STRUCTURAL STUDIES OF THE SARS VIRUS NSP15 ENDONUCLEASE AND THE HUMAN INNATE IMMUNITY RECEPTOR TLR3

A Dissertation

by

JINGCHUAN SUN

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

May 2006

Major Subject: Microbiology

STRUCTURAL STUDIES OF THE SARS VIRUS NSP15 ENDONUCLEASE AND THE HUMAN INNATE IMMUNITY RECEPTOR TLR3

A Dissertation

by

JINGCHUAN SUN

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
Jin Xiong
Deborah A. Siegele
Head of Department, Vincent Cassone

May 2006

Major Subject: Microbiology

ABSTRACT

Structural Studies of the SARS Virus Nsp15 Endonuclease and the Human Innate Immunity Receptor TLR3. (May 2006)

Jingchuan Sun, B.S., Peking University; M.S., Peking University

Chair of Advisory Committee: Dr. Andreas Holzenburg

Three-dimensional (3D) structural determination of biological macromolecules is not only critical to understanding their mechanisms, but also has practical applications. Combining the high resolution imaging of transmission electron microscopy (TEM) and efficient computer processing, protein structures in solution or in two-dimensional (2D) crystals can be determined. The lipid monolayer technique uses the high affinity binding of 6His-tagged proteins to a Ni-nitrilotriacetic (NTA) lipid to create high local protein concentrations, which facilitates 2D crystal formation. In this study, several proteins have been crystallized using this technique, including the SARS virus Nsp15 endonuclease and the human Toll-like receptor (TLR) 3 extracellular domain (ECD). Single particle analysis can determine protein structures in solution without the need for crystals. 3D structures of several protein complexes had been solved by the single particle method, including IniA from *Mycobacterium tuberculosis*, Nsp15 and TLR3 ECD. Determining the structures of these proteins is an important step toward understanding pathogenic microbes and our immune system.

ACKNOWLEDGEMENTS

I would like to thank the members of my committee, Drs. Andreas Holzenburg, Ryland Young, Jin Xiong and Deborah Siegele for their help and comments. I would especially like to thank Dr. Andreas Holzenburg for his patience, encouragement and effort. Many professors and researchers within Texas A&M and off the campus have given much help and assistance. Without them, it would have been impossible to continue and finish the projects presented here.

Finally I want to give many thanks to Dr. Cheng Kao. Although he is not on my committee, he gave me much guidance, help, and encouragement. I am very lucky to have met him. I admire his knowledge, ideas, problem-solving ability and confidence. Although I always think that God only helps those who help themselves, having a mentor like him is God's will.

NOMENCLATURE

CTF	Contrast Transfer Function
dsRNA	Double Stranded RNA
ECD	Extracellular domain
FFT	Fast Fourier Transform
IL-1	Interleukin 1
INH	Isoniazid
kDa	Kilo Dalton
LPS	Lipopolysaccharide
LRR	Leucine-Rich Repeat
Ni-NTA	Nickel Nitrilotriacetic Acid
NF-κB	Nuclear Factor Kappa B
NSP15	Nonstructural Protein 15
PAMP	Pathogen-Associated Molecular Pattern
PRR	Pathogen Recognition Receptor
R protein	Plant Disease Resistance Protein
SARS	Severe Acute Respiratory Syndrome
SNP	Single Nucleotide Polymorphism
TEM	Transmission Electron Microscopy
TLR3	Toll-Like Receptor 3

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

The new millennium began with the great achievement of sequencing the whole human genome. Many other genomic projects have been either finished or are in progress (Ureta-Vidal et al., 2003). The vast amount of sequence information is the basis for our understanding of the important processes of living organisms at the molecular level. A more challenging task in this post-genomic era is to study how the expression of genes is regulated and how the gene products (mostly proteins) work. The three-dimensional (3D) structures of the proteins and their complexes need to be determined and their functions and molecular mechanisms need to be characterized. Although DNA sequencing can be automated, the available methods for determining the 3D structures of proteins are time-consuming and technically demanding. The structures of some important proteins remain unknown even after many years of effort by many experienced researchers. We need to develop new techniques and new methods to solve the problems, especially for those proteins that are difficult to express or purify, or cannot be crystallized by traditional ways. Even though solving the 3D structure of proteins and their complexes is difficult, it is important and critical for us to understand their molecular mechanisms and how they work inside cells (Frank, 2002; Clemons et al., 2004).

This dissertation follows the style of Cell.

Methods for Protein Structure Determination

The traditional method used to determine the 3D structure of biological macromolecules is X-ray crystallography. This procedure requires proteins to form high-quality 3D crystals, but many proteins do not easily form crystals. This difficult and time-consuming crystallization step can be avoided by the nuclear magnetic resonance (NMR) method, which can determine protein structures in solution. However, the NMR method is applicable only to small proteins, which does not allow for the study of large multi-protein complexes. Transmission electron microscopy (TEM) provides an alternative approach for 3D structure determination.

TEM can work on either crystallized or non-crystallized samples. Instead of 3D crystals like those used for X-ray crystallography, two-dimensional (2D) crystals are the starting point for electron crystallography. The scattering ability of an electron beam is much higher than that of an X-ray, therefore, only a single layer of protein crystals (~ 10 nm, or one molecule thick) is needed for electron crystallography. If the 2D crystal is large enough and well ordered, the resolution of electron diffraction can reach 2 Å (Gonen et al., 2005). To achieve the same resolution from X-ray diffraction would require a 3D crystal 100 μm thick.

From raw X-ray diffraction patterns, only the intensity of the spots can be measured. The structural amplitude can be obtained directly, but the structural phase is lost. Several approaches are available to retrieve the structural phase indirectly, but all require crystals of high enough quality to diffract X-ray to high resolution. If the crystals are not good enough, no structural information can be extracted. On the other hand,

TEM can produce both diffraction patterns and images of 2D crystals as the electron beam can be focused in the electron magnetic field of the microscope lens. By combining the structural phase from the image and the structural amplitude from the diffraction pattern, the protein structure can be calculated directly. Even when 2D crystals are small and not perfect, valuable information can be extracted by electron crystallography. In the following sections, I first discuss the advantages of using the lipid monolayer technique for 2D crystallization of protein complexes, and then review the principles and advances of 3D reconstruction from electron micrographs. I conclude with a brief description of practical applications of these methods for several biomedically-relevant macromolecules.

2D crystallization by the lipid monolayer approach

The lipid monolayer method is a promising approach for 2D crystallization of proteins (Chang and Kornberg, 2000). This method requires only a small amount of protein at very low concentrations (less than 0.1 mg/ml) compared to >10 mg/ml usually required for 3D crystallization. There are also many other benefits of this approach that are related to the unique characteristics of phospholipids.

Phospholipids are amphiphilic molecules, each having a hydrophilic head and hydrophobic tails. In aqueous solution, the hydrophobic tails of phospholipids have the tendency to aggregate together to reduce surface exposure to water. When phospholipids in organic solvents are layered on the water surface, the lipids will spontaneously form a single layer with the hydrophobic tails facing air and hydrophilic heads facing water. If

positively or negatively charged groups are present in the lipid heads, opposite charged protein molecules in solution may be attracted to the lipid heads and will be concentrated there. The protein concentration can reach 100 mg/ml locally from a 0.1 mg/ml initial concentration.

For 3D crystallization, high initial protein concentrations and high salt concentrations are needed; however, many proteins precipitate at such high concentrations. In 2D crystallization using the lipid monolayer technique, the salt concentration and the initial protein concentration is much lower. Other benefits are that the experiment can be performed at low temperature (4 °C), the conditions for protein molecules to bind to the lipids are not harsh and will not denature the proteins, and the proteins become over-saturated locally at the lipid surface in their native conformation. Within the lipid monolayer at the air-water interface, each lipid molecule is free to move and rotate. Protein molecules attached to the lipid can also move and rotate. In these conditions, there is a good chance that the protein molecules will form crystals. For those proteins that are difficult to express, difficult to purify in large amounts, or easily denature in the presence of high salt, the lipid monolayer method is an attractive alternative for growing crystals.

One recent improvement to the lipid monolayer technique uses lipids with a modified head group that binds a target protein with high specificity/affinity (Chiu et al., 1997). This approach is similar to affinity chromatography, which utilizes the specific interaction between proteins and their ligands for purification. In order for the high-affinity lipid monolayer technique to be generalized for protein crystallization, we need

an easy and efficient way to produce proteins and lipids that specifically interact with each other. This process can be accomplished using the commercially available Ni-NTA (nickel-nitrilotriacetic) modified phospholipid, which binds proteins that carry a tag of six adjacent histidine residues specifically and with high affinity (Figure 1-1).

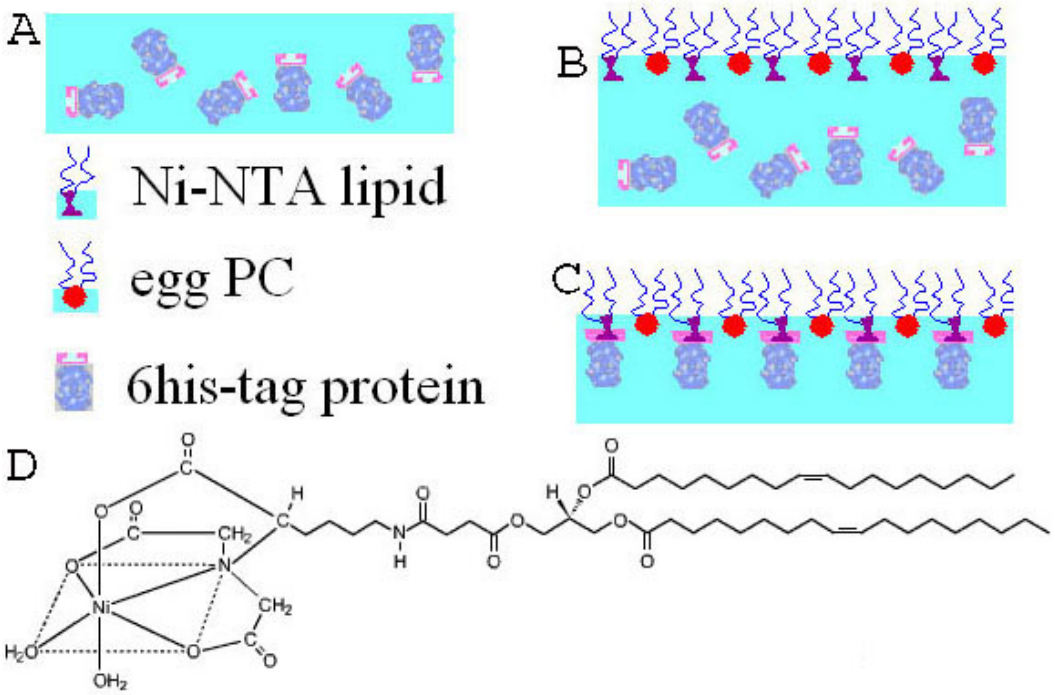

Figure 1-1. Diagram of 2D crystallization on lipid monolayer. (A) A Teflon well is loaded with a 6His-tagged protein solution. (B) Ni-NTA lipid and egg phosphocholine (PC) in chloroform/hexane is layered on top of the protein solution. The lipids form a monolayer at the air-water interface. (C) The protein molecules bind to the Ni-NTA lipid head group resulting in an increase in the local protein concentration at the lipid surface and the eventual formation of 2D crystals. (D) Structure of Ni-NTA-DOPS from Avanti Polar Lipids (Dang et al., 2005). Its full name is: 1,2-Dioleoyl-*sn*-Glycero-3-[(N-(5-amino-1-carboxypentyl) iminodiacetic acid) succinyl] (Nickel salt).

3D reconstruction from electron micrographs

The 3D structure determination of biological macromolecules from electron micrographs was started from the seminal paper of DeRosier and Klug (1968), which

provided the foundation for further development of more efficient algorithms and software. Since then, 3D reconstruction from 2D projections is not only used in scientific research but has many applications, such as computed tomography (CT) in medical diagnoses (Doi, 2005). Aaron Klug was awarded the Nobel Prize in 1982 for his contribution to the development of electron crystallography to elucidate the 3D structure of biologically important macromolecules.

3D reconstruction from electron micrographs is the reverse process of image formation in TEM. TEM can achieve very high magnification of an object, but what we can get is only a 2D projection of the 3D object. All of the atoms at different levels in the Z-direction are projected onto one plane and the information in the Z-direction is lost. To restore the information in the Z-direction, many projections from different orientations are needed. The number of projections required for a 3D reconstruction depends on the size of the object and the required resolution. The basic principle for 3D reconstruction from 2D projections is the "central section theorem", which states that the Fourier transformation of a 2D projection is a central section of the Fourier transformation of the 3D object that produces the 2D projection (DeRosier and Klug, 1968). If enough 2D projections at different orientations are collected, the central sections can fill the whole Fourier space (Amos et al., 1982). A 3D map of the object can be generated by interpolation of the missing data in Fourier space and applying 3D Fourier inversion.

There are several ways to collect the required data for 3D reconstruction in TEM, depending on the symmetry of the specimen. In addition to ultra-thin sectioning of thick

specimens that are used for electron tomography, many different kinds of macromolecule assemblies can be used for 3D reconstructions. The range of specimens that can be studied by TEM is nearly unlimited. Four kinds of specimens are summarized below: 2D crystals, helix filaments, icosahedral viruses, and low symmetry protein complexes.

(1) The best specimen for TEM is the 2D crystal, in which all the protein molecules are in the same orientation and ordered positions within a 2D plane. The only way to collect electron micrographs at different orientations of a 2D crystal is by tilting the specimen stage, however, it is difficult to record good images at high tilt angles (Sali et al., 2003). Furthermore, the highest tilt angles (between 75° to 90°) cannot be recorded in TEM. The Fourier space therefore has a "missing cone", which causes lower resolution in the Z-direction in the final 3D reconstruction as compared to the X- and Y-directions (Baumeister and Steven, 2000). Although there are still many inherent difficulties for the 2D crystal approach, electron crystallography has been used successfully to determine several protein structures near 3Å resolution, the highest attainable for the TEM method (Fujiyoshi, 1998; Walz and Grigorieff, 1998).

(2) The second preferred specimen for 3D reconstruction is a filament structure with helical symmetry. Some proteins assemble into helical filaments naturally, like actin, tubulin and flagellin (Padron et al., 1998; Amos, 2000; Yonekura et al., 2005), whereas others can artificially form helices (Unwin, 1998). In a perfect helix, subunits are positioned exactly around a central axis with a fixed angle and distance. All the orientations of the subunits required for 3D reconstruction are already present in a single

image of a helical filament, so there is no need to tilt the specimen. The Fourier space is covered evenly, and no "missing cone" problem exists. Data collection for a helical structure is relatively easier than for a 2D crystal, but the analysis of helical symmetry is more complex. In one recent report, 3D reconstruction of a helical structure had reached 4 Å resolution (Unwin, 2005).

(3) Viruses are huge macromolecule assemblies, but still lie beyond the detection limit of light microscopy. TEM has long been a primary tool for the identification, classification, and structural characterization of viruses. Icosahedral viruses, which have a defined set of symmetry elements that consist of six 5-fold axes, ten 3-fold axes, and fifty 2-fold axes, are the most favorable specimens for TEM studies (Baker et al., 1999). 3D reconstructions from TEM images have contributed in a large part to our understanding of viral structural biology, including the morphology of viruses, the stoichiometry and organization of their subunits, the interactions between their subunits, and the assembly of viruses.

(4) 3D reconstruction of large protein complexes without symmetry or with low symmetry was initially carried out by tilting one single protein molecule in TEM (Hoppe et al., 1974), just like collecting data from 2D crystals by electron crystallography. The resolution of such reconstructions was very low, because of serious radiation damage. Because purified protein molecules can be thought of as identical (although heterogeneity does exist), merging data from many molecules with the same conformation can improve the signal-to-noise ratio. Many different algorithms and computer program packages have been developed to align the particles (molecules) and

merge the data (Frank, 1975; Frank, 1996). This process is equivalent to moving the rate-limiting step of crystallization from “wet bench” to “in silico”. With faster computers and more efficient software, single particle 3D reconstruction from electron micrographs has become a powerful approach toward solving the structures of large protein complexes. Two protocols of data collection and image processing are described below.

Single particle reconstruction by random conical tilt

Biological macromolecules on carbon support film or in vitrified ice may have preferred orientations, or they may orient randomly (Frank, 1996). If only one or a few orientations exist, a data collection protocol called “random conical tilt” can collect enough orientations with minimal radiation damage. Each area is recorded twice in TEM: the first image is tilted at 45° and the second is not tilted (Figure 1-2A).

