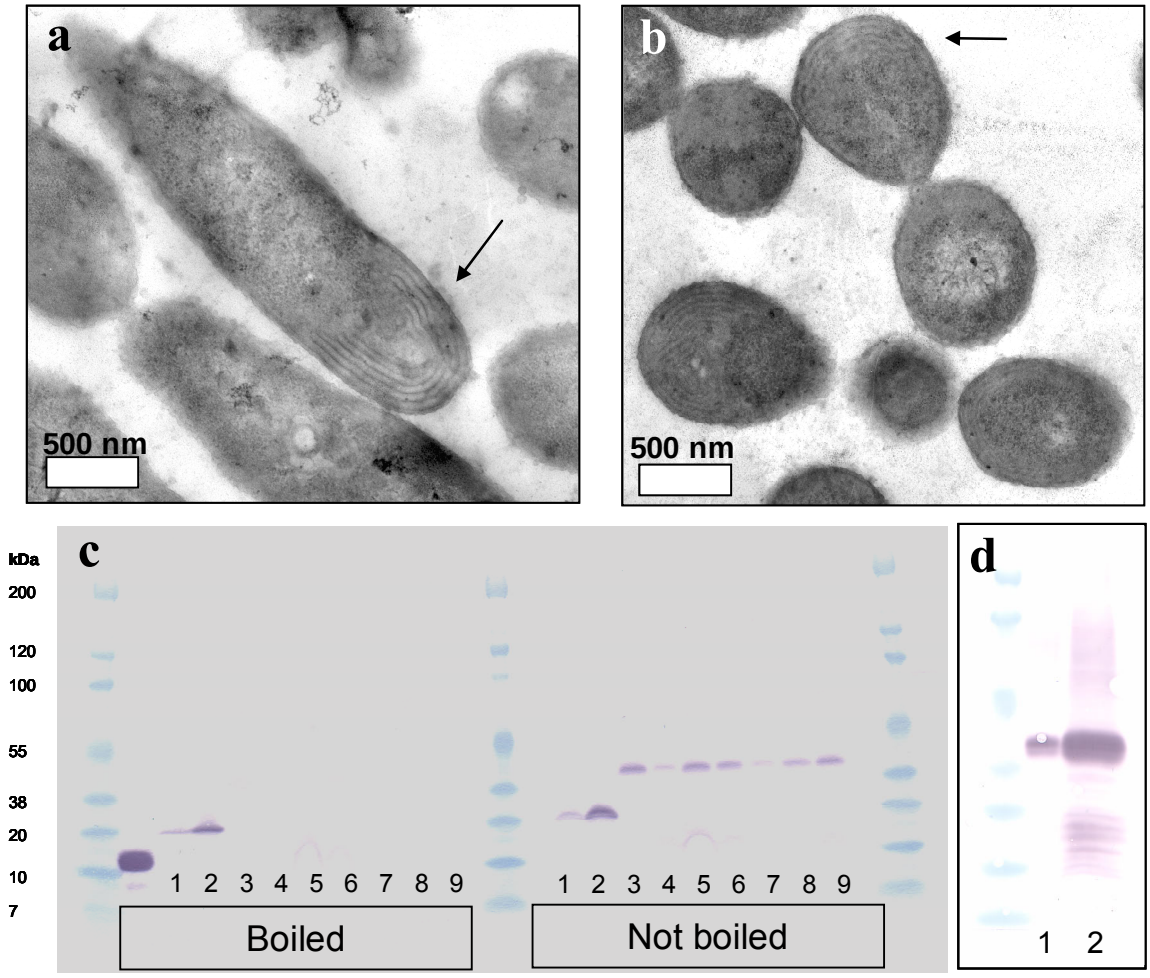

Figure 3.4 Expression studies of *tb secYE*
Electron micrographs of ultra-thin sectioned *E. coli* C43(DE3) cells expressing *uncF* from pAVD10 (a, b). Arrows indicate the intracellular membrane production. Anti-His western blot of media optimization for expression of *secY* (c, lanes 3-9 correspond to hyper, glucose M9Y, power, LB, TB , turbo and superior both). Post-induction temperature effect on the expression of *secE* (c, lanes 1, 2 correspond to 37 and 16 ºC respectively) and *secY* (d, lanes 1, 2 correspond to 37 and 16 ºC respectively)

expression of SecY was further optimized by growing cells in terrific broth (TB) instead of LB broth and inducing the cells at 16 °C for 18 hours instead of 37 °C for 3 hours (Figure 3.4 c, d).

After the optimization of *secY* expression, a third construct containing His_6 tags for both genes was used for detergent extraction and purification trials. The zwitterionic detergent EBB (Empigen BB) efficiently extracted both proteins from the membrane and was used throughout the purification procedure. Initial purification using metal affinity chromatography resulted in impure protein preparation. The purity was increased by passing the membrane extracts through anion and cation exchange columns prior to the metal affinity step (Figure 3.5a). The combined SecYE isoelectric point of 10.0 allowed passage of the complex through the anion exchange column while many contaminating proteins were restained. After elution from the cation exchange column and the subsequent metal affinity step (Figure 3.5b), purified SecYE was subjected to gel-permeation chromatography to isolate the SecYE complex from the excess SecE (Figure 3.5c). The final yield of SecYE complex was between 0.3 and 0.5 mg per liter of culture, while about 7 mg of uncomplexed SecE was isolated.

Expression of *secG* alone from a separate pACYC vector was also successful, however co-expression of SecG/pACYC along with SecYE/pETDuet resulted in decreased levels of SecY.

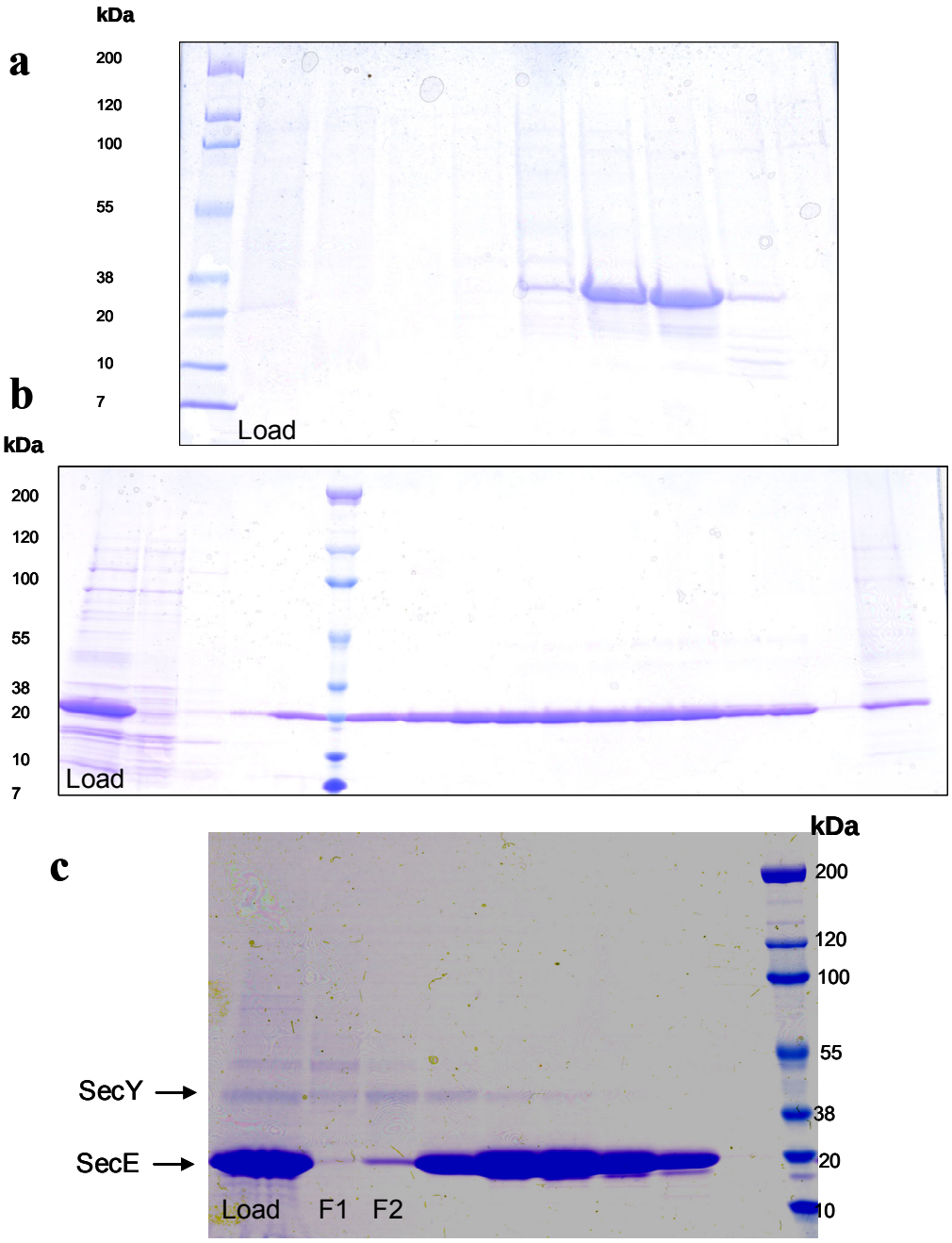

Figure 3.5 Purification of EBB solubilized SecYE
Coomassie blue stained SDS-PAGE of SecYE purification procedure. First elution from a cation exchange column (a). The peak fractions were then passed through a Ni-NTA column and eluted with a linear gradient of imidazole (b). The eluted protein was concentrated and subjected to gel-permeation chromatography (d). F1 corresponds to the 1:1 SecY:SecE complex.

3.3 DISCUSSION

The transport of proteins across the lipid bilayer is an essential biological process in all living organisms. The general secretory pathway or Sec system has been characterized by biochemical, genetic and biophysical methods since its discovery almost thirty years ago. These studies have identified the gene products and pathways involved in this highly complex process consisting of membrane embedded and cytosolic partners. Despite numerous attempts of the structural determination of the Sec components, significant progress in this effort has only been achieved in the past seven years. At the start of my experiments in this field, atomic resolution structures were only available for SecB (123) and a lower resolution structure of *E. coli* SecYEG which had been solved to 9 Å by electron crystallography (94). In this study we undertook the structural determination of the SecA and SecYEG components from *M. tuberculosis*.

The efforts to overexpress the *tb* SecYEG complex for structural studies were hampered by the low levels of SecY expression. Although significant progress was made after screening expression hosts, types of growth media and induction temperatures, the low expression levels of SecY, especially when co-expressed with SecE and SecG, made it extremely challenging to obtain a complex at defined stochiometric ratios. After gel-permeation chromatography in the detergent EBB, what seemed to be a 1:1 complex of SecY: SecE, was obtained. However, EBB interferes with EM visualization by displaying a pronounced lammelar phase, thereby obscuring other particles. The very low yield of protein and the nature of this detergent made this preparation unsuitable for

EM and X-ray structural studies. That said, these expression and purification trials will provide the foundations for future studies of the *tb* SecYEG complex.

An electron microscopic approach to the structural characterization of *tb* SecA was carried out in parallel to collaborate X-ray crystallographic studies. Favorable conditions for single particle analysis were obtained in buffer containing high concentrations of sodium formate, a prerequisite for the growth of *tb* SecA 3D crystals. The relationship observed between conditions suitable for both single particle analysis and 3D crystallization was observed again for the S105 holin (Chapter II). The EM structure obtained from these conditions was of a tetrameric form of SecA with D2 symmetry. The tetramer exhibited large cavities that were present at two of the symmetry axes. A smaller cavity at the third symmetry axis that was present in the class averages, was not visible in the final reconstruction. The tetrameric assembly observed was in contrast to previous biophysical studies that indicated *E. coli* SecA was present *in vitro* as a dimer (113,114) but consistent with small angle X-ray scattering studies of a C-terminal truncated form of *E. coli* SecA that formed tetramers (124). Furthermore it was shown that detergent solubilized *E. coli* SecYEG assembled into tetramers upon interaction with SecA (95). This raises the question of a symmetry match that may facilitate docking of a tetrameric SecA to a tetrameric SecYEG. However, the unusual buffer condition that resulted in monodisperse behavior of SecA on the carbon support film raises questions about the effect of the high salt on the oligomerization. Interestingly, the 3D crystals that formed in the same buffer conditions also have a tetrameric form in the lattice. Since the tetramer is observed in solution (single particles),

it is unlikely that the oligomer formation is induced by crystal packing. Examination of the crystal structure of *tb* SecA reveals that one of the two dimer interfaces is formed by non-conserved residues and residues that resulted from the expression vector itself. Thereby it was concluded that the tetramer may be due to the interaction of the biologically relevant dimers through non-physiological interactions.