The zero-tilted image is processed first to classify all the particles into different groups according to their orientations. Only particles in one orientation are used for each 3D reconstruction. The particle coordinates from the zero-tilted image are used to locate the same particles in the tilted image. The in-plane rotation angles (azimuths) obtained from aligning the particles from a zero-tilted image are applied to the particles from a tilted image and their orientations in the 3D Euclidian space (as defined by Euler angles) are determined. The Euler angles of the particles from the tilted image will form a cone with a fixed tilt angle and random azimuths around the cone (Figure 1-2B). The Fourier Transform of these particles can cover all the Fourier space except a small “missing

cone". All the later processing is actually performed on the particles from the tilted image, which is exposed to the electron beam only once, so the radiation damage is minimal. The SPIDER software package, which is based on the above philosophy, has been successfully applied to the 3D reconstruction of numerous protein complexes (Penczek et al., 1994; Guenebaut et al., 1997). The best-known example is the 3D reconstruction of the 70S ribosome (Figure 1-2C).

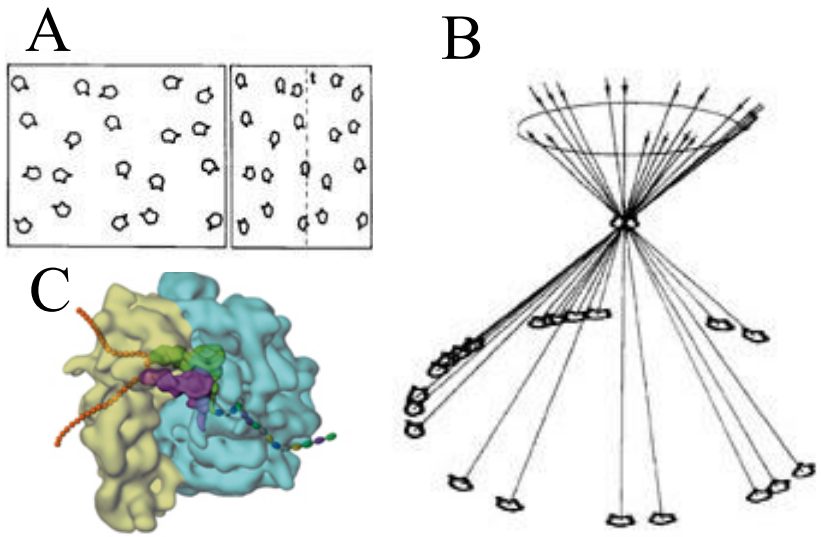

Figure 1-2. Random conical tilting. The principle of random conical tilting is explained in (A) and (B) (Radermacher et al., 1987) (A) Untilted (left) and corresponding tilted (right) projections of protein particles with the tilt axis indicated by a dashed line. (B) In the untilted image, the particles have the same orientation at the Z-axis, but have random rotation in the plane of X- and Y-axis. In the tilted image, the Euler angles of the particles form a 'cone' around the Z-axis. (C) Example of a 3D reconstruction using random conical tilt method: the 70S ribosome (Gabashvili et al., 2000).

Angular reconstruction from zero-tilt images

If the protein molecules (particles) are oriented randomly in TEM, all the orientations that are needed for 3D reconstruction are already present in the zero-tilted

image. The problem is to find the relative orientation of each particle (Euler angles). Marin van Heel (1987) proposed the "angular reconstruction" algorithm based on the "common-line-projection theorem". This real space theorem is equivalent to the "central section theorem" in Fourier space. It states that two different 2D projections of a 3D object always have a 1D projection (line) in common. 2D projection can be calculated from a 3D object; in the same way, we can calculate 1D projection from a 2D image. Stacking all the 1D projections can produce an artificial image called a "sinogram", in which the X-axis is the same as the original 2D image and the Y-axis is the tilting angle from 0° to 360° (the direction that produces the 1D projection). Any two sinograms must have at least one line in common according to the common-line-projection theorem. Comparing each line in one sinogram with each line in another sinogram produces a "sinogram correlation function", which can also be visualized as an artificial image. The position of the maximum value in the "sinogram correlation function" corresponds to the common line of the two sinograms, from which the relative orientation (angles) between the 2D projections can be determined. For asymmetric particles, three 2D projections (not related by tilting around a single axis) are required to determine their relative orientation (Euler angles).

The IMAGIC software package based on this strategy has been successfully applied to the 3D reconstruction of several macromolecules (van Heel et al., 2000). The program uses a robust algorithm to classify the particles in different orientations, and to align and average particles in each class. Euler angles are assigned by the sinogram method to the averaged particles, which have a much higher signal-to-noise ratio than

raw particles and reduces the risk of assigning the wrong Euler angles (van Heel et al., 1996). Another software package (EMAN) was developed using similar strategies to IMAGIC, but many processes have been automated and have graphical user interfaces (Ludtke et al., 2004).

Perspective of single particle method

Although crystallography methods are still the first choice if the proteins can be crystallized, we may depend more on single particle 3D reconstruction in the future. The simplicity of sample preparation and data collection has paved the way for automation. The only disadvantage of single particle 3D reconstruction is the demand for a large amount of computer time. With faster computers and more efficient software, single particle 3D reconstruction from zero-tilted images will have more impact on determining the structure of complex biological macromolecules. The resolution of single particle 3D reconstruction is not as high as crystallography or NMR, but is improving constantly.

One benefit of the single particle method is that the dynamic states of proteins are more easily captured. Crystallizing proteins in one conformation alone can be difficult and it is less likely that other conformations and intermediate states can be captured. In solution, the dynamic conformational changes of the proteins that occur in a very short time after substrate binding can be captured by fast freezing and imaging them at liquid nitrogen temperature. The cryo-electron microscopy (cryo-EM) single particle technique, combined with cryo-sectioning and electron tomography, has the

potential to solve the complex biological structures, from whole living cells down to a single macromolecular machine (Baumeister and Steven, 2000).

CTF corrections

The contrast transfer function (CTF) is part of the microscope transfer function that modifies the amplitudes and phases of the electron micrographs (Misell, 1978). The density distribution in the images may therefore not represent the true structure of the original object. Depending on the choice of defocus, the image contrast reverses at different spatial frequencies. When reconstruction is performed at low resolution from electron micrographs that are imaged at close to focus, there is no need to correct the CTF. But when aiming at higher resolution (e.g. < 15 Å) reconstruction, the CTF shows oscillations resulting in contrast reversals necessitating the correction of the CTF effects. CTF correction of electron micrographs may be the most difficult step in single particle 3D reconstruction. In electron crystallography and helical reconstruction, CTF correction is relatively easy, as the diffraction peaks of the 2D crystal or helix are discrete, whereas noise and CTF are continuous. The Fourier transform of a single particle image is continuous, so it is impossible to separate the structural information of the particle from noise and apply the CTF correction. The EMAN program adapts a strategy to manually find the eight parameters (A, C_A, ΔZ, B, and n1 through n4) for CTF correction by fitting a power spectrum curve of each particle set in a graphical utility, using a 1D structural factor as a constraint (Ludtke et al., 1999). After phase flips, the amplitude correction is performed by a weighted average of the images in Fourier space, where the weights vary

with spatial frequency according to the eight parameters determined. When an image contains no information at a particular spatial frequency (zero in CTF), it does not contribute to the final average.

The 1D structural factor can be determined experimentally by performing a solution X-ray scattering of the target protein, which requires a high concentration of protein and special equipment. Another approach is to calculate the 1D structural factor from several data sets at different defocus (Ludtke et al., 1999). This method assumes that the structural factor is constant for the same protein imaged at different defocus. As long as the distributions of particle orientations remain consistent, any micrographs of the same protein will have the same 1D structural factor. Another assumption is that the high frequency portion of the 1D structural factor is similar for different proteins with similar folding. The final 1D structural factor is produced by combining the low frequency portion calculated from the target protein cryo-EM images with the high frequency portion from a protein whose structure factor had been determined experimentally.

Single particle 3D reconstruction from cryo-EM images of GroEL, whose crystal structure has also been solved by X-ray crystallography, matches very well with the crystal structure (Ludtke et al., 2004). GroEL has three domains, which are proposed to have domain-domain rotation and movement during substrate binding. In crystals, GroEL molecules are locked in one conformation. Cryo-EM solution structures have captured conformational change during ATP and substrate protein binding (Ranson et al., 2001).

Studies of Biomedically Important Macromolecules

In cooperation with several laboratories, I have successfully crystallized several biomedically important protein complexes (including Nsp15 from SARS virus and TLR3 ECD from humans) with the lipid monolayer method. Each protein was expressed with a 6His-tag, purified with Ni-NTA chromatography, and crystallized using Ni-NTA lipids. Single particle 3D reconstruction was also applied, as this approach can yield fast, reliable results without the need for crystallization. This thesis first briefly introduces IniA from *Mycobacterium tuberculosis*, and then focuses on Nsp15 and TLR3. Several other projects, including studies of the circadian clock proteins KaiC and KaiA from cyanobacteria and complexes of membrane proteins with the molecular chaperonin GroEL are also discussed.

IniA from Mycobacterium tuberculosis

M. tuberculosis is an important pathogenic bacterium that causes millions of deaths worldwide each year (Smith and Sacchettini, 2003). Even more serious is the emergence of drug-resistant bacilli. More research is needed to find new drugs to combat the bacteria (Wade and Zhang, 2004). Traditional drugs are found by chance or by screening millions of different chemicals, either extracted from natural sources or artificially synthesized (Sun and Cohen, 1993). Although this kind of blind search has led to many important discoveries to treat various human diseases, it takes a long time to find a possible candidate.

Structure-based drug design adopts a new strategy and provides another way for drug discovery (Klebe, 2000). First, the pathogenesis process of a human disease is studied and potential drug targets are identified. The next step is to clone the genes, purify the proteins, study their biochemical characteristics, and determine their structures. The third step is in silico virtual screening to dock ligands into the binding sites of the target protein. This process can narrow down the potential drug candidates from millions to tens. New molecules can be designed according to the structure of the target protein. Lastly, the effects of the drugs are tested at the molecular and cellular level, and finally in model animals and in human population.

In *M. tuberculosis*, the IniA protein is a potential target for the structure-based drug design strategy. When treating *M. tuberculosis* with the antibiotic isoniazid (INH), the expression level of IniA was increased greatly (Alland et al., 1998). INH is an inhibitor of *M. tuberculosis* cell wall biosynthesis. From amino acid sequence analysis, no homologs can be found and no function can be inferred for the IniA proteins. Overexpression of the iniA gene in *Mycobacterium bovis* BCG confers tolerance to both INH and ethidium bromide (Colangeli et al., 2005). Reserpine, an inhibitor of multidrug resistance pumps, reverses the tolerance to INH and ethidium bromide in BCG (Colangeli et al., 2005). These results indicate that IniA may be a pump component and participate in the development of tolerance to INH in *M. tuberculosis*.

Inhibitors of IniA may be used as new drugs to treat resistant bacteria and aid in combating tuberculosis. Determining the structure of IniA will be a key step in finding inhibitors of IniA. To this end, we have expressed IniA in *E. coli* and purified the protein

by affinity chromatography. Different methods were attempted to solve the structure of IniA.

Nsp15 from SARS virus

Severe Acute Respiratory Syndrome (SARS) virus, a newly identified coronavirus, caused a worldwide pandemic in 2003 (Wenzel et al., 2005). It is a deadly threat globally, as no effective treatment is available. Since the studies of coronavirus have a long history, comparative analysis of the SARS virus with other coronaviruses gave many clues to its molecular mechanisms and pathogenesis. Like other coronaviruses, SARS is an enveloped virus with a long positive-stranded RNA genome (Snijder et al., 2003). Its genome is ~30 kb long with 14 open reading frames (ORFs). ORF '1a' encodes the polyprotein '1a', and a frameshift produces the longer polyprotein '1ab' (Figure 1-3A). Polyproteins '1a' (> 4000 amino acid residues) and '1ab' (> 7000 amino acid residues) are processed by two viral proteinases to produce 16 replication-related enzymes, called non-structural proteins (Nsp) 1-16. The ORFs downstream of ORF1b encode the structural and accessory proteins that are synthesized from subgenomic mRNAs. These subgenomic RNAs have the same 5' and 3' termini as the viral genome. It is still not clear how the subgenomic RNAs are produced.

The functions of most Nsp proteins are unknown. As the genomic sequence of SARS is available, sequence analysis can provide some insight. Comparative genomics of the coronaviruses had identified several distant homologs of cellular enzymes that are involved in RNA processing (Snijder et al., 2003). Of these, Nsp15 is most closely

related to the *Xenopus laevis* endoribonuclease XendoU (Figure 1-3C). Biochemical experiments showed that Nsp15 is a Mn^{2+}-dependent endoribonuclease that preferentially cleaves uridylates of RNAs (Bhardwaj et al., 2004). The Nsp15 orthologs are conserved among all coronaviruses, and are predicted to be important enzymes in the life cycle of these viruses. Characterizing the biochemical and biophysical properties of Nsp15 is essential to understand the role of Nsp15 and may lead to effective treatments of the virus. The oligomeric state and 3D structure of Nsp15 is the focus of this study.

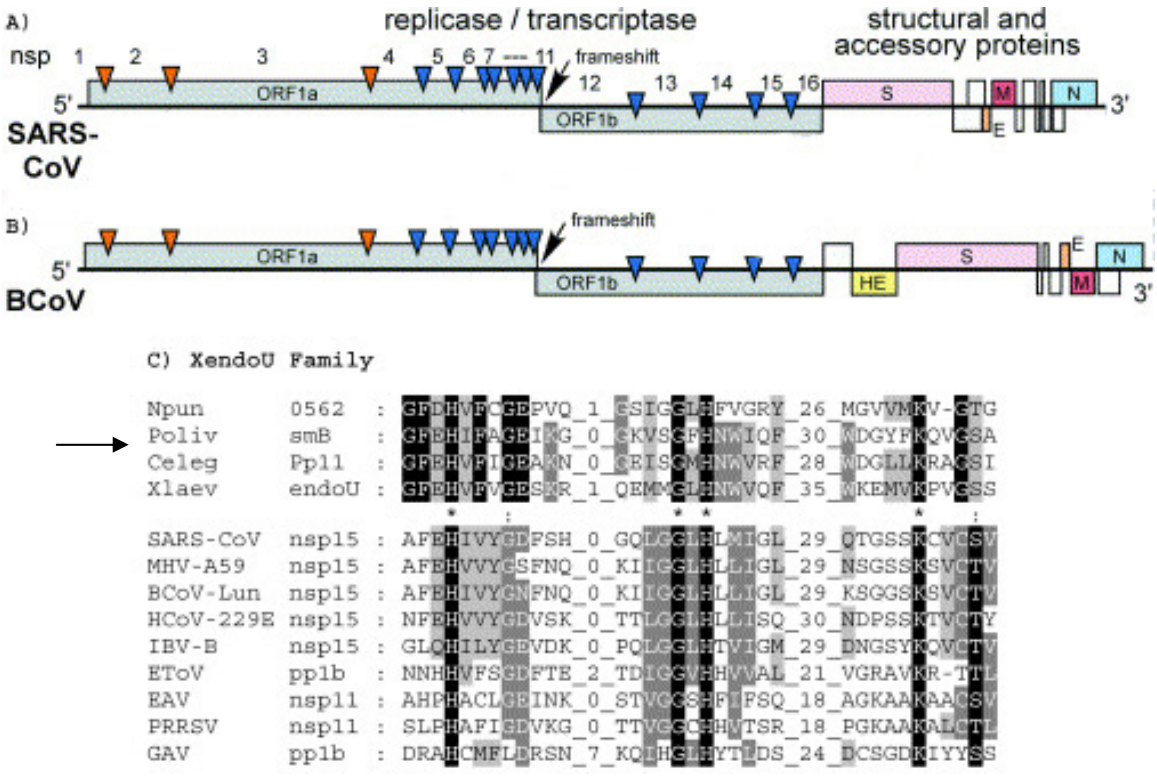

Figure 1-3. SARS virus and Nsp15. (A) The genome organization of the SARS virus and (B) the bovine coronavirus (BcoV). (C) Sequence alignments of SARS Nsp15 (arrow) with other enzymes involved in RNA processing (Snijder et al., 2003).

Several lines of experimental evidence, including gel filtration, analytical ultracentrifugation, native gel and UV crosslinking/SDS-PAGE studies, indicate that Nsp15 exists as both monomers and hexamers (Guarino et al., 2005). A single amino acid substitution (K289A) causes the protein to predominantly exist as hexamers, whereas another substitution (E3A) renders it mainly monomers. The molecular mass of the Nsp15 monomer is 40 kDa and a hexamer is 240 kDa, which is large enough to perform TEM single particle 3D reconstruction. Not only can the oligomeric state be determined from TEM reconstruction, but also the subunit organization can be visualized directly. Because Nsp15 was expressed and purified with a 6His-tag, 2D crystallization was also attempted with the lipid monolayer method using the Ni-NTA lipid.