Shortly after the *tb* SecA EM and crystal structures were published, another group reported EM data of *E. coli* SecA in the presence of lipid monolayers (125). SecA assumed two main forms, a dumbbell shaped form very similar in size and shape to our tetramers, and a ring shaped form, widely different than any form seen in *M. tuberculosis* or *B. ssubtilis* SecA.

The biological assembly of SecA seems dynamic and dependent on several factors including the interaction with SecB (126), the preprotein substrate (127) and association with SecYEG (128) as well as the lipid bilayer (127,129). It is also important to point out that the described structures depict a snapshot of the molecular assembly and although the fold of the monomer might not change overall globally, small changes could have a significant effect on oligomerization. It is therefore rash to speculate that any one of the SecA assemblies bears biologically significance, as the nature of this assembly may change in response to its immediate environment or ligand. It may also be possible that the oligomeric assembly of SecA is genus specific and that SecA in *M. tuberculosis* may be involved in other biological roles that require higher oligomers, including a role as a SecB substitute that is missing in mycobacteria. Unusual as this may seem, it is not without precedence. In *E. coli* the chaperone GroEL exists as a

tetradecamer but in mycobacteria which carry two homologues, GroEL1 and GroEL2 exist as dimers with no higher oligomers being formed (130). In addition, *M. smegmatis* GroEL1 has been implicated in the formation of biofilm by interacting directly with and regulating the FAS-II protein KasA, a role beyond the typical chaperone function (131). The significance of the *tb* SecA tetramer or higher SecA oligomers from other organisms may be revealed by future studies.

3.4 EXPERIMENTAL PROCEDURES

3.4.1 Electron microscopy and single particle analysis

SecA was diluted to a concentration of 30 μg/ml in buffer containing 20 mM Tris, 4M sodium formate, 150 mM NaCl, 1 mM DTT pH 8.0. Specimens were prepared according to the method of Valentine *et al.* (107) using an aqueous solution of uranyl acetate (2% w/v, pH 4.25) and omitting any washing steps. Samples were observed in a Zeiss 10C transmission electron microscope operated at 80 kV. Electron micrographs were recorded on Kodak 4489 image sheet films at a calibrated magnification of 37,000 and developed using the D-19 developer, diluted threefold. Selected micrographs were digitized using a Leafscan 45 microdensitometer at a 10 μm scan step corresponding to 2.7 Å/pixel at the specimen level. 4390 particles from 15 micrographs were analyzed using the IMAGIC-5 software package (40). The selected particles were normalized and band-pass filtered (150 and 10Å) to remove low and high frequency noise, respectively. Multivariate statistical analysis (MSA) followed by hierarchical ascendant classification

(HAC) (41) was used to provide the first reference-free class averages from which references were selected for multi-reference alignment (MRA). The process was iterated until stable classes were obtained based on visual inspection of class members and the intraclass variance output. A three-dimensional reconstruction was calculated using angular reconstitution and the filtered back projection algorithm (40). For rendering and visualization, the SITUS (132) and VMD (133) packages were employed using a protein density of 0.844 Da/Å^3 (109).

3.4.2 Cloning of SecYE complex

The genes for *secY* and *secE* were amplified from *M. tuberculosis* H37RV genomic DNA using the polymerase chain reaction. The restriction enzyme recognition sites 5'BamHI-3'HindIII and 5'NdeI-3'XhoI were introduced in the amplified products for *secY* and *secE* respectively. Additionally a hexahistidine tag sequence was introduced at the 5' end of *secE*. PCR products were digested with appropriate restriction endonucleases and ligated into the pETDuet-1 vector (Novagen), a vector that allows the co-expression of two target genes each under the control of its own T7 promoter. *secY* was introduced into the first cloning cassette in which it acquires a hexahistidine tag as part of the pETDuet vector. *secE* was cloned into the second cloning cassette. After both genes were cloned into the pETDUET vector, the two genes were subcloned into the pET28a vector by mobilizing both MCS under the BamHI-XhoI restriction sites. This was done due to a frame-shift that occurs when using the BamHI sites in some of the pET vectors (the b series and some pETDuet vectors).

3.4.3 Purification of SecYE complex

A single colony from plates of fresh transformants was used to create a lawn. The next day the lawn was scraped and used to inoculate 300 ml of TB medium supplemented with 50μg/ml kanamycin and 34 μg/ml chloramphenicol. The culture was grown to OD_{600} of 0.7 and 50 ml of this culture was used to inoculate 2 L of TB (terrific broth) medium. Cells were then grown to OD_{600} of 0.7, cooled on ice for 15 minutes followed by induction with 1 mM IPTG. The cultures were then grown at 16 °C for 18 hours after which cells were harvested by centrifugation and lysed by two passes through a high pressure French-press cell. The lysate was diluted so that one gram of the original wet-cell paste is finally diluted in 10ml of lysis buffer (10mM Tris pH 8.0, 140mM NaCl, 10% glycerol, 3mM 2-mercaptoethanol, EDTA-free protease inhibitor cocktail). The lysate was spun at 10,000 x g for 15 minutes at 4 ºC to remove unlysed cells and debris. The dilution allows for better separation of unlysed cells from membranes during the 10,000 x g spin. The supernatant was then spun at 100,000 x g in a Beckman Ti50.2 rotor for 90 minutes at 4 ºC to harvest the total membranes. Membranes were solubilized overnight in 20 mM Tris pH 8.0, 10mM $MgCl_2$, 10 mM NaCl, 20% (v/v) glycerol, 2% (v/v) Empigen BB, 3 mM 2-mercaptoethanol. Unsolubilized material was removed by centrifugation at 100,000 x g for 60 minutes at 4 ºC. Solubilized membranes were loaded onto a Q-Sepharose anion column connected to an SP-Sepharose cation exchange column equilibrated with 20mM Tris pH 8.0, 10mM NaCl, 0.5% EBB, 20 % glycerol (v/v) and 3 mM 2-mercaptoethanol. SecYE has a theoretical combined isoelectric point of 10 and in practice the complex does not bind to the Q-Sepharose column but remains

bound to the SP-Sepharose. In this way major contaminants can be removed as they bind to the Q-Sepharose. SecYE was eluted from the SP column using a linear gradient of NaCl.

Fractions from the SP-column containing SecYE were pooled and loaded onto a 5ml Ni-NTA column equilibrated with 20mM Tris pH 8.0, 150mM NaCl, 0.5% EBB, 20 % (v/v) glycerol, 10 mM imidazole and 3 mM 2-mercaptoethanol and eluted using a linear gradient of imidazole. The peak fractions from the Ni-NTA elution were pooled, concentrated to 2 ml and loaded onto a 200 ml S-200 gel-permeation chromatography column for separation of oligomeric species.

CHAPTER IV

CONCLUSIONS AND FUTURE DIRECTIONS

4.1 ELECTRON MICROSCOPY AS A STRUCTURAL CHARACTERIZATION METHOD

Cell biology, genetics, biochemistry and other biological sciences strive to identify and determine the role of genes and their products in the greater context of a living organism. By observing phenotypes associated with certain genes and their alleles, the biological function of these genes can be identified. Isolation and characterization of the gene products can then provide further information on their function and even allow the *in vitro* reconstitution of the biological system in question. Ultimately, however, a thorough understanding at the molecular level requires a look at the structure of the macromolecule.

The information offered by structural analysis of macromolecules is invaluable to the further understanding of how these molecules perform their biological roles and how form fits to function. The three-dimensional arrangement of polypeptide chains and the sequence and chemistry of amino acids ultimately define the macromolecule and how it carries out its catalytic, structural, signaling or mechanical function. Methods such as X-ray crystallography and NMR allow for structural determination at atomic resolution and a detailed examination of macromolecular architecture. Frequently though, single polypeptides will assemble into larger complexes to perform their biological function

and the study of such complexes as a whole is beyond the size-restrictions associated with crystallography and NMR. Electron microscopy can provide structural information for such assemblies albeit at lower resolution. Electron crystallography can be used to investigate the structures of membrane proteins in the context of the lipid bilayer and can achieve resolutions below 2Å.

Single particle analysis offers an escape from 2D and 3D crystallization. Using computational techniques, single complexes can be visualized and used to reconstruct a structure representing an average of the population. This technique can accommodate virtually any size assembly with only smaller (< 200 kDa) complexes posing technical limitations. Furthermore, the amounts of biological material required are 10-100 fold less than those required for x-ray crystallography and NMR, thus allowing the study of macromolecules isolated from their native host and those which pose overexpression difficulties.

The structural studies performed on the assemblies formed by the bacteriophage lambda S holin (>1.7 Mda) and the *M. tuberculosis* SecA translocase (400 kDa) illustrate the dynamic size range of single particle EM. The structural determination of the large and slightly heterogeneous ring assembly formed by a small, 12 kDa membrane protein, could only have been achieved in a reasonable time frame by single particle EM. The computational separation of varying sized rings made it possible to obtain a structure from a relatively heterogeneous dataset and provided the first ever look into a multimeric assembly formed by a bacteriophage holin.

4.2 S105 FORMS RINGS OF LARGE DIAMETER

Membrane proteins such as the lambda holin S, pose difficulties for molecular characterization, because of the requirement of detergents to be kept in solution. The chemical nature of the detergent can affect the stability and oligomeric state of the protein. Traditionally non-ionic detergents have been employed in biochemical and structural studies because they are less destructive to the protein-protein interactions that hold the assembly together, unlike their ionic and zwitterionic counterparts. The lambda S holin which only consists of three TMDs must oligomerize in the host membrane in order to create a lesion large enough to disrupt the membrane and allow the endolysin exit.In order to isolate these proteins from the membrane bilayer and keep them thereon in solution, detergents have to be used. S had previously been isolated in the non-ionic detergent, OG, but its stability was short term. The zwitterionic detergent EBB could keep S stably solubilized, but as shown in this study, it ultimately affected its ability to oligomerize into larger assemblies that were known to be present in the membrane. In order to preserve the protein-protein interaction of the S oligomer, other non-ionic detergents were employed for the structural characterization in this study.

The non-ionic detergent DDM was found to extract and solubilize S for biochemically useful periods of time. Furthermore using gel-permeation, EM and single particle analysis it was shown that S forms large ring shaped multimers, which were solved to a resolution 2.6 nm. An estimated 72 S105 molecules contribute to one ring that measures 23 and 8.5 nm in outer and inner diameters respectively and has a height

of 4 nm, matching the estimated thickness of a lipid bilayer. Each ring displays an 18-fold rotational symmetry with an estimated four S105 molecules per asymmetric unit or protomer. To query the biological relevance of these assemblies we employed a variety of biochemical approaches. Using chemical crosslinking we showed that structures in the ring size range can be captured in the membrane. Moreover, the cysteine accessibility and protease digestion results for S in IMVs and in detergent showed almost identical patterns, indicating that the protein-protein interactions and the oligomeric assembly observed in detergent was similar if not the same as the one found in the membrane. Further evidence arises from the fact that the oligomerization pattern is *S* allele dependent with S_{a52v} and $S_{\Delta TM1}$ unable to form these multimers.

The finding that S105 can assemble into rings follows the pattern of other pore-forming toxins and cytolysins that assemble into small or even larger than 30 nm ring assemblies. Fundamental differences exist in respect to holins since these toxins are otherwise soluble proteins that undergo a secondary structure rearrangement prior to membrane insertion. The holins are the only examples of true integral membrane proteins with lethal pore forming ability and therefore should display fundamental differences in the process of membrane lesion formation. The S rings offer a first glance at what this process may involve and provide the foundation for further investigation of holin mechanics.