Receptors for pathogenic microbes

Mammalian immunity has two arms, the innate and adaptive immune systems, both of which are important to protect the host against infections of pathogenic microbes (Diamond et al., 2003). In the adaptive immune system, there are nearly unlimited kinds of T- and B-cell receptors, produced by somatic recombination and mutation of the receptor genes (Livak et al., 2000). Each kind of T- or B-cell receptor recognizes one type of antigen with high specificity/affinity, but recognizing antigens itself is not enough to activate the T- or B-cells. Activation of T- or B-cells also requires co-stimulatory molecules induced by the innate immune system (Pasare and Medzhitov, 2003). A limited number of receptors in the innate immune system, called pattern-

recognizing receptors (PRRs), can sense pathogens and initiate the inflammatory responses. The most important PRRs are Toll-like receptors (TLRs), which recognize pathogen associated molecule patterns (PAMPs). More than ten different TLRs have been identified in the human genome, called TLR1-10 (Bowie and Haga, 2005). TLR4 recognizes lipopolysaccharide (LPS, also called endotoxin) of Gram-negative bacteria, whereas TLR1, 2, and 6 respond to lipoprotein of Gram-positive bacteria. TLR5 is the receptor for bacterial flagellin, whereas TLR3 responds to double stranded RNA, TLR7 and 8 to single stranded RNA, and TLR9 to DNA of bacteria and viruses that contains unmethylated CpG (Janssens and Beyaert, 2003).

The discovery of TLR is an interesting story. A mutation that disrupted the dorsal-ventral patterning during *Drosophila* embryonic development was named Toll (Anderson et al., 1985). Many components of the Toll signal transduction pathway, including a transcriptional activator similar to nuclear factor kappa B (NF-κB), resemble components of the interleukin 1 (IL-1) receptor signal transduction pathway, which activates the inflammatory response in mammals. This similarity led to the proposal that Toll might be involved in immune responses of *Drosophila*. Mutant flies that lack the Toll pathway components were indeed highly sensitive to fungal infection (Lemaitre et al., 1996). The important role of Toll in *Drosophila* immunity prompted a search for human homologs of Toll protein (Medzhitov et al., 1997). TLR4 was identified as the long-sought receptor for endotoxin, a lethal agent of Gram-negative bacteria. This discovery stimulated great interest from both the academy and pharmaceutical companies to study the TLR pathway and find new therapies targeting this pathway.

During the last seven years, rapid progress has been made in understanding endotoxin-TLR4 signaling and the molecular basis of the lethal effect of endotoxin. Upstream of TLR4, MD-2 is needed for endotoxin recognition (Ishihara et al., 2004). Downstream of TLR4, a MyD88 dependent pathway leads to the activation of NF-κB, and also induces cytokines, chemokines (Palsson-McDermott and O'Neill, 2004). The inflammatory responses of the endotoxin-TLR4 pathway are important for host defense against Gram-negative infection, but overwhelming inflammatory responses may lead to endothelial damage, coagulopathy, tissue hypoperfusion, and multiple-organ failure (Miyake, 2004). Proper response of TLR4 signaling to endotoxin is critical. Animals that cannot sense endotoxin may die if infected by Gram-negative bacteria. Animals that respond too vigorously to endotoxin may also die because of septic shock. It is estimated that 20,000 people die each year in the United States as a result of septic shock caused by Gram-negative infection (Pinner et al., 1996). Understanding the molecular mechanism of the TLR4 signaling pathway may provide an opportunity for therapeutic intervention.

Many polymorphisms of TLR genes exist in the human population, especially single nucleotide polymorphisms (SNPs) (Cook et al., 2004). It is not surprising that mutations in such critical genes have serious effects, such as increased susceptibility to bacterial and viral infections (Figure 1-4). Indeed, polymorphisms of TLRs have been related to many human diseases, including sepsis, immunodeficiencies, atherosclerosis and asthma (Cook et al., 2004). The complexity of many human diseases demands

multiple approaches to study the pathogenic process. Structural and functional analysis of purified proteins may give some clues as to how they work in humans.

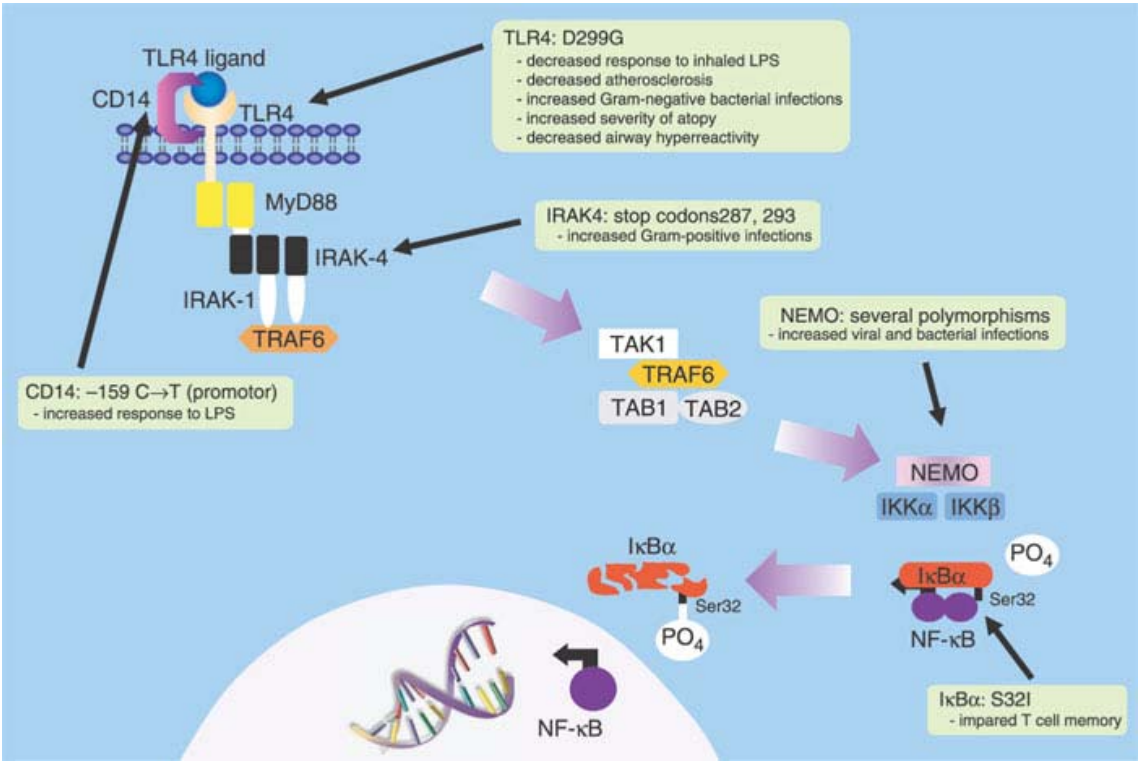

Figure 1-4. Polymorphisms of TLR4 signaling molecules. The polymorphic genes and associated human diseases and phenotypes in the TLR4 signaling pathway are shown (Cook et al., 2004).

Human TLR3

It has long been realized that animal cells infected with virus can produce interferons that inhibit virus replication (Matsumoto et al., 2004). Later it was shown that it is the double-stranded (ds) RNAs of the virus that induce this response (Isaacs et al., 1961). Synthetic dsRNA, polyriboinosinic polyribocytidylic acid or polyIC, can also

induce the production of interferons (Matsumoto et al., 2004). How animal cells recognize dsRNA and stimulate anti-viral responses remained unanswered for more than 40 years. The discovery that TLR4 is the receptor for bacterial endotoxin led Schulz et al. (2001) to test all of the TLRs in the human genome for their response to dsRNA. It was found that TLR3 is the only receptor among the TLRs that responds to dsRNA. The *in vivo* role of TLR3 may be to induce maturation of dendritic cells (DCs) of the innate immune system when viral dsRNA is sensed (Schulz et al., 2005; Salio and Cerundolo, 2005). The maturation of DCs is characterized by increased expression of co-stimulatory molecules and secretion of cytokines, which in turn activate virus-specific T-cell responses.

Like other TLRs, human TLR3 has an extracellular leucine-rich repeat (LRR) domain and a cytoplasmic Toll-Interleukin 1 receptor (TIR) domain, connected by a single transmembrane helix (Matsumoto et al., 2004). Full-length TLR3 is difficult to purify in large amounts for biochemical and structural studies, so this thesis focuses on the TLR3 extracellular domain (ECD). 2D crystallization and single particle 3D reconstruction were applied to study the structure of the TLR3 ECD. A large horseshoe shaped molecule was visualized by both methods. In solution, TLR3 ECD was shown to exist as a monomer by single particle analysis. From the 2D crystal projections, several possible dimeric forms might exist. Through bioinformatic analysis, site-directed mutation, and cell-based activity assays, the functional role of N-glycosylation of TLR3 was studied. Several SNPs of TLR3 in the human population were also characterized by the cell-based functional assay.

CHAPTER II

3D RECONSTRUCTION OF INIA FROM *MYCOBACTERIUM TUBERCULOSIS*

Introduction

Tuberculosis caused by the Gram-positive *Mycobacterium tuberculosis* is the leading cause of death in the world. The disease affects about one-third of the entire world population (Zhang, 2005). In AIDS patients, it is the most important and deadly agent among the opportunistic bacterial infections. Repeated infection and prolonged use of drugs to treat tuberculosis in AIDS patients have given rise to drug-resistant bacteria. The HIV virus can only be transmitted through sex or blood transfusion, but *M. tuberculosis* may be transmitted through the air (Castro, 1995). If these drug-resistant bacteria become prevalent, it would be a great threat to public health. New drugs are urgently needed to prevent the transmission and proliferation of the resistant bacteria. More research is needed to understand the biological processes of bacterial infection and find new drug targets. IniA may be such a drug target (Alland et al., 2000). The expression level of IniA is greatly increased after exposure of the bacillus to the antibiotics isoniazid and ethambutol (Alland et al., 1998). Genetic and biochemical studies indicate that IniA may be part of an efflux pump that transports these drugs out of the bacteria (Colangeli et al., 2005). In order to further understand its structure and mechanism, we collaborated with James Sacchettini's and David Alland's labs to clone

and express a 6His-tagged IniA of *M. tuberculosis* in *E. coli* and purify it with Ni-NTA affinity chromatography.

Results

Expression and purification of Mycobacterium tuberculosis IniA

IniA is expressed in very low amounts in its native host and special requirements are needed for handling laboratory strains of *M. tuberculosis*. It would be difficult, if not impossible, to express and purify IniA from *M. tuberculosis* cultures. For structural studies, a large amount of purified protein is needed. The easiest approach is to clone the target gene, insert it into an expression vector, which will express the protein at high levels from *E. coli*, and purify the protein with affinity chromatography.

Low temperature induction (18 °C) is needed to express IniA in soluble form. Cell growth and protein translation may slow down at low temperature, allowing proteins to fold properly and avoiding their aggregation into inclusion bodies (Hidari et al., 2005). SDS-PAGE shows that even though IniA is predicted to have a single trans-membrane domain, a ~ 73 kDa band appears in the supernatant after high-speed centrifugation of French press lysates (Figure 2-1). Ni-NTA chromatography with step-wise washes of 50 and 100 mM imidazole followed by elution with 200 mM imidazole yields protein that is 95% pure as judged by SDS-PAGE. The 100 mM imidazole wash also elutes some IniA, but contains some low molecular mass contaminates. This protocol for the expression and purification of IniA is very efficient, and one-step purification had been achieved.

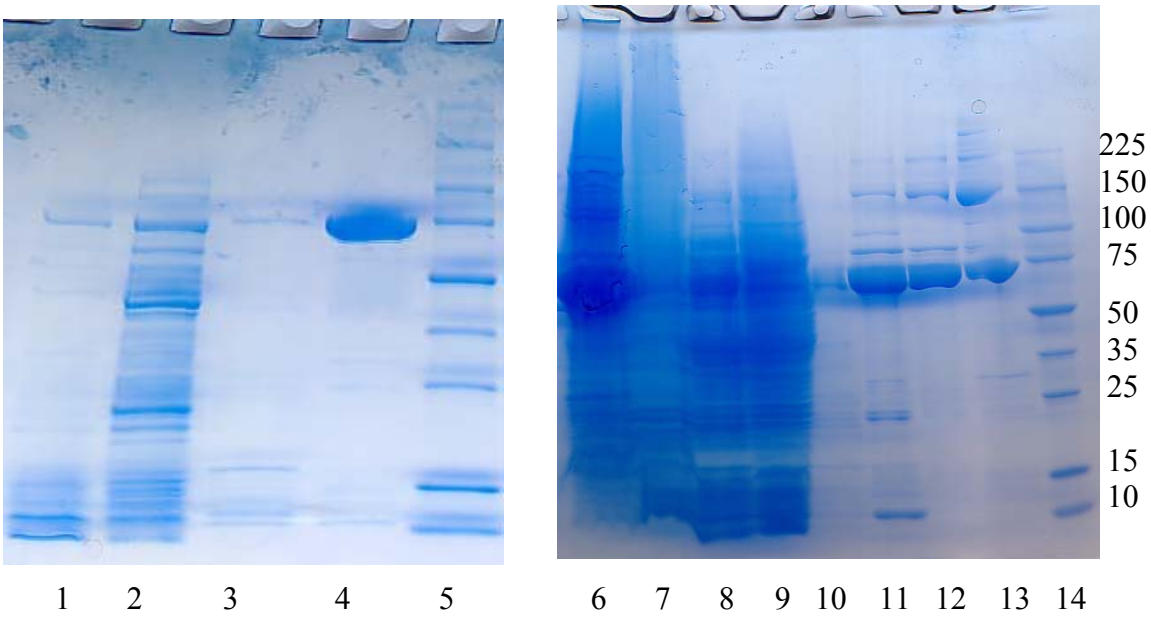

Figure 2-1. SDS-PAGE of IniA. Lanes 1-4: boiled the sample for 1 minute. Lanes 6-13: incubated for 30 minutes at 50 °C. (1) Supernatant after the first high-speed spin. (2) Ni-NTA column flow-through. (3) 50 mM imidazole wash. (4) 200 mM imidazole eluate. (5) Molecular marker. (6) Low-speed pellet after French press passage. (7) High-speed pellet and (8) corresponding supernatant. (9) Flow through. (10) 50 mM imidazole wash. (11) 100 mM imidazole eluate. (12) 200 mM imidazole eluate (peak fractions, first half). (13) 200 mM imidazole eluate (peak fractions, second half). (14) Molecular marker. The masses of the molecular markers are given in kDa.

Boiling or not boiling the sample in the loading buffer has a profound effect on the mobility of IniA in SDS-PAGE. Purified IniA appears as a single band when the sample is boiled for 1 minute, but as multiple bands if the sample is incubated at 50 °C for 30 minutes before loading. The higher molecular weight bands of the non-boiled sample correspond to SDS-resistant but heat-sensitive dimers or higher oligomers of IniA. Longer boiling time results in no band at all.

For growth of 3D crystals, high protein concentration is required. Purified IniA was unstable and precipitate easily during the buffer exchange and concentration steps.

After many trials, we found that IniA could be concentrated to ~10 mg/ml in 40% glycerol. Several hundreds of conditions were tried with crystallization screen kits (Hampton Research) either automatically using a crystallization robot (Hydra-one-plus, Robin Scientific Instruments) or manually, using the sitting drop, hanging drop or microbatch method. IniA forms obvious precipitation in most conditions. We did see several small crystals that were more likely to be salt crystals, as they appeared in a short time and did not grow larger. Further more these crystals did not absorb the protein-crystal test dye (izit, Hampton), indicating that the molecules in the crystal are very tightly packed and do not allow penetration of the dye in the crystals. Cracking the crystals with a needle also indicated that the crystals might be salt.

3D reconstruction of IniA

Triangular-shaped particles of IniA were sometimes observed on carbon films in TEM (Figure 2-2). The particles resembled wheels but with sharp edges. Each triangular particle has three cavities around a central density. Some long, square-shaped particles were also observed, which may represent the sideview of the protein. 3D reconstruction from 1000 particles with both the EMAN and IMAGIC single particle analysis software packages successfully built a 3D model of the triangular-shaped IniA. The resolution is 28 Å from EO-test. Figure 2-2D is a surface rendering of the 3D volume with a molecular mass threshold of 438 kDa. If the mass threshold is set to 219 kDa, as is expected for a trimer (IniA monomer is 73 kDa), some densities disappear in the 3D structure and the 3D model do not match the 2D class averages.