Future efforts will include investigating the molecular arrangement of the ring subunits. Interesting questions include how curvature is achieved by a possible asymmetry between molecules and what surfaces of the helices are exposed in the inner

hole versus the outer lining of the ring. Promising data that may answer some of these questions are the preliminary crystallization and X-ray diffraction results from crystals of $S_{\Delta TM1}$ and S_{a52v} which were crystallized in the presence of DDM. An atomic resolution structure of S could be combined with the lower resolution EM data of the rings to provide some answers to these questions.

Ultimately S has to be observed in the context of the lipid bilayer. Visualization of liposome reconstituted S was hampered by the ability of S to form water soluble structures, thus escaping insertion into the bilayer or in some cases inserting but having no effect on liposomes as judged by EM. Observation of S enriched IMVs also indicated the presence of soluble ring structures, which raised questions as to their biological relevance. These assemblies should be investigated further as they may represent an alternate form of S, i.e. a different stage during S function. Moreover, intermediate "honeycomb" structures could be captured *in vitro* at low detergent concentrations. These rafts of rings may be involved in lysis timing and coordinating hole formation between the rings and ultimately between the 1000-3000 S molecules produced during infection. Finally, observation of the large S assemblies in intact cells by means of cryo-electron tomography will provide the most biologically pertinent structural information of such assemblies. The large library of mutants characterized over the past years and the ability to capture intermediates by instant vitrification promises new insights into the mechanism of lesion formation by holins.

4.3 THE STRUCTURE OF *tb* SecA

The electron microscopic studies described in this dissertation provided the first structural information for *M. tuberculosis* SecA in solution. Our EM studies on *tb* Sec A indicated that under the experimental conditions used, *M. tuberculosis* SecA formed a tetramer, contrary to previous reports from other organisms that indicate dimer formation. Whether this assembly bears biological significance will require further investigation. Inspection of the crystal structure of *tb* SecA suggested that the dimer-dimer interface of the tetramer was created by non-conserved residues and residues from the expression vector itself. Constructs devoid of foreign sequence as well as other methods for determining oligomeric states need to be used to address the oligomerization properties of *tb* SecA

The crystal structures of *Bacilus subtilis* SecA (134) and *tb* SecA (116) show an overall similar fold for the monomers, but the dimeric interfaces formed by these two homologues are different. Both structures describe antiparallel dimers (a head to tail or N to C arrangement) but the contacts involved in dimerization are comprised of different parts of the respective polypeptide sequences. More recently, the crystal structure of *E. coli* SecA indicated a dimeric association completely different from both *B. subtilis* and *tb* SecA in which only the N-terminal domains are involved in dimerization while the C-terminal domains are facing away from each other (135). Finally, the crystal structure of *Thermus thermophilus* SecA provides a third distinct dimer in which the two subunits form a parallel dimer (head to head or N to N) (136). In light of these observations one

cannot rule out that the SecA oligomeric state may be organism specific. Furthermore, crystal packing effects on the quaternary arrangement of molecules have to be accounted for and resolved by other means. The quaternary structure of the *tb* SecA crystal structure was confirmed by comparison to the solution EM structure of *tb* SecA and hence cannot be a result of crystal packing forces.

Efforts to study the SecYEG heterotrimeric complex *M. tuberculosis* were also carried out but were hampered by *secY* expression problems. Significant progress was made however after optimization of the expression hosts and growth conditions. These conditions allowed the co-expression of secY and SecE but not in stochiometric amounts. Further optimization studies to deal with the low levels of SecY are required to obtain quantities of the stochiometric complex useful for structural characterization. The problems encountered here are not unfamiliar to membrane protein overexpression, but new developments in *in vivo* and *in vitro* expression systems promise ways to overcome these obstacles.

REFERENCES

1. Ureta-Vidal, A., Ettwiller, L., and Birney, E. (2003) *Nat Rev Genet* **4**(4), 251-262
2. Voet, D., Voet, J. G., and Prat, C. W. (eds). (2002) *Fundamentals of Biochemistry*, John Wiley and Sons, Inc, New York
3. Lundstrom, K. (2006) *Mol Biotechnol* **34**(2), 205-212
4. Lundstrom, K. (2006) *Cell Mol Life Sci* **63**(22), 2597-2607
5. Lucic, V., Forster, F., and Baumeister, W. (2005) *Annu Rev Biochem* **74**, 833-865
6. Bazzola, J. J., and Russel, L. D. (eds). (1998) *Electron Microscopy*, Jones and. Bartlett Publishers, Inc, Sudbury
7. Misell, D. L. (ed) (1978) *Image Analysis Enhancement and Interpretation*, North-Holland Publishing Company, New York
8. Frank, J. (ed) (1996) *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*, Academic Press, London
9. Ludtke, S. J., Baldwin, P. R., and Chiu, W. (1999) *J Struct Biol* **128**(1), 82-97
10. Fujiyoshi, Y. (1998) *Adv Biophys* **35**, 25-80
11. Henderson, R., and Unwin, P. N. (1975) *Nature* **257**(5521), 28-32
12. Kimura, Y., Vassylyev, D. G., Miyazawa, A., Kidera, A., Matsushima, M., Mitsuoka, K., Murata, K., Hirai, T., and Fujiyoshi, Y. (1997) *Nature* **389**(6647), 206-211
13. Kuhlbrandt, W., Wang, D. N., and Fujiyoshi, Y. (1994) *Nature* **367**(6464), 614-621

14. Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J. B., Engel, A., and Fujiyoshi, Y. (2000) *Nature* **407**(6804), 599-605

15. Gonen, T., Cheng, Y., Sliz, P., Hiroaki, Y., Fujiyoshi, Y., Harrison, S. C., and Walz, T. (2005) *Nature* **438**(7068), 633-638

16. Uzgiris, E. E., and Kornberg, R. D. (1983) *Nature* **301**(5896), 125-129

17. Darst, S. A., Ahlers, M., Meller, P. H., Kubalek, E. W., Blankenburg, R., Ribi, H. O., Ringsdorf, H., and Kornberg, R. D. (1991) *Biophys J* **59**(2), 387-396

18. Bischler, N., Balavoine, F., Milkereit, P., Tschochner, H., Mioskowski, C., and Schultz, P. (1998) *Biophys J* **74**(3), 1522-1532

19. Frank, J. (2002) *Annu Rev Biophys Biomol Struct* **31**, 303-319

20. Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J. (1986) *J Microsc* **141**(Pt 1), RP1-2

21. Van Heel, M. (1987) *Ultramicroscopy* **21**, 111-124

22. Frank, J., Penczek, P., Grassucci, R., and Srivastava, S. (1991) *J Cell Biol* **115**(3), 597-605

23. Frank, J., Zhu, J., Penczek, P., Li, Y., Srivastava, S., Verschoor, A., Radermacher, M., Grassucci, R., Lata, R. K., and Agrawal, R. K. (1995) *Nature* **376**(6539), 441-444

24. Malhotra, A., Penczek, P., Agrawal, R. K., Gabashvili, I. S., Grassucci, R. A., Junemann, R., Burkhardt, N., Nierhaus, K. H., and Frank, J. (1998) *J Mol Biol* **280**(1), 103-116

25. Mitra, K., Schaffitzel, C., Shaikh, T., Tama, F., Jenni, S., Brooks, C. L., 3rd, Ban,

N., and Frank, J. (2005) *Nature* **438**(7066), 318-324

26. Tilley, S. J., Orlova, E. V., Gilbert, R. J., Andrew, P. W., and Saibil, H. R. (2005) *Cell* **121**(2), 247-256

27. Cheng, Y., Zak, O., Aisen, P., Harrison, S. C., and Walz, T. (2004) *Cell* **116**(4), 565-576

28. Savva, C. G., Holzenburg, A., and Bogner, E. (2004) *FEBS Lett* **563**(1-3), 135-140

29. Sun, J., Duffy, K. E., Ranjith-Kumar, C. T., Xiong, J., Lamb, R. J., Santos, J., Masarapu, H., Cunningham, M., Holzenburg, A., Sarisky, R. T., Mbow, M. L., and Kao, C. (2006) *J Biol Chem* **281**(16), 11144-11151

30. Zhang, W., Chipman, P. R., Corver, J., Johnson, P. R., Zhang, Y., Mukhopadhyay, S., Baker, T. S., Strauss, J. H., Rossmann, M. G., and Kuhn, R. J. (2003) *Nat Struct Biol* **10**(11), 907-912

31. Ludtke, S. J., Chen, D. H., Song, J. L., Chuang, D. T., and Chiu, W. (2004) *Structure* **12**(7), 1129-1136

32. Ranson, N. A., Clare, D. K., Farr, G. W., Houldershaw, D., Horwich, A. L., and Saibil, H. R. (2006) *Nat Struct Mol Biol* **13**(2), 147-152

33. Chen, D. H., Song, J. L., Chuang, D. T., Chiu, W., and Ludtke, S. J. (2006) *Structure* **14**(11), 1711-1722

34. Henderson, R. (2004) *Q Rev Biophys* **37**(1), 3-13

35. Stagg, S. M., Lander, G. C., Pulokas, J., Fellmann, D., Cheng, A., Quispe, J. D., Mallick, S. P., Avila, R. M., Carragher, B., and Potter, C. S. (2006) *J Struct Biol*

155(3), 470-481

36. Frank, J., Wagenknecht, T., McEwen, B. F., Marko, M., Hsieh, C. E., and Mannella, C. A. (2002) *J Struct Biol* **138**(1-2), 85-91

37. Zhang, P., Khursigara, C. M., Hartnell, L. M., and Subramaniam, S. (2007) *Proc Natl Acad Sci U S A* **104**(10), 3777-3781

38. Sui, H., and Downing, K. H. (2006) *Nature* **442**(7101), 475-478

39. Nicastro, D., McIntosh, J. R., and Baumeister, W. (2005) *Proc Natl Acad Sci USA* **102**(44), 15889-15894

40. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R., and Schatz, M. (1996) *J Struct Biol* **116**(1), 17-24

41. Van Heel, M. (1989) *Optik* **82**, 114-126

42. White, H. E., Saibil, H. R., Ignatiou, A., and Orlova, E. V. (2004) *J Mol Biol* **336**(2), 453-460

43. Crowther, R. A. (1971) *Proc. R. Soc. London.* **261**, 221-230

44. Harauz G. Van Heel, M. (1986) *Optik* **73**, 146-156

45. van Heel, M., Gowen, B., Matadeen, R., Orlova, E. V., Finn, R., Pape, T., Cohen, D., Stark, H., Schmidt, R., Schatz, M., and Patwardhan, A. (2000) *Q Rev Biophys* **33**(4), 307-369

46. Snyder, L., and Champness, W. (eds). (1997) *Molecular Genetics of Bacteria*, ASM Press, Washington

47. Bernhardt, T. G., Wang, I. N., Struck, D. K., and Young, R. (2001) *Science* **292**(5525), 2326-2329

48. Bernhardt, T. G., Struck, D. K., and Young, R. (2001) *J Biol Chem* **276**(9), 6093-6097

49. Wang, I. N., Smith, D. L., and Young, R. (2000) *Annu Rev Microbiol* **54**, 799-825

50. Taylor, A., Benedik, M., and Campbell, A. (1983) *Gene* **26**(2-3), 159-163

51. Kedzierska, S., Wawrzynow, A., and Taylor, A. (1996) *Gene* **168**(1), 1-8

52. Young, R., Way, J., Way, S., Yin, J., and Syvanen, M. (1979) *J Mol Biol* **132**(3), 307-322

53. Loessner, M. J. (2005) *Curr Opin Microbiol* **8**(4), 480-487

54. Evrard, C., Fastrez, J., and Declercq, J. P. (1998) *J Mol Biol* **276**(1), 151-164

55. Jespers, L., Sonveaux, E., and Fastrez, J. (1992) *J Mol Biol* **228**(2), 529-538

56. Garrett, J., Fusselman, R., Hise, J., Chiou, L., Smith-Grillo, D., Schulz, J., and Young, R. (1981) *Mol Gen Genet* **182**(2), 326-331