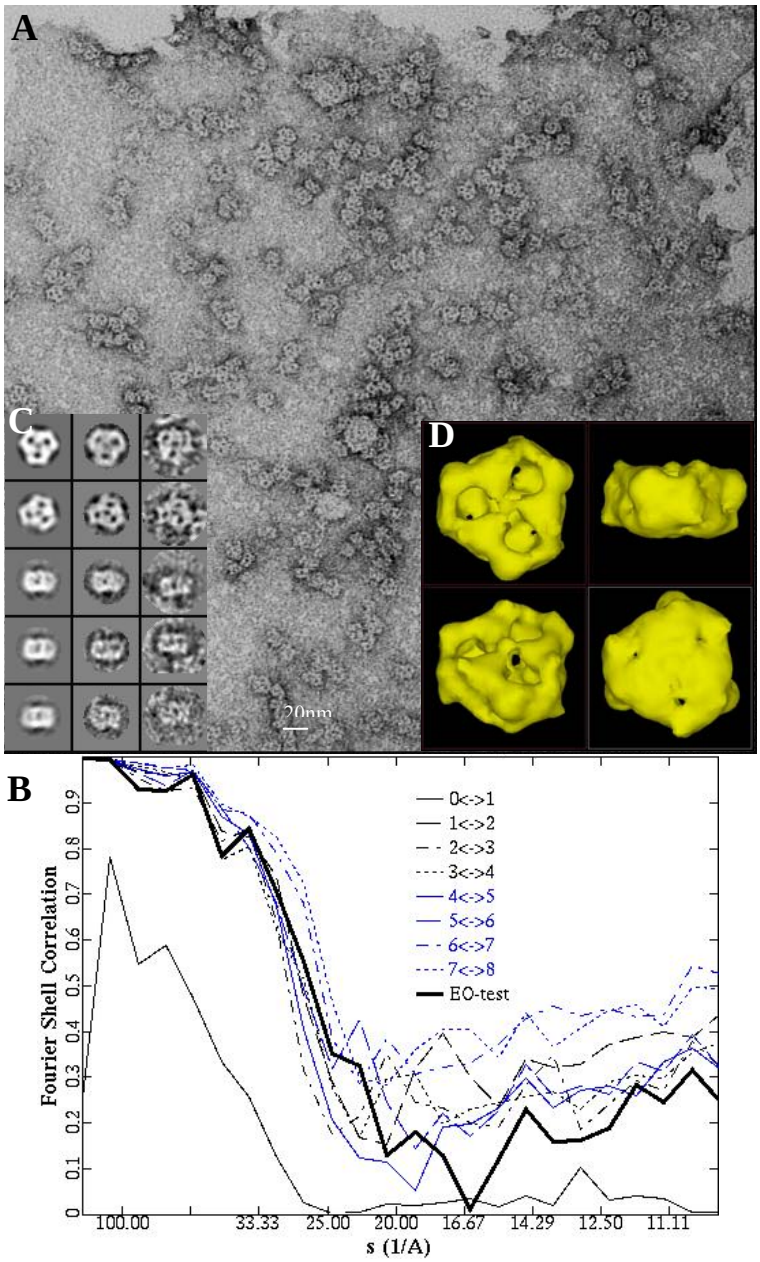

Figure 2-2. Electron microscopy of IniA. (A) Electron micrograph of negatively stained IniA single particles. (B) Fourier Shell Correlation (FSC) monitoring the 3D reconstruction process over 8 refinement cycles. EO-test is resolution assessment of the final reconstruction. (C) Right column: raw particles. Middle column: class averages. Left column: re-projections from the final 3D reconstruction. (D) Surface representation of the final 3D reconstruction highlighting different orientations. The threshold for surface rendering was set to correspond to a molecular mass of 418 kDa.

From the single particle 3D reconstruction, it is more likely that two IniA monomers first form dimers, and then three dimers form a triangular shaped hexamer. The trimer of dimers structure is very tight that the monomers could not be distinguished at this resolution. There are three cavities on top and bottom, but no opening on the side. The three cavities at the top are larger than at bottom and they are inter-connected, forming a big empty space in the center of the particle.

Discussion

To get large amount of proteins for structural and functional studies, expressing them in cultured cells with a tag and extracting them with affinity chromatography was the first choice. The expression level of IniA in *E. coli* was high, and purification by Ni-NTA column was quick and easy. From 2 liters culture, more than 10 mg IniA could be extracted with 95% purity in less than a week. The problem is the instability of IniA.

3D crystallization of the purified IniA was not successful after several years' efforts. Attempts to grow 2D crystals yielded small ordered patches, but they were not reproducible. NMR method could not be used for IniA because it is too large. Thus, the only 3D structure information of IniA obtained by now is from the TEM single particle analysis, showing that this method is very powerful and can be applied to problematic protein complexes.

The triangular shaped IniA seems to be a trimer in projection, but it is in fact a trimer of dimers. Hexameric IniA, with true 6-fold symmetry, was also observed (Colangeli et al., 2005). It is unclear how one protein can form different structures.

The IniA protein purified using the Ni-NTA column was easily denatured. The predicted transmembrane domain in the IniA sequence may be the reason for its instability in solution. In fact, it was unexpected that IniA could be purified in soluble form without detergent. However, there is another report of a membrane protein that can remain soluble in the absence of detergent. The S105 protein from Lambda phage, a three transmembrane domain protein, can form soluble ring-shaped structures under certain conditions (Savva, unpublished data).

One approach to increase the stability of the IniA protein is to remove the transmembrane helix, but the functionality needs to be tested to assure that the truncated protein retain its function. More genetic, biochemical, and biophysical studies are needed to characterize IniA and find conditions in which it is stable. The information obtained from such experiments will be of great value toward the crystallization of IniA and the subsequent high-resolution structural analysis.

It is possible to get high-resolution structure with single particle cryo-EM without the need for crystals, but the problem is still the instability of IniA. Even with negatively stained preparation, most time only aggregation or denatured protein was observed. The focus should be how to keep the protein stable.

Materials and Methods

IniA expression and purification

The *M. tuberculosis* H37RV iniA ORF (nt 410818-412756) was cloned into the BamH1/Nco1 sites of pET-30b, creating pET-iniA (Colangeli et al., 2005). Rosetta

(DE3) pLysS cells (Novagen, #70956) were used as the host for expression. Single colonies of the cells with pET-iniA were inoculated into 5 ml TB broth, grown at 37 °C overnight and diluted into 2L TB broth. The cells were grown at 37 °C to OD600 ~ 0.6, cooled to 18 °C, and then 1 mM IPTG was added to induce expression for 20 hours at 18 °C. Cells were harvested by low speed centrifugation at 8000g for 15 minutes with a Sorvall RC-5B centrifuge at 4 °C, washed with cold buffer A (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 1 mM mercaptoethanol) and resuspended in 40 ml buffer A. DNase (Roche Applied Science, #10104159001) was added to a final concentration of 10 μg/ml and phenylmethylsulfonyl fluoride (PMSF) to 1 mM before the cells were lysed with the French press at 16000 psi. Unbroken cells and debris were removed by low speed centrifugation, and supernatant was further clarified by ultra-centrifugation (100,000g for 1 hour with a Beckman Ti50.2 rotor). The supernatant was filtered through a 0.45 μm filter and imidazole was added to a final concentration of 10 mM to reduce non-specific binding to the Ni-NTA column. The sample was loaded into a HisTrap nickel-chelating column (Amersham, #17-5247-01) that had been equilibrated with buffer A. The column was then washed with buffer A until the OD280 returned to zero, then washed with 50 and 100 mM imidazole in buffer A. IniA protein was eluted with 200 mM imidazole in buffer A and dialyzed against 50 mM Tris-HCl, pH 8.0, 100 mM Li_2SO_4, 1 mM DTT. A 100 kDa MWCO 20 ml spin-concentrator was used to concentrate IniA by spinning at 3000g at 4 °C. Glycerol was added to 40% to keep IniA in soluble form during concentration and storage.

SDS-PAGE

Equal volumes of reducing sample buffer (3% SDS, 125 mM Tris-HCl, pH 6.8, 15% glycerol, 0.01% bromophenol blue, 0.7 M mercaptoethanol) were mixed with the protein solution, boiled for 1 minute at 100 °C or incubated for 30 minutes at 50 °C. Samples were spun at max speed for 3 minutes in a desktop centrifuge to remove the undissolved pellet. The samples were loaded onto a 4-24% gradient polyacrylamide gel (BioRad Criterion) and stained with Coomassie blue R250 dye for 2 hours at room temperature. Destaining was performed using a solution of 40% ethanol, 10% glacial acetic acid. Gels were then scanned for record.

Preparation of carbon-coated grids

Pure carbon-coated grids were prepared with the following protocol (Figure 2-3). (1) 400 mesh copper grids were cleaned by sonication for 10 seconds in acetone followed by 10 seconds of sonication in ethyl alcohol. After removal from the solvent, the grids were dried on a filter paper at 60 °C. (2) A thin layer of carbon (~ 10 nm) was deposited on a freshly cleaved mica sheet. (3) A container with an adjustable outlet at the bottom was filled with distilled water, and a glass beaker was put upside-down into the container. A stainless steel wire and a filter paper were placed on top of the beaker below the water surface. (4) The cleaned copper grids were placed on the filter paper. (5) The carbon-coated mica was submerges in the water to allow the carbon film to flow on the water surface. (6) The outlet of the container was opened, and the water level was lowered slowly while making sure the carbon film was kept on top of the copper grids.

After the water level was lowered below the steel wire and the carbon film was deposited onto the grids, the "steel wire-filter paper-copper grid-carbon film" was removed together to a covered petri dish and dried.

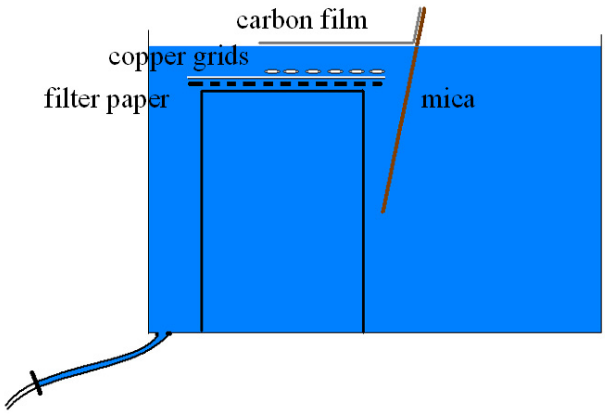

Figure 2-3. Diagram depicting the batch preparation of carbon-coated grids. A container with a bottom outlet was filled with water, and a beaker was placed upside-down inside the container. A stainless steel mesh and filter paper were placed on top of the beaker, below the water surface. Cleaned copper grids were placed on top of the filter paper. Under an angle, a carbon-coated mica sheet was inserted into the water in order to almost completely float the carbon film off. The remaining connection between the carbon and the mica enabled a controlled positioning of the carbon film over the copper grids. Subsequent lowering of the water level resulted in a deposition of the carbon film onto the copper grids.

Negative staining

A 2.5 µl solution of the purified protein complex at 0.1 mg/ml in TBS buffer (20 mM Tris-HCl, pH 6.9, 150 mM NaCl) was applied directly to a freshly glow-discharged carbon-coated 400-mesh grid, absorbed for 1 minute and then washed with 20 µl distilled water on a piece of parafilm. The grid was blotted briefly with filter paper to remove excess water and stained with 3 µl of an aqueous solution of uranyl acetate (1%, w/v, pH 4.25) for 1 minute, blotted dry and stored in a dust-free environment.

Alternatively, negative staining was performed with the following protocol that did not need pre-made carbon-grids (Valentine et al., 1968). Briefly, a ~ 4×4 mm piece of carbon-coated mica was inserted into 50 µl protein solution for 1 minute. The carbon film with the mica was transferred to 100 µl distilled water, and then floated off on 10 µl 1% uranyl acetate for 1 minute. The carbon film was transferred to a clean copper grid and blotted dry with filter paper.

Transmission electron microscopy and image digitizing

Specimens were observed in a JEOL 1200EX transmission electron microscope (TEM) operated at an acceleration voltage of 100 kV and images were recorded at calibrated magnifications of 39,800× on Kodak 4489 electron image sheet film.

TEM films were scanned using a Leafscan-45 microdensitometer (Mitsuoka et al., 1997) with 1270 DPI step size, corresponding to 5.14 Å/pixel at the specimen level. After the initial preview, the area to be scanned was chosen and the contrast was adjusted. The scanned images were saved as uncompressed TIF files for image analysis.

Single particle 3D reconstruction with EMAN

Most of the single particle image analysis and 3D reconstructions were performed using the EMAN software package (Ludtke et al., 1999). Single particle images of protein molecules were selected from electron micrographs with the 'boxer' utility. The selected particles were filtered to remove high and low frequency noise (only

frequencies between 15 – 200 Å were kept), and then centered by aligning with a reference that was produced by the rotational average of all the particles. The centering process was repeated several times. The particles that could not be centered were considered 'bad' particles and were removed from further processing. The 'good' particles were clipped to a box size about 20% larger than the dimension of the molecule and saved in a file named as 'start.hed/img', which was used as the default input file by the 'refine' program.

For particles of C3 or higher symmetry, 'startcsym' was used to build a first 3D model, which only used the best 5% top view and side view particles (ring-shaped particles and particles exhibiting a mirror or *pseudo*-mirror symmetry).

The 3D model was then projected into different directions to obtain a defined number of 2D projections (stored in 'proj.hed/img' file). These 2D projections were used as references to classify the raw particles in 'start.hed/img'. Each of the raw particles was compared to each of the 2D projections, and the best match was put into the group (class) that belongs to that reference. All the particles in each group were aligned to the reference to produce an averaged map. Depending on the purity of the protein and noise level of the image, a portion of the particles that did not match the averaged image were discarded and a new average was made with only the remaining particles. Then the class average was used as a new reference to align the raw particles in that class and produced another class average. This process was repeated several times to avoid the risk of the initial model bias.

The class averages and their corresponding Euler angles (the same angles as the references) were used to build a new 3D model. This new 3D model was then used as a starting point for the next round of refinement. The refinement process was repeated many times until no improvement or change in the 3D model was observed, as judged by the convergence of the Fourier Shell Correlation (FSC). FSC was produced by cross-correlation of two 3D models in Fourier space. Each point of a 3D model in Fourier space was compared to the corresponding point of another 3D model in Fourier space, and an averaged value was calculated for each shell of frequency in Fourier space. A curve was plotted with the cross-correlation values from each frequency shell as the Y-axis and the frequency as the X-axis.

For asymmetric molecules, 'startnrclasses' was used to classify the raw particles into ~ 50 classes. About 8 different classes were selected and saved in 'good.hed/img' file. The 'startany' program was used to produce an initial 3D model with the common line principle similar to the IMAGIC software. Then the models were refined as described above but instead with C1 symmetry.

Resolution was estimated by EOTEST. First the data set was split into an even half and an odd half. 3D reconstruction was performed separately, and then the similarity of the halves in Fourier space was compared. The frequency value that corresponding to cross-correlation value of 0.5 in the FSC curve was used as resolution in this study.

Visualizing 3D volumes

Chimera software was used for the visualization of 3D volumes in the '.mrc' file format (Pettersen et al., 2004). If the 3D model was built with IMAGIC software, the '.hed/img' format was changed to '.mrc' format using the 'em2em' program in IMAGIC. A molecular mass was assigned to the 3D model with the 'volume' program of the EMAN package. Alternatively, the 'vis5d' program was used for display the 3D volume (http://www.ssec.wisc.edu/~billh/vis5d.html, by Bill Hibbard at University of Wisconsin). The 'vis5d' format, '.v5d', was transformed from the '.mrc' file by the 'mrc2v5d' program in EMAN.

CHAPTER III

CHARACTERIZATION OF NSP15 FROM SARS VIRUS*

Introduction

The Severe Acute Respiratory Syndrome (SARS) epidemic emerged from south China in 2003 and caught the world with a surprise. It spread throughout the world in a very short time. More than 8000 people were affected with mortality rates reaching over 40% in certain populations (Stark and Atreya, 2005). This epidemic reflects how poorly our understanding to microbes and how un-prepared our human being to pathogens. There are no special drugs to combat the SARS virus. The existing treatment is not efficient at deterring the disease and has severe side effects (Tsang and Seto, 2004). Although the spread of the disease was controlled by strict quarantine and patient isolation, it may resurface at anytime. If it reappears, there is still no effective treatment to prevent another SARS pandemic. Better understanding of the infection and replication cycle of the SARS virus can reveal potential drug targets.

Nsp15 is one potential drug target that worth more investigation (Bhardwaj et al., 2004). It is a novel endoribonuclease that is unique to this family of viruses and is essential for the virus replication (Ivanov et al., 2004). Determining the structure of Nsp15 will help us to understand its mechanism and design drugs to inhibit its activity. In collaboration with Drs. Cheng Kao, Linda Guarino and James Sacchettini, the

---

* Part of the data reported in this chapters is reprinted with permission from "Mutational analysis of the SARS virus Nsp15 endoribonuclease: identification of residues affecting hexamer formation." by Guarino, L.A., Bhardwaj, K., Dong, W., Sun, J., Holzenburg, A., and Kao, C. (2005). J. Mol. Biol. *353*, 1106-1117. Copyright (2005) by Elsevier.

biochemical and biophysical characters of Nsp15 were studied. Lipid monolayer technique was used to crystallize Nsp15. Single particle analysis was used to obtain the first 3D structure of Nsp15.

Results

2D crystallization of Nsp15 on a lipid monolayer

Determining the oligomerization state of Nsp15 is the initial goal of this study. Most endoribonucleases function as dimers (Chen et al., 2005), but Nsp15 can exist as either a monomer or hexamer based on analytical ultra-centrifugation, gel filtration and UV cross-linking/SDS-PAGE studies (Guarino et al., 2005). One mutant Nsp15 with a single amino acid substitution E3A is predominately monomeric. Another mutant protein K289A is predominately hexameric. The H234A mutation is similar to wild type with the same ratio of monomeric and hexameric forms. The molecular mass of an Nsp15 monomer is 40 kDa, and a hexamer is 240 kDa. A hexamer of this size should be visible directly in TEM, but several attempts by different researchers had failed to see any hexameric particles in TEM.