57. Reader, R. W., and Siminovitch, L. (1971) *Virology* **43**(3), 623-637

58. Garrett, J. M., and Young, R. (1982) *J Virol* **44**(3), 886-892

59. Altman, E., Altman, R. K., Garrett, J. M., Grimaila, R. J., and Young, R. (1983) *J Bacteriol* **155**(3), 1130-1137

60. Altman, E., Young, K., Garrett, J., Altman, R., and Young, R. (1985) *J Virol* **53**(3), 1008-1011

61. Blasi, U., Fraisl, P., Chang, C. Y., Zhang, N., and Young, R. (1999) *J Bacteriol* **181**(9), 2922-2929

62. Graschopf, A., and Blasi, U. (1999) *Arch Microbiol* **172**(1), 31-39

63. Grundling, A., Blasi, U., and Young, R. (2000) *J Bacteriol* **182**(21), 6082-6090

64. Deaton, J. (2004) *Solubilization and Functional Analysis of the Lambda Holin*. Ph.D Dissertation, Texas A&M University, College Station

65. Blasi, U., Nam, K., Hartz, D., Gold, L., and Young, R. (1989) *Embo J* **8**(11), 3501-3510

66. Chang, C. Y., Nam, K., and Young, R. (1995) *J Bacteriol* **177**(11), 3283-3294

67. Grundling, A., Smith, D. L., Blasi, U., and Young, R. (2000) *J Bacteriol* **182**(21), 6075-6081

68. Blasi, U., Chang, C. Y., Zagotta, M. T., Nam, K. B., and Young, R. (1990) *Embo J* **9**(4), 981-989

69. Wang, I. N., Deaton, J., and Young, R. (2003) *J Bacteriol* **185**(3), 779-787

70. Smith, D. L., Chang, C. Y., and Young, R. (1998) *Gene Expr* **7**(1), 39-52

71. Smith, D. L., and Young, R. (1998) *J Bacteriol* **180**(16), 4199-4211

72. Smith, D. L., Struck, D. K., Scholtz, J. M., and Young, R. (1998) *J Bacteriol* **180**(9), 2531-2540

73. Deaton, J., Savva, C. G., Sun, J., Holzenburg, A., Berry, J., and Young, R. (2004) *Protein Sci* **13**(7), 1778-1786

74. Sun, J., Savva, C. G., Deaton, J., Kaback, H. R., Svrakic, M., Young, R., and Holzenburg, A. (2005) *Arch Biochem Biophys* **434**(2), 352-357

75. de Keyzer, J., van der Does, C., and Driessen, A. J. (2003) *Cell Mol Life Sci* **60**(10), 2034-2052

76. Bolhuis, A. (2004) *Philos Trans R Soc Lond B Biol Sci* **359**(1446), 919-927

77. Van den Berg, B., Clemons, W. M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C., and Rapoport, T. A. (2004) *Nature* **427**(6969), 36-44

78. Romisch, K. (1999) *J Cell Sci* **112** 4185-4191

79. Scotti, P. A., Urbanus, M. L., Brunner, J., de Gier, J. W., von Heijne, G., van der Does, C., Driessen, A. J., Oudega, B., and Luirink, J. (2000) *Embo J* **19**(4), 542-549

80. Samuelson, J. C., Jiang, F., Yi, L., Chen, M., de Gier, J. W., Kuhn, A., and Dalbey, R. E. (2001) *J Biol Chem* **276**(37), 34847-34852

81. Nagai, K., Oubridge, C., Kuglstatter, A., Menichelli, E., Isel, C., and Jovine, L. (2003) *Embo J* **22**(14), 3479-3485

82. Emr, S., and Hanley-Way, S. (1981) *Cell* **23**, 79-88

83. Ito, K., Wittekind, M., Nomura, M., Shiba, K., Yura, T., Miura, A., and Nashimoto, H. (1983) *Cell* **32**(3), 789-797

84. Oliver, D. B., and Beckwith, J. (1981) *Cell* **25**(3), 765-772

85. Kumamoto, C. A., and Beckwith, J. (1983) *J Bacteriol* **154**(1), 253-260

86. Gardel, C., Benson, S., Hunt, J., Michaelis, S., and Beckwith, J. (1987) *J Bacteriol* **169**(3), 1286-1290

87. Riggs, P. D., Derman, A. I., and Beckwith, J. (1988) *Genetics* **118**(4), 571-579

88. Nishiyama, K., Hanada, M., and Tokuda, H. (1994) *Embo J* **13**(14), 3272-3277

89. Valent, Q. A., Kendall, D. A., High, S., Kusters, R., Oudega, B., and Luirink, J. (1995) *Embo J* **14**(22), 5494-5505

90. Randall, L. L., and Hardy, S. J. (2002) *Cell Mol Life Sci* **59**(10), 1617-1623

91. Zhou, J., and Xu, Z. (2003) *Nat Struct Biol* **10**(11), 942-947

92. Mori, H., and Ito, K. (2001) *Trends Microbiol* **9**(10), 494-500

93. Yamada, H., Matsuyama, S., Tokuda, H., and Mizushima, S. (1989) *J Biol Chem* **264**(31), 18577-18581

94. Breyton, C., Haase, W., Rapoport, T. A., Kuhlbrandt, W., and Collinson, I. (2002) *Nature* **418**(6898), 662-665

95. Manting, E. H., van Der Does, C., Remigy, H., Engel, A., and Driessen, A. J. (2000) *Embo J* **19**(5), 852-861

96. Tuberculist, Institut Pasteur, http://genolist.pasteur.fr/tuberculist/. Accessed on 07-10-07

97. Braunstein, M., Brown, A. M., Kurtz, S., and Jacobs, W. R., Jr. (2001) *J Bacteriol* **183**(24), 6979-6990

98. Lenz, L. L., and Portnoy, D. A. (2002) *Mol Microbiol* **45**(4), 1043-1056

99. Braunstein, M., Espinosa, B. J., Chan, J., Belisle, J. T., and Jacobs, W. R., Jr. (2003) *Mol Microbiol* **48**(2), 453-464