My initial attempt of negatively stained Nsp15 on glow discharged carbon films also failed to detect any ring-like particles. As the other experimental data reliably indicated that Nsp15 could exist as a hexamer, it is likely that the protein was not stable, or did not stick to the carbon film, or any other reasons. Changing sample preparation methods may help. Previous experience showed that some proteins are better preserved on a lipid monolayer. Another benefit of using the lipid monolayer method is that protein

molecules can be concentrated on the lipid surface from a very dilute solution and can eventually form 2D crystals. If crystals do form, even not perfect, we should be able to analyze them by electron crystallography.

Nsp15 was expressed with a 6His-tag and purified with a Ni-NTA affinity column; thus, it should also be able to efficiently bind to the Ni-NTA lipid. At certain conditions, many particles of long, square shape were observed in TEM, but hexameric rings could not be detected. It is unlikely that those particles were contaminates, as the sample was very pure. After trying many conditions, 2D crystal was finally obtained. Averaged projection shows that each asymmetric unit is composed of one 'X-shaped' particle (Figure 3-1). The size of the particle (94 × 81 Å) can accommodate a protein of 240 kDa if the third dimension is of similar size. It is not clear from this crystal how the six subunits are arranged in the X-shaped particle.

2D crystallization of Nsp15 on carbon film

Crystals of the ring-shaped particle was obtained on carbon film directly in an effort to prepare sample for single particle imaging (Figure 3-2). Image analysis of the crystals show that the phase residue with 3-fold symmetry (p3) is smaller than with 6-fold symmetry (p6), and the projection map with p6 symmetry smears out the details that could be seen with p3 symmetry. If Nsp15 is a hexamer but has 3-fold symmetry, the most possible organization is a dimer of trimers. The 'X-shaped' projection of the former crystal can be thought as the side view of the particles with 2-fold symmetry. From the two crystal forms, we can conclude that the symmetry of Nsp15 molecule is

D3: one 3-fold axis and one 2-fold axis, and the two axes are perpendicular to each other. Larger and better crystals are needed to determine the 3D structure of Nsp15. Single particle method can also determine the 3D structure, although at relatively lower resolution.

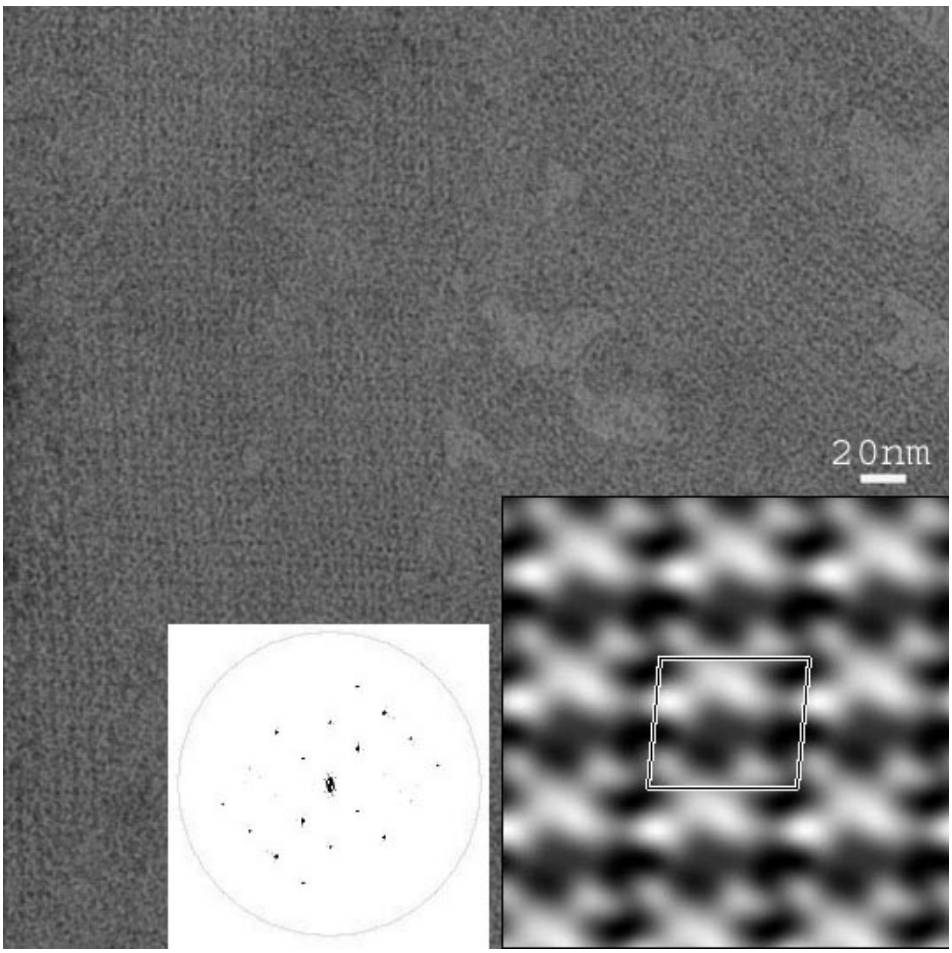

Figure 3-1. 2D crystal of NSP15 K289A grown on a lipid monolayer. Inserts show a representative FFT of the crystal (left) and the corresponding Fourier projection map (right) calculated in p1 with a = 93.5 Å, b = 80.5 Å, and γ = 95.7º. One unit cell is highlighted. Protein density appears white against a darker background.

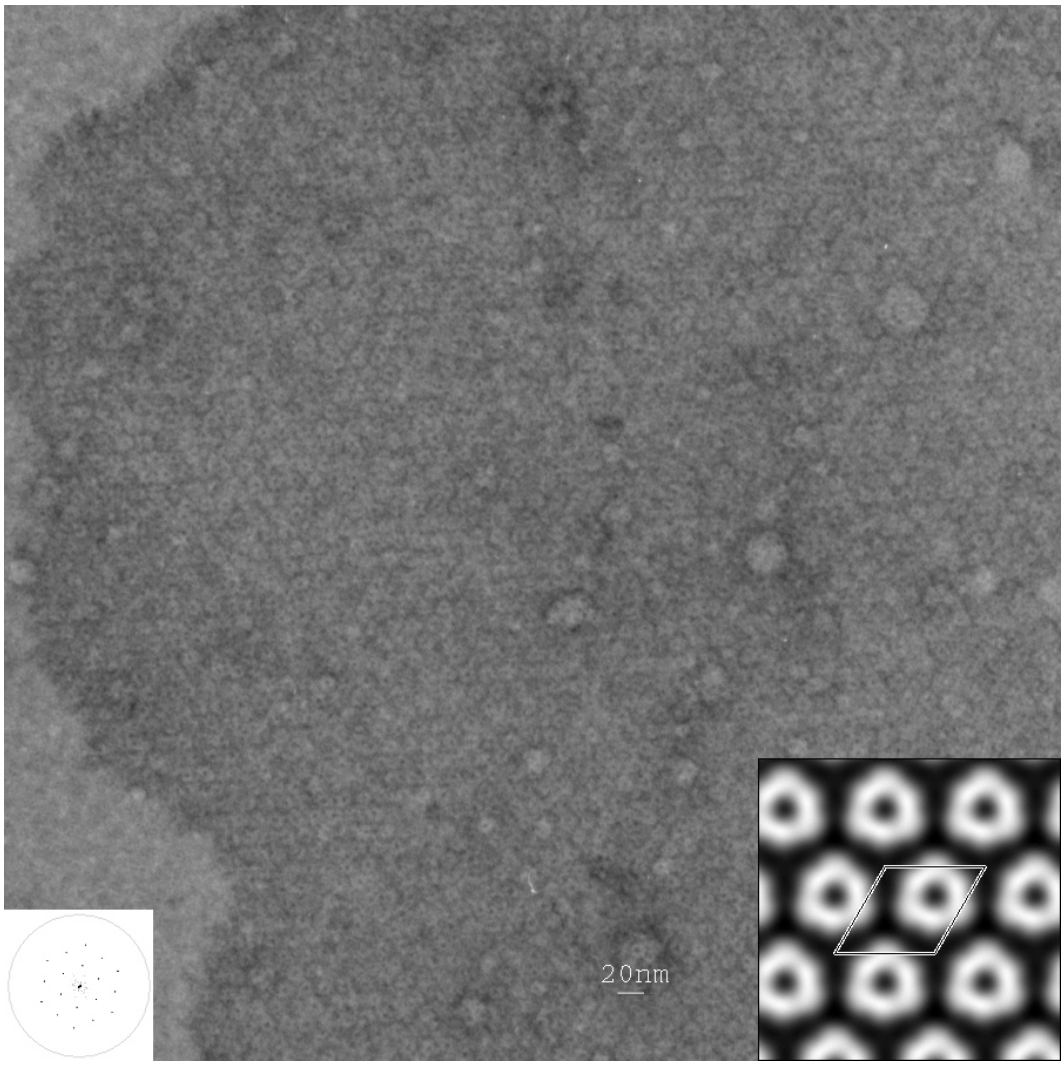

Figure 3-2. 2D crystal of NSP15 K289A grown epitaxially on carbon film. Inserts show a representative FFT of the crystal (left) and the corresponding Fourier projection map (right) calculated in p3 with a = b = 90.6 Å, and γ = 120º. One unit cell is highlighted. Protein density appears white against a darker background.

Nsp15 forms a dimer of trimers

If the triangular-shaped crystal is the top view and the long, square-shaped crystal is the side view, we can predict the arrangement of the six subunits of Nsp15 hexamer. The side view has two-fold symmetry, indicating the top and the bottom of the particle are identical. The most probable arrangement is that three monomers form a trimer, and two trimers form a hexamer. The 3-fold axis is likely to form along the long axis of the square-shaped side view, so that the surface area of the top view is smaller than the side view. This difference in size may be the reason that the ring-shaped top view is rare in single particle images.

With this knowledge in mind, we started to build a 3D model by single particle method. Images of monodispersed particles were prepared on carbon film (Figure 3-3). Class averages of selected particles show several different long, square-shaped particles. Some of the side views resemble the "X-shaped" projection of the 2D crystal on lipid monolayer; the others may be side views on other orientations. Ring-shaped particles are very rare, less than 0.1%, and no features could be seen from the class average.

An initial 3D model was produced from the top view and side view with C3 symmetry. Although this model is crude and very noise, I continued with the refinement process. After several cycles of refinement with D3 symmetry, the 3D models were improved significantly, and some interesting features began to appear. The structures converged after eight cycles of refinement (Figure 3-3B). The resolution of the final 3D reconstruction is 18 Å, as judged by EO-test, which has reached the limitation of negative stain.

One indication of the reliability of the structure is the perfect match of the class averages and the reprojections from the 3D model (Figure 3-3C). Surface displays of the final 3D model clearly show the dimer of trimers architect, with many holes or channels within and between the trimers (Figure 3-3E). These channels may be the sites for RNA substrate binding. Each monomer can be separated into three domains, with the inner domain of all six subunits fused together to form a large central mass.

From the asymmetric triangle (Figure 3-3D), it could be seen that the orientations of the particles are heavily biased to the side view, which was already obvious in the raw images and class averages. It becomes clear now that the width of the particle is smaller than the height. When the particle rests on the carbon film in the side view, it has more contact surface and is more stable than when it sets in the top view.

The cylinder shape of the particles favors them to 'roll' on the carbon film, which can give many different side views. Although top views are rare, this obstacle does not preclude a successful 3D reconstruction. Many different side views provide sufficient coverage of the Fourier space, which is equivalent to single axis tilting from -90° to +90°, so no missing cone problems existed.

Although the size of Nsp15 is small, close to the suggested limitation of single particle analysis method (200 kDa), the final 3D reconstruction is nearly perfect. The characteristic shape of Nsp15 hexamer with so many channels within and between the two trimers reveals a unique structure of a new type of endoribonucleases that has not been seen before.

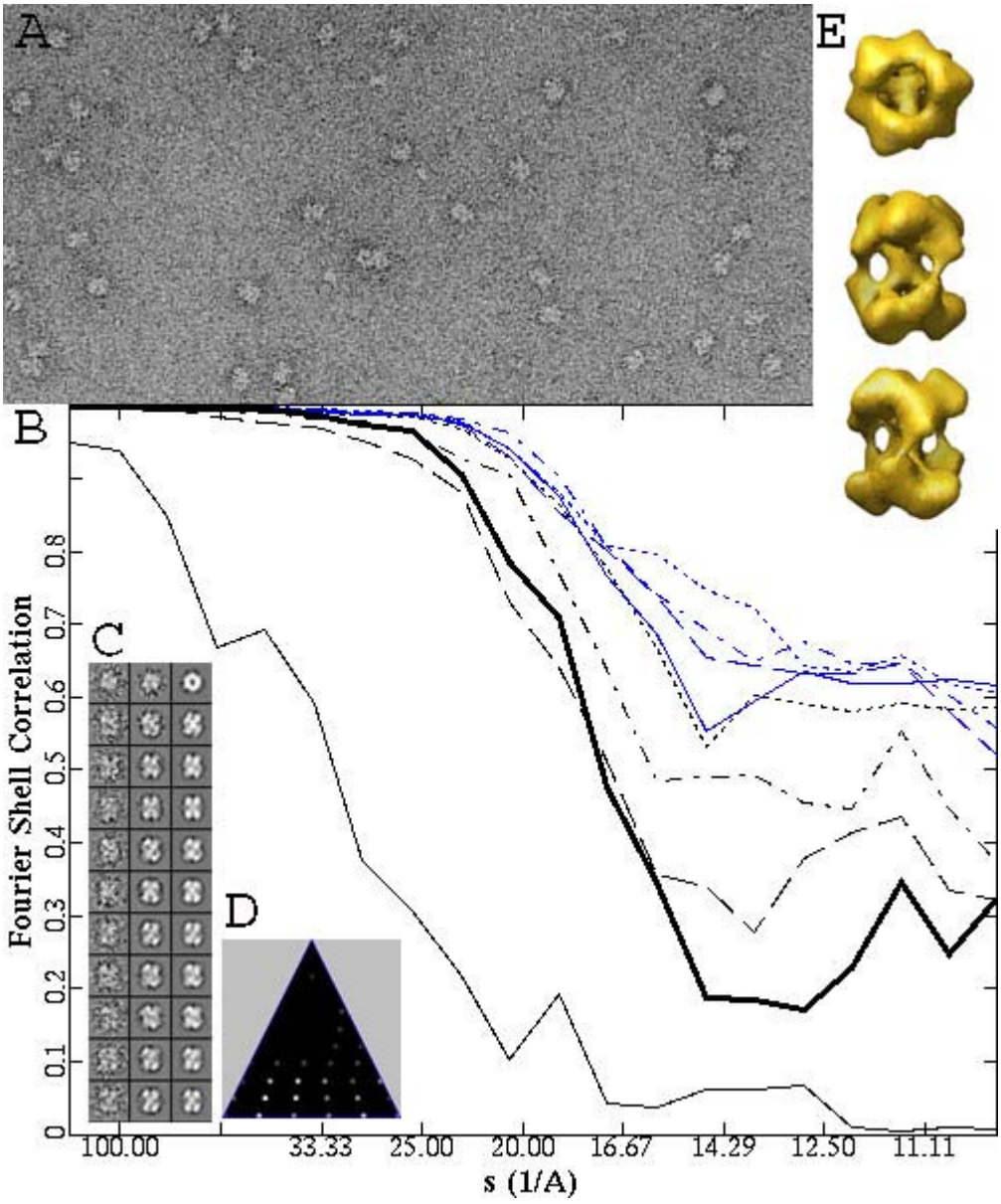

Figure 3-3. 3D reconstruction of NSP15 K289A. (A) Electron micrograph of negative stained Nsp15 K289A. (B) FSC showing the progress in the refinement process. The EO-test indicates a resolution of approximately 2 nm. (C) Left column: raw particles; middle column: class averages; right column: re-projections. (D) Asymmetric triangle highlighting the fact that the data set is dominated by side-on views. (E) Surface representation of the 3D structure with the molecular mass set to 240 kDa.

2D crystallization of Nsp15 with RNA substrate

RNA substrate binding was captured in a crystal formed by the 'X-shaped' side views on a lipid monolayer (Figure 3-4). An extra density is visible at an empty space of the "X-shaped" side view, which correspond to a gap between the two trimers in the 3D model. Only one extra density is observed in each particle, but this does not exclude the possibility that each hexamer could potentially bind more than one RNA substrate. The size of the particle with RNA binding is increased by 10%, which may be caused by a rearrangement of subunits of the dimer of trimers.

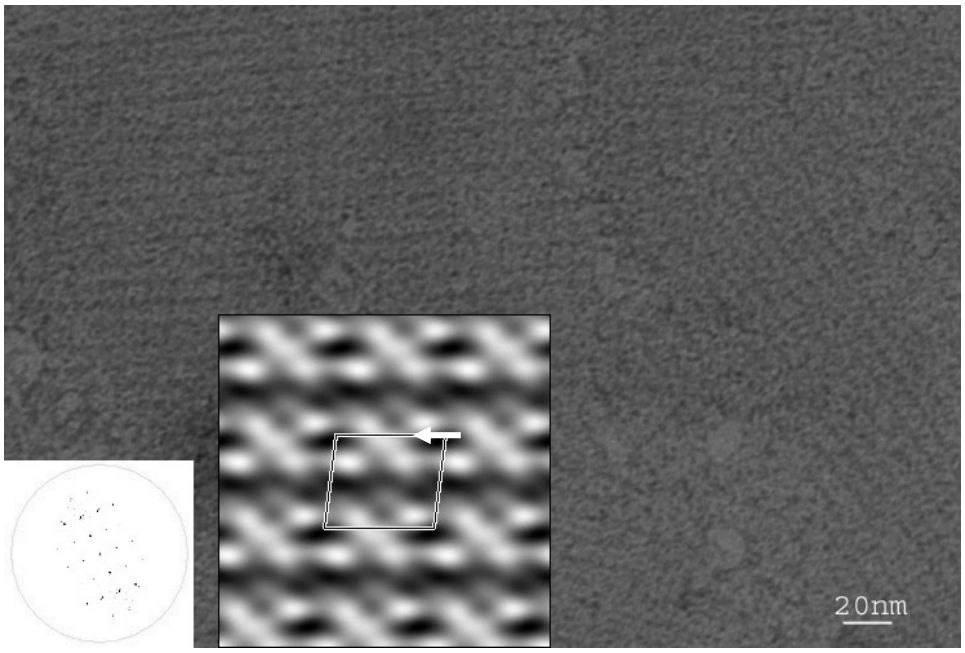

Figure 3-4. 2D crystal of NSP15 K289A complexed with RNA substrate. Inserts show a representative FFT (left) and the corresponding Fourier projection map (right) calculated in p1. One unit cell is highlighted with a = 104.7 Å, b = 88.8 Å, and γ = 96.6º. Protein density appears white against a darker background. Compared to Figure 3-1, an extra density can be identified (arrow).

Cryo-EM of Nsp15

Fast freezing a very thin layer (~20 nm) of aqueous protein solution in holey carbon film can vitrify the sample without the formation of ice crystals (Frank, 2002). The protein molecules can then be imaged in cryo-EM and potentially, high-resolution information can be recorded and dynamic conformational changes of the protein molecules can be captured.

The first attempt of imaging Nsp15 K289A molecules in cryo-EM was done in the same buffer used for crystallization. No particles could be detected in either close-to-focus or far-defocus images. As particles of similar size had been imaged before, we tried different buffers and protocols to prepare the samples. Figure 3-5 shows a focus pair images of Nsp15 K289A recorded with a 4K×4K CCD camera. Although the particles are barely visible in the close-to-focus image, they are far easier to identify in the far-defocus image. Some cryo-EM images of Nsp15 K289A have structural information that extends to 10 Å (Figure 3-6B). The structural factor calculated from four particle sets is not perfect (Figure 3-6A), as can be seen from the CTF fitting curve (Figure 3-6B). An initial 3D model was constructed by C3 symmetry with 50 best top and side views selected from 8000 particles. Although the first 3D model is featureless, after eight cycles of refinement with D3 symmetry, the final 3D model resembles the negatively stained reconstruction (Figure 3-6I). The resolution is ~ 25 Å from EO-test with 0.5 cut off. Although some top views and intermediate views are present in the data sets, side views are the predominate orientation (Figure 3-6E and F).

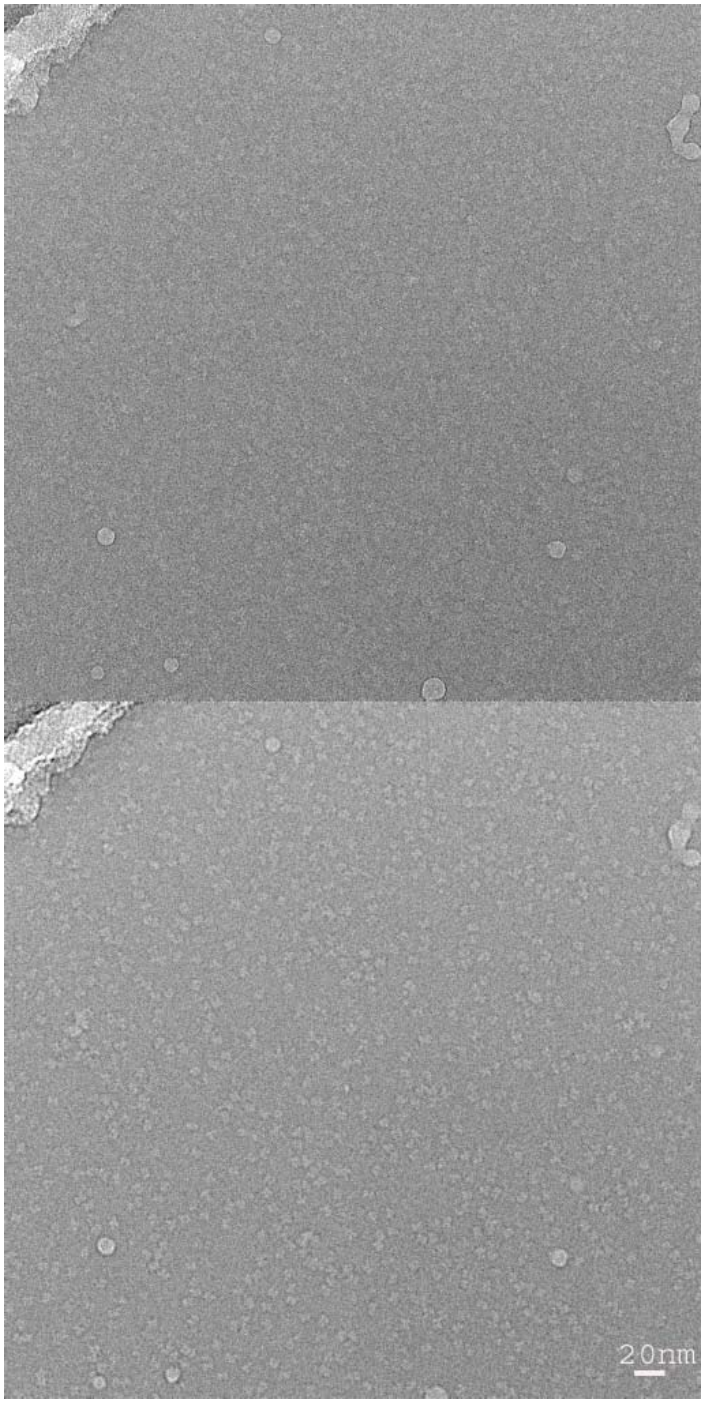

Figure 3-5. Cryo-electron micrographs of Nsp15. The same area is imaged twice at two different defocus settings, i.e.1 μm defocus (top) and 3 μm defocus (bottom). The contrast has been reversed so that protein density appears white against a darker background.

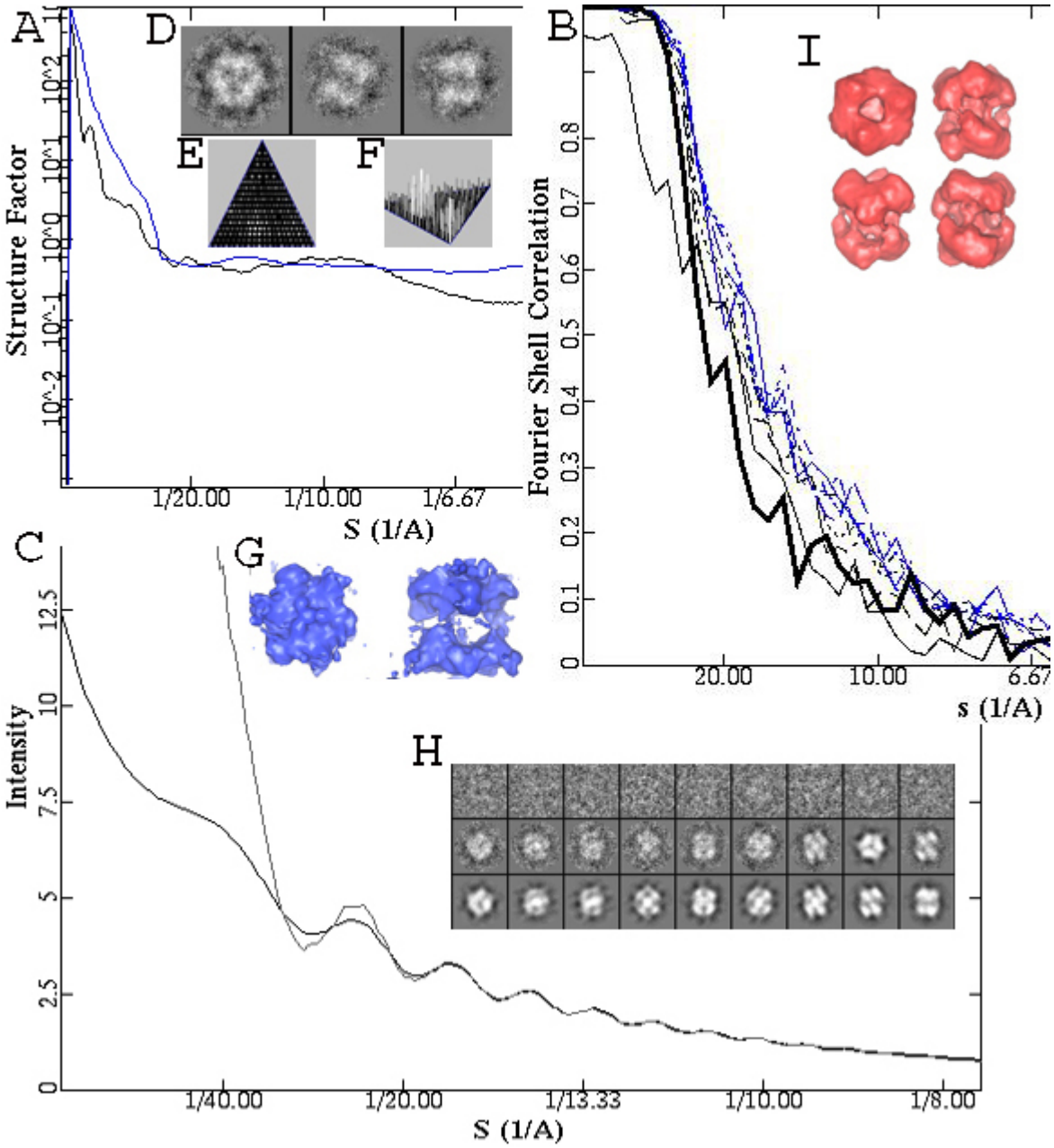

Figure 3-6. Cryo-EM 3D reconstruction of Nsp15. (A) Structure factor intensity plotted vs. spatial frequency. The black line was measured for GroEL and the blue line was calculated for Nsp15. (B) Power spectrum (light line) and fitted curve (dark line) (3) FSC monitoring the refinement process and EO-test (bold line). (D and G) Class averages and surface representation of the initial 3D reconstruction. (E and F) Asymmetric triangle shows the distribution of projections. (H) Raw particles, class averages, and re-projections (left to right). (I) Surface representation of the final 3D reconstruction at different orientations.

Discussion

Many genetic, biochemical, and biophysical methods have been applied to characterize the Nsp15 endoribonuclease from SARS virus (Bhardwaj et al., 2004; Guarino et al., 2005), but its 3D structure is the most necessary information. It provides the basis for more in-depth experimental design and is essential for structure-based drug discovery. X-ray crystallography was the first choice for trying to determine the structure of Nsp15. 3D crystals of Nsp15 have been obtained, but the diffraction was not good enough to solve the phase problem and no structural information could be extracted from the X-ray diffraction patterns. In contrast, 2D crystal analysis provided useful information despite the smaller crystal size as compare to the 3D crystals. From the projection of the side view crystals, Nsp15 has 2-fold symmetry. From the top view crystals, the Nsp15 has 3-fold symmetry. The information from 2D crystal analysis is important for the choice of D3 symmetry in single particle 3D reconstruction. It is possible to determine symmetry from single particle analysis *ab initio*, but the missing top view in electron micrograph renders the symmetry determination difficult.

2D crystallization experiments by lipid monolayer method include changes in many parameters, including protein purity and concentration, pH, buffer constitution, temperature, the type of dilution lipid, and the ratio of dilution lipid to Ni-NTA lipid. 2D crystals of Nsp15 were detected using a buffer that contains 10 mM Tris-HCl, pH 6.5, 150 mM NaCL, 5 mM $MgCl_2$, 40 mM Na_2SO_4, 50 mM $(NH_4)_2SO_4$ and 2 % (w/w) PEG3350 with 0.1 mg/ml Ni-NTA-DOPS and 0.4 mg/ml egg PC. Some divalent-ions and salts that promote 3D crystal formation also increase the chance of 2D crystal

formation, but at relatively lower concentration. There are other methods to grow 2D crystals. In some cases, 2D crystals formed just by chance, like the p3 crystal of Nsp15 K289A.

It is possible to determine the 3D structure from the 2D crystal images collected at different orientations. This is a time consume process, and the result will not be satisfied as the 2D crystals are not good enough. Single particle 3D reconstruction is more powerful in elucidation of the subunit arrangement and the architecture of Nsp15 with less effect. Although top view is missing, the different side views are enough to cover the whole Fourier space to perform 3D reconstruction successfully.

Even at this low resolution, the information obtained from the single particle 3D reconstruction has begun to reveal some interesting characters of Nsp15. The dimer of trimers organization of Nsp15 subunits is very unique, and the multiple channels within and between the trimers are especially rare. This resembles the structure of GroEL to some degree, although Nsp15 is only one-third the mass of GroEL. GroEL is functioning as a molecular chaperonin and need co-operation of the intra- and inter- subunit rings in the process of refolding protein (Ranson et al., 2001). The symmetric arrangement of Nsp15 dimer of trimers indicates that the subunits may also co-ordinate for its function. It is possible that endoribonuclease is not the only activity of Nsp15. Other RNA processing enzyme, including Nsp9, Nsp14 and Nsp16, may interact with Nsp15. They may even form a large protein complex to perform the RNA replication and processing co-ordinately.

Cryo-EM holds the promises to reach higher resolution (Chiu et al., 2005). Our initial attempt with 8000 particles of Nsp15 cryo-EM images produced a 3D model similar to negatively stained reconstruction. While this conformed our early result, it is also disappointing that we did not get structure at higher resolution. There are many possible reasons why we did not get better structure. One unexpected thing is the uneven distribution of orientations of Nsp15 molecules in cryo-EM. The reason for the uneven distribution is not known, as the vitrified sample should have the same orientation as in the solution, which should randomly oriented. It is possible that the orientations of Nsp15 molecules were not evenly distributed in the thin layer solution within the holey carbon before frozen. Uneven distribution of surface charged groups might cause the biased orientation at the surface in the thin layer solution.

As the cryo-EM images have very low contrast and are very noisy, much more particles are needed to improve the signal-to-noise ratio and reach higher resolution. We have collected 30,000 particles and the intermediate result of refinement is promising. Even with the fastest computer in our lab, one circle of refinement takes one week. We need many circle of refinement, which will take months to get a converged structure. We also need to use different initial models and parameters for refinement to verify that the reconstruction process does not create artifacts.

Nsp15 may have multiple conformations in solution. If that is the case, using multiple initial models for ‘multirefine’ at the same time may solve this problem. The bright side is that we may get different conformations at the same time. However, this procedure requires much more particles, and also much long time to process. Performing

an X-ray solution scattering experiment to acquire a 1D structural factor curve for CTF correction may also help, as the structural factor produced from cryo-EM data sets does not match well when used to fit the power spectrum curves of the particles.

One final note is that other researchers have met similar problem that some organic compounds, like glycerol, PEG, and detergent, which are good to stabilize the protein sample, are not suitable for cryo-EM imaging. This may reflect the low contrast between protein molecules and the background in solutions with organic compounds, which have similar density with biological macromolecules.

Materials and Methods

2D crystallization on a lipid monolayer

15 μl protein solution at 0.1 mg/ml in 20 mM Hepes, pH6.9, 100 mM NaCl, 1 mM NaN_3 was added to a custom-designed Teflon well measuring 4 mm wide and 1 mm high. 0.3 μl functional lipid (Ni-NTA-DOPS) diluted in egg PC lipid (0.1 and 0.4 mg/ml, respectively) in a chloroform/hexane (1/1, v/v) solution was spread over the 6His-tagged protein solution and incubated overnight at 4 °C in a humid environment. A 400-mesh grid with a hydrophobic carbon film (more than two weeks old) was placed on top of the solution for several minutes, washed with distilled water and negatively stained with 1% uranyl acetate for 1 minute. This was often performed by transferring the grids to drops of water (50 μl) and stain solution (5 μl) on parafilm. Excess water was blotted briefly after water wash, but did not let the grid completely dry before staining.