100. Grundling, A., Blasi, U., and Young, R. (2000) *J Biol Chem* **275**(2), 769-776

101. Wei, Y., Li, H., and Fu, D. (2004) *J Biol Chem* **279**(38), 39251-39259

102. Soskine, M., Steiner-Mordoch, S., and Schuldiner, S. (2002) *Proc Natl Acad Sci U S A* **99**(19), 12043-12048

103. Nield, J., Morris, E. P., Bibby, T. S., and Barber, J. (2003) *Biochemistry* **42**(11), 3180-3188

104. Grigorieff, N. (1998) *J Mol Biol* **277**(5), 1033-1046

105. Gimpelev, M., Forrest, L. R., Murray, D., and Honig, B. (2004) *Biophys J* **87**(6), 4075-4086

106. Tweten, R. K. (2005) *Infect Immun* **73**(10), 6199-6209

107. Valentine, R. C., Shapiro, B. M., and Stadtman, E. R. (1968) *Biochemistry* **7**(6), 2143-2152

108. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) *J Comput Chem* **25**(13), 1605-1612

109. Zipper, P., Kratky, O., Herrmann, R., and Hohn, T. (1971) *Eur J Biochem* **18**(1), 1-9

110. Futai, M. (1974) *J Membr Biol* **15**(1), 15-28

111. Herskovits, A. A., Seluanov, A., Rajsbaum, R., ten Hagen-Jongman, C. M., Henrichs, T., Bochkareva, E. S., Phillips, G. J., Probst, F. J., Nakae, T., Ehrmann, M., Luirink, J., and Bibi, E. (2001) *EMBO Rep* **2**(11), 1040-1046

112. Top, D., de Antueno, R., Salsman, J., Corcoran, J., Mader, J., Hoskin, D., Touhami, A., Jericho, M. H., and Duncan, R. (2005) *Embo J* **24**(17), 2980-2988

113. Akita, M., Shinkai, A., Matsuyama, S., and Mizushima, S. (1991) *Biochem Biophys Res Commun* **174**(1), 211-216

114. Driessen, A. J. (1993) *Biochemistry* **32**(48), 13190-13197

115. Goodsell, D. S., and Olson, A. J. (2000) *Annu Rev Biophys Biomol Struct* **29**, 105-153

116. Sharma, V., Arockiasamy, A., Ronning, D. R., Savva, C. G., Holzenburg, A., Braunstein, M., Jacobs, W. R., Jr., and Sacchettini, J. C. (2003) *Proc Natl Acad*

Sci U S A **100**(5), 2243-2248

117. Stoylova, S. S., Flint, T. D., Kitmitto, A., Ford, R., and Holzenburg, A. (1998) *Micron* **29**, 341-348

118. Holzenburg, A., Bewley, M. C., Wilson, F. H., Nicholson, W. V., and Ford, R. C. (1993) *Nature* **363**, 470-472

119. Kihara, A., Akiyama, Y., and Ito, K. (1995) *Proc Natl Acad Sci U S A* **92**(10), 4532-4536

120. Miroux, B., and Walker, J. E. (1996) *J Mol Biol* **260**(3), 289-298

121. Arechaga, I., Miroux, B., Karrasch, S., Huijbregts, R., de Kruijff, B., Runswick, M. J., and Walker, J. E. (2000) *FEBS Lett* **482**(3), 215-219

122. Lakey, D. L., Voladri, R. K., Edwards, K. M., Hager, C., Samten, B., Wallis, R. S., Barnes, P. F., and Kernodle, D. S. (2000) *Infect Immun* **68**(1), 233-238

123. Xu, Z., Knafels, J. D., and Yoshino, K. (2000) *Nat Struct Biol* **7**(12), 1172-1177

124. Dempsey, B. R., Economou, A., Dunn, S. D., and Shilton, B. H. (2002) *J Mol Biol* **315**(4), 831-843

125. Wang, H. W., Chen, Y., Yang, H., Chen, X., Duan, M. X., Tai, P. C., and Sui, S. F. (2003) *Proc Natl Acad Sci U S A* **100**(7), 4221-4226

126. Chen, Y., Tai, P. C., and Sui, S. F. (2007) *J Struct Biol* **159**(1), 149-153

127. Benach, J., Chou, Y. T., Fak, J. J., Itkin, A., Nicolae, D. D., Smith, P. C., Wittrock, G., Floyd, D. L., Golsaz, C. M., Gierasch, L. M., and Hunt, J. F. (2003) *J Biol Chem* **278**(6), 3628-3638

128. Natale, P., den Blaauwen, T., van der Does, C., and Driessen, A. J. (2005)

Biochemistry **44**(17), 6424-6432

129. Ding, H., Mukerji, I., and Oliver, D. (2001) *Biochemistry* **40**(6), 1835-1843

130. Qamra, R., Srinivas, V., and Mande, S. C. (2004) *J Mol Biol* **342**(2), 605-617

131. Ojha, A., Anand, M., Bhatt, A., Kremer, L., Jacobs, W. R., Jr., and Hatfull, G. F. (2005) *Cell* **123**(5), 861-873

132. Wriggers, W., Milligan, R. A., and McCammon, J. A. (1999) *J Struct Biol* **125**(2-3), 185-195

133. Humphrey, W., Dalke, A., and Schulten, K. (1996) *J Mol Graph* **14**(1), 33-38, 27-38

134. Hunt, J. F., Weinkauf, S., Henry, L., Fak, J. J., McNicholas, P., Oliver, D. B., and Deisenhofer, J. (2002) *Science* **297**(5589), 2018-2026

135. Papanikolau, Y., Papadovasilaki, M., Ravelli, R. B., McCarthy, A. A., Cusack, S., Economou, A., and Petratos, K. (2007) *J Mol Biol* **366**(5), 1545-1557

136. Vassylyev, D. G., Mori, H., Vassylyeva, M. N., Tsukazaki, T., Kimura, Y., Tahirov, T. H., and Ito, K. (2006) *J Mol Biol* **364**(3), 248-258

VITA

Name: George Christos William Savva

Address: 3 Aristomenous Street, Larnaca, Cyprus

Email address: csavva@mail.bio.tamu.edu

Education: B.S Joint honours in genetics and medical microbiology, University of Leeds, England. 2000