Crystallographic image processing

The CRISP software package (Hovmoller, 1992) was used for the analysis of 2D crystals. The crystal image was imported from a digitized TIF file; an area of the crystal was selected and used to calculate Fourier transform (FFT). The cut-off in the FFT was chosen so that it corresponded to the first zero in the contrast transfer function (CTF) of the image. The resulting patterns were indexed manually by selecting two reflections and the corresponding (h, k) values were assigned. These two reflections were not within the same line from the origin. The crystallographic symmetry was selected according to the unit cell parameters, systematic absences and the phase residuals of symmetry-related reflections. A projection map with the allowed symmetry was then produced and evaluated.

Cryo-EM

The cryo-EM data were collected at the National Center for Macromolecular Imaging at Houston (Booth et al., 2004). Briefly, 2.5 μl purified protein (0.2 mg/ml) solution was applied to a glow-discharged holey carbon grid (Quantiforil R2-1), blotted with filter paper to remove extra buffer, and flash frozen by plunging into liquid ethane cooled by liquid nitrogen. The grid was transferred into a grid box under liquid nitrogen and kept in liquid nitrogen in all the later steps to avoid ice crystal formation. A Gatan cryostage was used to keep the specimen below –160 °C for data collection in a JEOL 2010F cryo-EM. After the cryostage stabilized (~ 1 hour), the microscope was aligned, the sample was focused and the control software was set up. JEOL automated

microscopy expert system (JAMES), a specially designed software package (Booth et al., 2004), was used for semi-automatic data collection. First, the grid was examined at low magnification to select holes with the proper ice thickness. Then low-dose procedure was using to record the image (12 $e^-/Å^2$) on a 4K × 4K CCD camera (Gatan US4000) at 60,000× magnification. Each area was imaged twice. A close-to-focus image was recorded first (1 to 2.5 μm); after a 1-minute interval, a far-defocus image (3 to 4 μm) was recorded.

The close-to-focus image had high-resolution information, but very low contrast and the particles were difficult to recognize. The far-defocus image was used to select particles semi-automatically with the 'boxer' program of the EMAN software package. The coordinates of the particles from the far-defocus image were then used to locate particles from the close-to-focus image. As there was movement and rotation between the two images, the far-defocus image was first aligned to the near-focus image with 'alignhuge' program. The other steps of 3D reconstruction process were similar to the EMAN instruction described before but contrast transfer function (CTF) correction was applied to each data set.

(1) 1D structural factor used for CTF correction was derived from the cryo-EM images. First several particle sets from images of different defocus were processed with 'ctfit' program in 'Change background mode' of 'Advanced' menu'. Set 'Amp' to 0, adjusted 'N/A' to a continuous line, and then adjusted other parameter to fit the power spectrum curve of each data set. The 'Struct Fac' button was selected to display the calculated structural factor for all the data sets, readjusted the CTF parameter to allow

the low-resolution range of calculated structural factors match each other, while keeping the curve in the main window also fit well. On the ‘Advanced’ menu, selected ‘Save 1 Column’ with option ‘11’. ‘sfmerge.py’ program was then used to combine the calculated structural factor at low-frequency range with experimental determined structure factor (groel.sm) at high frequency range.

(2) Fit the power spectrum curve of each particle sets by ‘ctfit’ program. First input the 1D structural factor created above, set ‘Amp’ to 0, adjusted noise parameters to fit the baseline of the power spectrum curve. Then adjusted the defocus and other parameters to get the best fit. The “B-factor” was 100 to 200 for JEOL2010F cryo-EM, but differed for each micrograph. After flipped the phase of the particle set, ‘.fix.’ was inserted into the filename in the phase flipped particle set. The CTF parameters were also stored in the header file of the particle set.

Alternatively, ‘fitctf’ program was used to calculate the CTF parameters automatically, but the program did not work well at some cases, so the results were checked with ‘ctfit’.

(3) Symmetry of the particles was determined first by ‘startnrclasses’. An initial 3D model was produced with ‘startcsym’ program for particles with higher than C3 symmetry, or with ‘startAny’ program for particles without symmetry.

(5) In ‘refine’ options, ‘ctfcw=’ was used for full CTF correction with the calculated structural factor. Alternatively, ‘ctfc=’ option was used with a threshold of resolution, but no correction for amplitude was applied.

CHAPTER IV

STRUCTURAL AND FUNCTIONAL ANALYSIS OF THE HUMAN TLR3 LIGAND BINDING DOMAIN

Introduction

Microbes are everywhere in the environment and many of them are pathogenic to human beings. In order to survive in the sea of microbes, a sound immune system is needed. The mammalian immune system can be divided into two parts: innate and adaptive (Pasare and Medzhitov, 2004). Innate immunity, which also exists in other more primitive organisms, is the first line of defense. If the innate immune system is not sufficient to eradicate the invaders, the adaptive immune system is called to enhance the battle with the pathogens. T- and B-lymphocyte receptors of the adaptive immune system can recognize large amounts of different antigens with high specificity/affinity (Tonegawa, 1983), but recognizing antigen is not sufficient to activate T- or B-cells. Activation of T- and B-lymphocytes requires two signals. Antigen recognition provides the first signal. If the lymphocytes only receive the first signal without proper second signal, they do not respond to the antigen. The second signal is produced by the innate immune response to the pathogens. This signal is initiated by germline-encoded receptors that recognize pathogen associated molecular patterns (PAMPs). PAMPs are common structural components of microbes, but are not produced by mammalian cells. The most important PAMP receptors are Toll-like receptors (TLRs), which were first demonstrated in *Drosophila* to be essential for sensing and defending against fungi

(Lemaitre et al., 1996). More than ten TLRs that recognize different PAMPs have been identified in the human genome (Akira et al., 2003).

All TLRs have a single transmembrane α-helix that connects the extra- and intra-cellular domains. The intracellular domain, called TIR for Toll/Interleukin-1 receptor, is conserved among all TLRs, as well as some interleukin receptors and adaptor proteins (Yammamoto et al., 2004). The extracellular domain (ECD) has tandem leucine-rich repeats (LRRs) with N- and C- terminal caps (Takeda and Akira, 2005). The proposed function of LRRs is to provide a frame for substrate binding and protein-protein interactions (Bell et al., 2003). It is proposed that PAMPs either bind directly to the LRRs or bind indirectly through adaptor proteins, which cause a conformational change in TLR that triggers signal transduction inside the cell and initiates immune responses.

Full length TLR was difficult to express and purify, so we worked with the ECD to study its structure and function. TLR3 ECD (28-702 aa) has 23 LRRs with a consensus motif of $xL^{2}xxL^{5}xL^{7}xxN^{10}xL^{12}xxL^{15}xxxxF^{20}xxL^{23}x$ (Choe et al., 2005). The N-terminal 27 amino acids encode a signal peptide that is cleaved and is not present in the mature protein (Bell et al., 2005).

TLR3 ECD is highly glycosylated, about 35% of its mass is sugar (76/115 kDa) when expressed in human cells. 2D crystallization of TLR3 ECD by the lipid monolayer technique demonstrates that this method can be applied to highly glycosylated proteins. Single particle 3D reconstruction was also applied to TLR3 ECD. Both methods reveal a large horseshoe shape of TLR3 ECD.

Polymorphisms of TLR genes have been shown to correlate to many human diseases, including sepsis, immunodeficiencies, atherosclerosis and asthma (Cook et al., 2004; Rifkin et al., 2005). Some single nucleotide polymorphisms (SNPs) of TLR3 were found during a genomic-wide sequencing of several ethnic groups (Lazarus et al., 2002). Four of the SNPs cause amino acid changes at the coding region of TLR3. Although no disease has been directly related to these SNPs, in vitro analysis may provide some clues of their effect in vivo.

Results

2D crystallization of TLR3 ECD

TLR3 ECD was purified as a 6His-tagged protein from human embryonic kidney cells (HEK293T). The lipid monolayer method with Ni-NTA-DOPS lipid, which has high affinity for the 6His-tagged protein, was used to screen different conditions for 2D crystallization of TLR3 ECD. The crystals shown in Figure 4-1 has $p2_12_12$ symmetry with unit cell dimension of a = 275.7 Å, b = 131.6 Å, and $\gamma = 90^{o}$. This symmetry means that there is a 2-fold axis in the Z direction, whereas the X and Y direction each has a 2-fold screw axis. The symmetric operation for a 2-fold screw axis is to rotate 180º around the axis, and then to translate half of the unit cell dimension along that axis.

The 2D projection of the crystal reveals that TLR3 ECD forms a characteristic horseshoe-shaped molecule, which is consistent with other LRR proteins (Bell et al., 2003), but much larger (~ 90 Å). The large surface of the horseshoe provides enough

space for ligand binding and protein-protein interaction, which is related to its function as the recognition site for pathogenic molecules.

Each asymmetric unit has four horseshoe-shaped molecules. Two kinds of dimer could be proposed. One dimer is related by 2-fold symmetry, with the corresponding terminal domains interacting with one another. The other dimer is formed by interaction of the terminal domain of one horseshoe with the middle domain of the other horseshoe. The crystal packing is very loose; nearly half of the space is empty.

In another crystal form, the particles packed more densely (Figure 4-2). The crystal has p2 symmetry with unit cell dimensions of a = 96.5 Å, b = 182.0 Å, and γ = 90º. Each unit cell has two horseshoe-shaped particles, or one possible dimer. Although this dimer also has a 2-fold axis, the position of interaction is different from the previous crystal. The middle of each horseshoe has less mass than the “arms”, so it can be concluded that the bright spot in the middle of each horseshoe in the previous crystal does not correspond to more mass. Instead, it is caused by stacking of the middle part of one horseshoe with the terminal domain of another horseshoe.

Dimer formation is thought to be important for TLR function (Choe et al., 2005). The packing of the crystals reveals several possible interactions between TLR3 ECD and several possible dimer formations. Whether these interactions exist in vivo requires further experiments. Structure analysis of full length TLR3 will be more meaningful in this aspect.

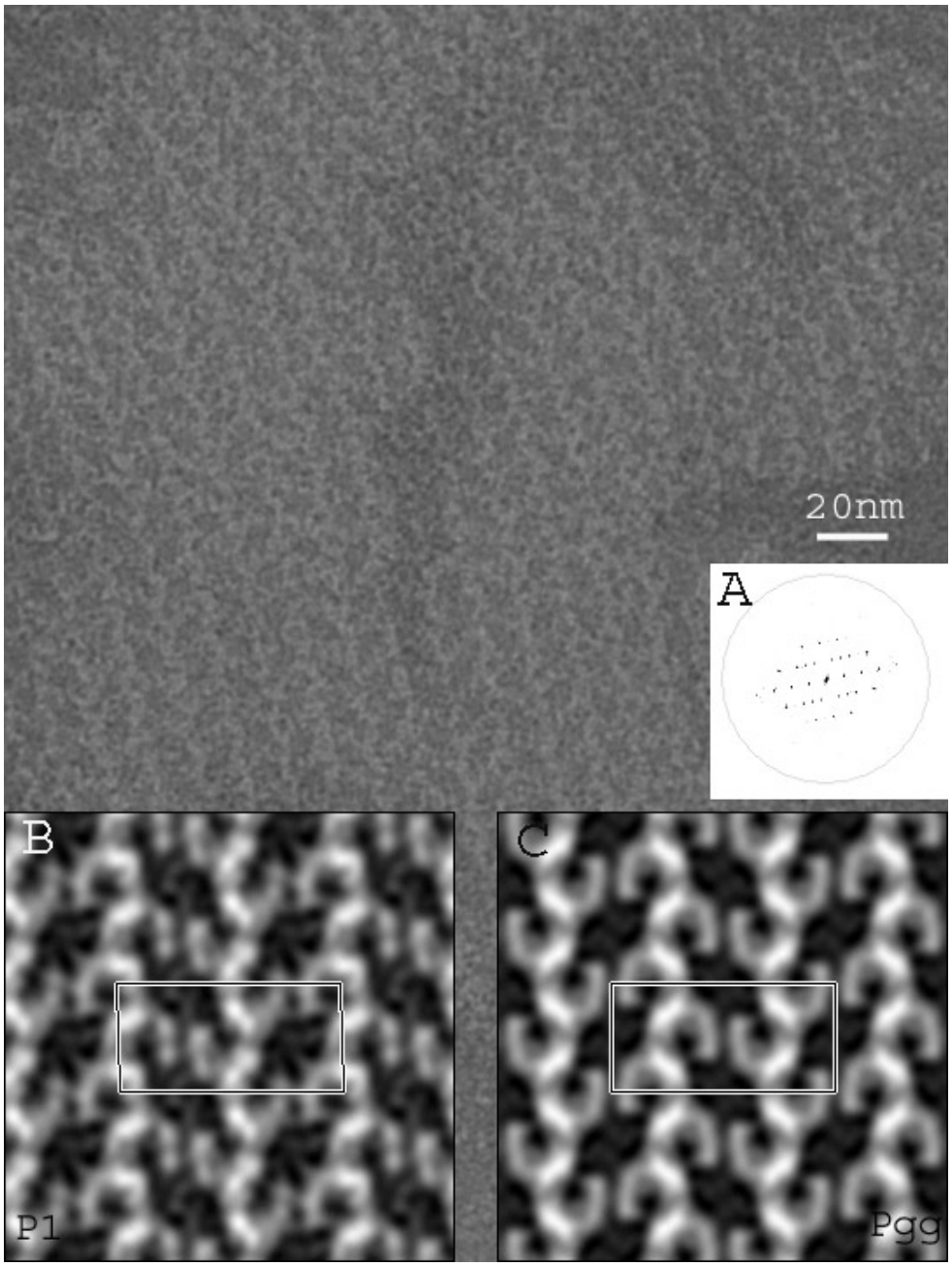

Figure 4-1. TLR3 ECD 2D crystal obtained by the lipid monolayer method. (A) FFT. (B) Projection map calculated with no symmetry imposed (p1). (C) Projection map calculated in $p2_12_12$. (a = 275.7 Å, b= 131.6 Å, and $\gamma = 90º$). One unit cell is highlighted in each map.

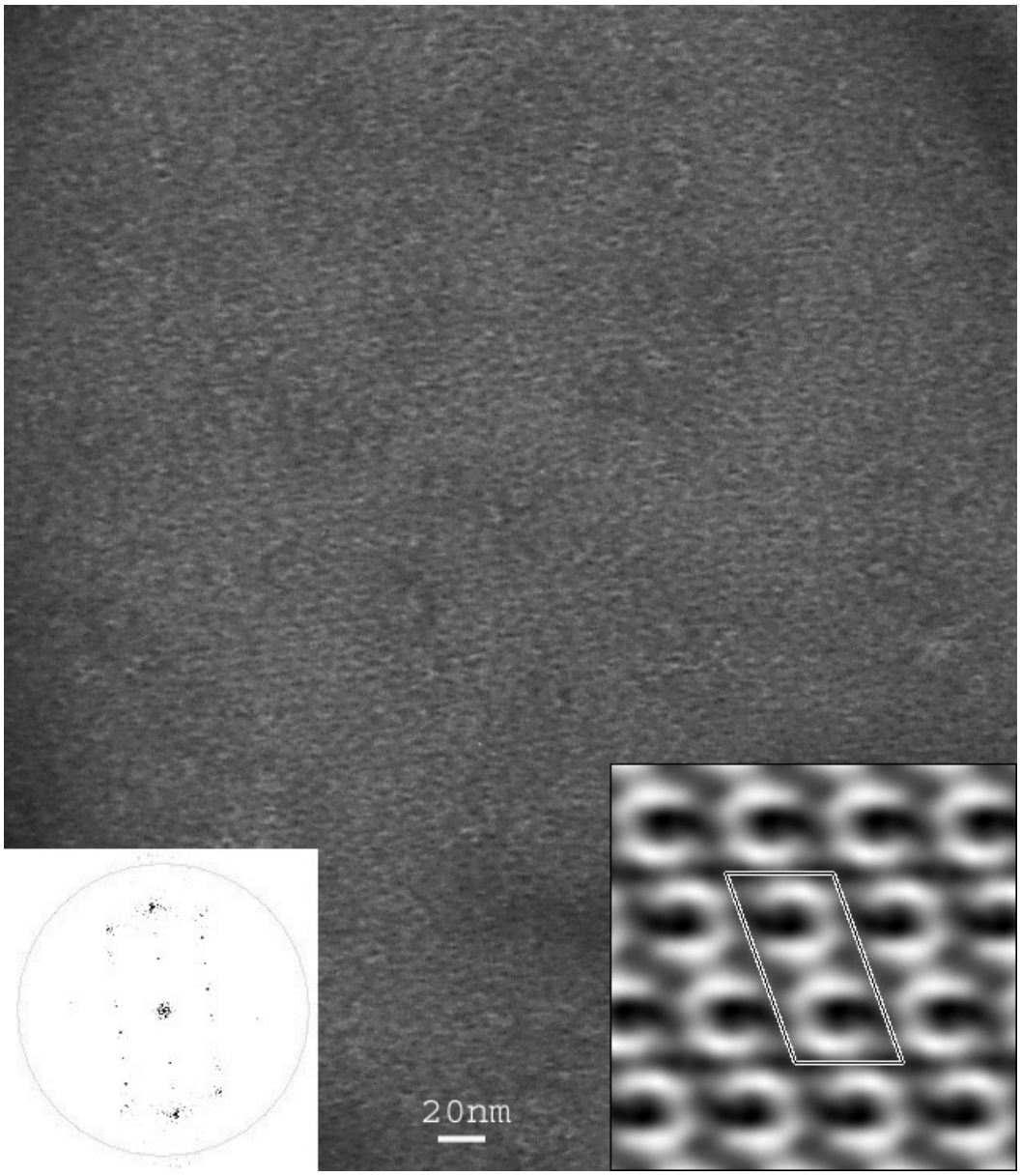

Figure 4-2. 2D crystal of TLR3 ECD with p2 symmetry. Inserts show FFT (left) and projection map (right). Unit cell parameters are: a = 96.5 Å, b = 182.0 Å, and γ = 109.9º. One unit cell is marked.

3D reconstruction of TLR3 ECD

Several thousand TLR3 ECD particles were automatically selected from electron micrographs of negatively stained sample. After classification and averaging, we observed not only the horseshoe-shaped particles that were seen in the 2D crystals, but also I-shaped particles (Figure 4-3), which correspond to the side view of the TLR3 ECD. Several horseshoe and I-shaped particles were selected and an initial 3D model was built with C1 symmetry using 'startany' of the EMAN software package. After several cycles of refinement, the 3D models were converged and the resolution of the final 3D reconstruction is about 25 Å from EO-test. Some orientations are missing in the asymmetric triangle, which is often the case for asymmetric particles on carbon film. The 3D reconstruction of the TLR3 ECD represents the smallest asymmetric molecules that have been successfully reconstructed using the single particle method.

The TLR3 ECD was also imaged in holey carbon film with trehalose/ammonium molybdate. Classification and averaging of the particles using the IMAGIC software package reveals that the particles are heterogeneous. Most of the particles resemble the previously mentioned horseshoe shape, but some particles look like a ring with only a small opening, or with no obvious opening at all (Figure 4-4B). The differences in shape may reflect flexibility of the horseshoe frame in solution.

Lectin, which can bind sugar motif of glycoprotein, was added to the TLR3 ECD and imaged in TEM. Extra density can be seen around the horseshoe shaped particles (Figure 4-4A), conforming that TLR3 ECD has multiple glycosylation sites.

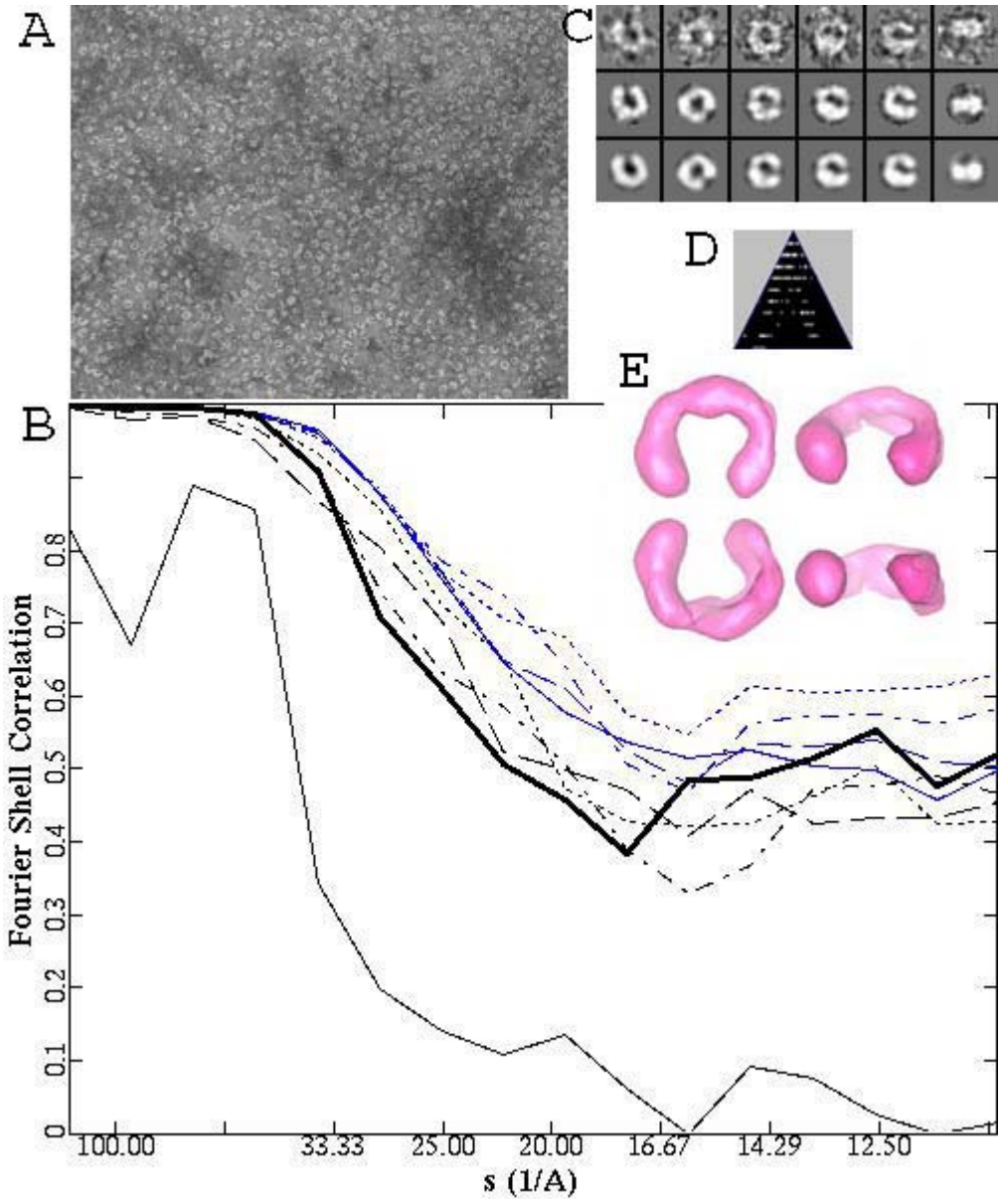

Figure 4-3. 3D reconstruction of TLR3 ECD. (A) Electron micrograph of negatively stained TLR3 ECD on a carbon support film. (B) FSC and EO-test (bald line). (C) Single particle analysis (top row: raw particles; middle row: class averages; bottom row: re-projections.) (D) The asymmetric triangle of the data set. (E) Surface representation of the 3D structure with the molecule mass set to 100 kDa.

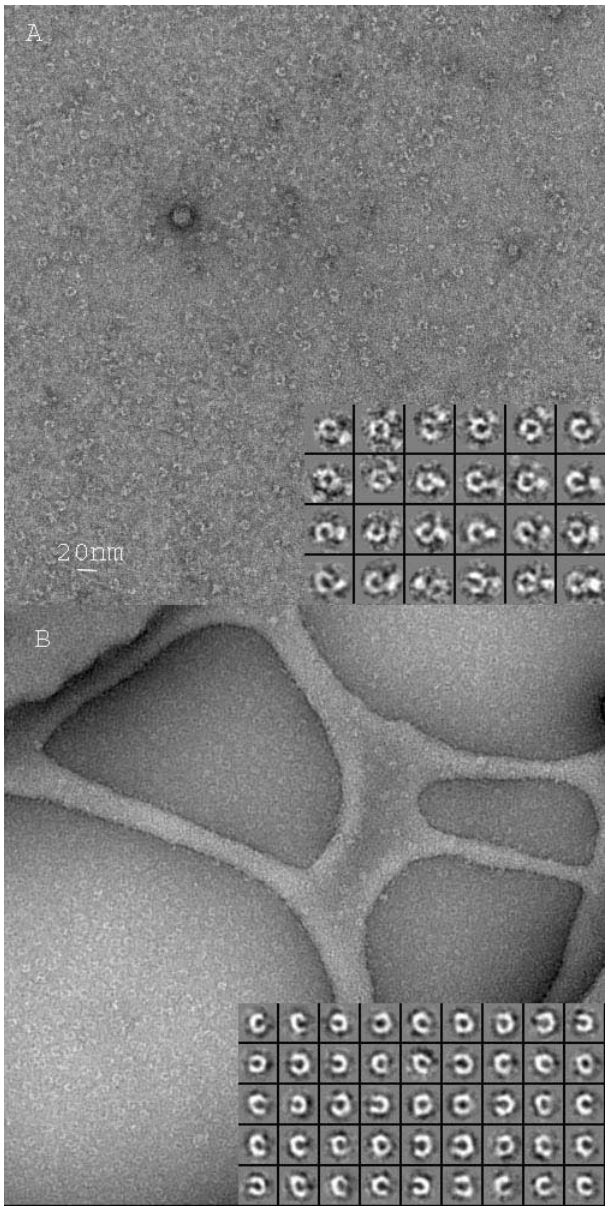

Figure 4-4. Single particle analysis of TLR3 ECD. (A) TLR3 ECD molecules on a continuous carbon support film after incubation with *Ricinus communis* lectin and negative stained with uranyl acetate. The gallery highlights projections that show extra densities around the outer perimeter of TLR3 ECD. (B) TLR3 ECD molecules embedded /negatively stained with trehalose/ammonium molybdate and suspended over a holey carbon film. The insert shows a gallery of horseshoe-shaped projections.

3D model of TLR3 ECD

TEM method can be thought as ‘top down’, which is good at determining low resolution structure, like the shape of the protein and the arrangement of domains. Although the resolution of TEM method is constantly improving, it is still difficult to see atomic details at this time. Homology modeling can be thought as ‘bottom up’, which creates 3D structure at atomic resolution from protein sequence directly. The modeling method is accurate in predicating local peptide folding (secondary structure), but is not reliable for long-range interaction. The 3D structure may be completely wrong at “low-resolution” (shape, domain arrangement). Using low-resolution TEM 3D reconstruction as constrain, we can produce more reliable 3D model at both high and low resolution.

The TLR3 ECD sequence (gi4507531) was used to search the Protein Data Bank (PDB) and a small LRR proteoglycan was found to have 25.9% amino acid identity with one-third of TLR3 ECD (K27 - F294) (Figure 4-5). Homology modeling with similarity at this level usually produces a reliable 3D model. Using different regions of TLR3 ECD, templates for L253 – L471 and I461 - I668 were found with 22.6% and 29.6% identity, respectively (Figure 4-5). The three parts cover the majority of the TLR3 ECD with some overlapping sequences, which were used as a guide for aligning the three models together. The final 3D model resembles the TEM structure, but there are some noticeable differences (Figure 4-6 and 4-7). These differences may be due to glycosylation of TLR3 ECD that was expressed in HEK293T cells. Homology modeling does not consider protein glycosylation.

```
Entry: 1xcd  Identity: 25.9 %
Query: 1   KCTVSHEVADCSHLKLTQVPDDLPTNITVLNLTHNQLRRLPAANFTRYSQLTSLDVGFNT 60
           +C     V  CS L L +VP DLP +  +L+L +N++  +   +F     L +L +  N
Templ: 1   RCQCHLRVVQCSDLGLEKVPKDLPPDTALLDLQNNKITEIKDGDFKNLKNLHTLILINNK 60

Query: 61  ISKLEPELCQKLPMLKVLNLQHNELSQLSDKTFAFCTNLTELHLMSNSIQKIKNNPFVKQ 120
           ISK+ P     L  L+ L L  N+L +L +K       L EL +  N I K++ + F
Templ: 61  ISKISPGAFAPLVKLERLYLSKNQLKELPEK---MPKTLQELRVHENEITKVRKSVFNGL 120

Query: 121 KNLITLDLSHNGLSSTKL--GTQVQLENLQELLLSNNKIQALKSEELDIFANSSLKKLEL 180
             +I ++L  N L S+ +  G    ++ L  + +++  I  +           SL +L L
Templ: 121 NQMIVVELGTNPLKSSGIENGAFQGMKKLSYIRIADTNITTIPQG-----LPPSLTELHL 180

Query: 181 SSNQIKEFSPGCFHAIGRL--FGLFLNNVQLGPSLTEKLCLELANTS-IRNLSLSNSQLS 240
             N+I +        +  L   GL  N++    + +      LANT  +R L L+N++L
Templ: 181 DGNKITKVDAASLKGLNNLAKLGLSFNSISAVDNGS------LANTPHLRELHLNNNKLV 240

Query: 241 TTSNTTFLGLK-WTNLTMLDLSYNNLNVVGNDSF 274
                 GL     + ++ L  NN++ +G++ F
Templ: 241 KVPG----GLADHKYIQVVYLHNNNISAIGSNDF 274

Entry: 1xcd  Identity: 22.6 %
Query: 1   LSLSNSQLSTTSNTTFLGLKWTNLTMLDLSYNNLNVVGNDSFAWLPQLEYFFLEYNNIQH 60
           L L N++++   +  F  LK  NL  L L  N ++ +   +FA L +LE  +L  N ++
Templ: 1   LDLQNNKITEIKDGDFKNLK--NLHTLILINNKISKISPGAFAPLVKLERLYLSKNQLKE 60

Query: 61  LFSHSLHGLFNVRYLNLKRSFTKQSISLASLPKIDDFSFQWLKCLEHLNMEDNDIP--GI 120
           L              + ++  +  +    + K+    F  L  +  + +  N +   GI
Templ: 61  LPE------------KMPKTLQELRVHENEITKVRKSVFNGLNQMIVVELGTNPLKSSGI 120

Query: 121 KSNMFTGLINLKYLSLSNSFTSLRTLTNETFVSLAHSPLHILNLTKNKISKIESDAFSWL 180
           ++  F G+  L Y+ +++  T++ T+      SL       L+L  NKI+K++ + +  L
Templ: 121 ENGAFQGMKKLSYIRIAD--TNITTIPQGLPPSLTE-----LHLDGNKITKVDAASLKGL 180

Query: 181 GHLEVLDLGLNEIGQELTGQEWRGLENIFEIYLSYNKYLQL 221
            +L  L L  N I     G       ++ E++L+ NK +++
Templ: 181 NNLAKLGLSFNSISAVDNGS-LANTPHLRELHLNNNKLVKV 221

Entry: 1ook  Identity: 29.5 %

Query: 1   IYLSYNKYLQLTRNSFALVP--SLQRLMLRRVALKNVDSSPSPFQPLRNLTILDLSNNNI 60
           ++LS N  L  T +   L+P   L +L L R  L  +    +    L  L  LDLS+N +
Templ: 1   LHLSEN--LLYTFSLATLMPYTRLTQLNLDRAELTKLQVDGT----LPVLGTLDLSHNQL 60

Query: 61  ANINDDMLEGLEKLEILDLQHNNLARLWKHANPGGPIYFLKGLSHLHILNLESNGFDEIP 120
            ++   + + L  L +LD+  N L  L        P+  L+GL  L  L L+ N    +P
Templ: 61  QSL-PLLGQTLPALTVLDVSFNRLTSL--------PLGALRGLGELQELYLKGNELKTLP 120

Query: 121 VEVFKDLFELKIIDLGLNNLNTLPASVFNNQVSLKSLNLQKNLITSVEKKVFGPAFRNLT 180
             +     +L+ + L  NNL  LPA + N   +L +L LQ+N + ++ K  FG     L
Templ: 121 PGLLTPTPKLEKLSLANNNLTELPAGLLNGLENLDTLLLQENSLYTIPKGFFGSHL--LP 180

Query: 181 ELDMRFNPFDCTCESIAWFVNWINETHTNI 210
              +  NP+ C CE I +F  W+ +   N+
Templ: 181 FAFLHGNPWLCNCE-ILYFRRWLQDNAENV 210
```

Figure 4-5. Sequence alignments for homology modeling.

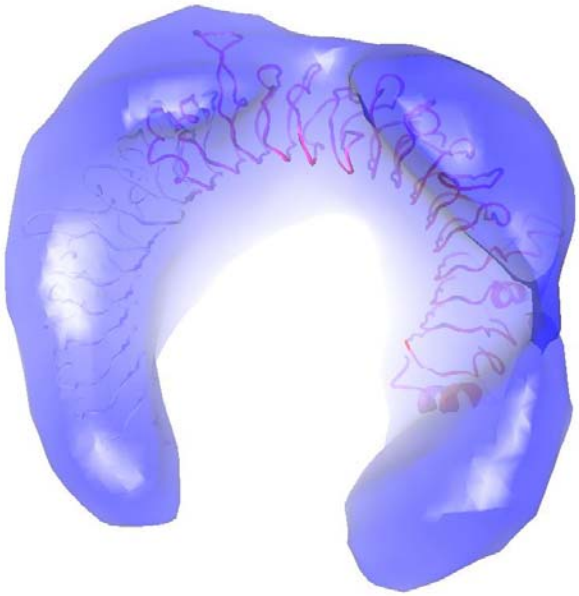

Figure 4-6. TLR3 ECD model docked to TEM reconstruction. The ribbon diagram is from homology modeling, whereas TEM reconstruction is shown as continues volume.

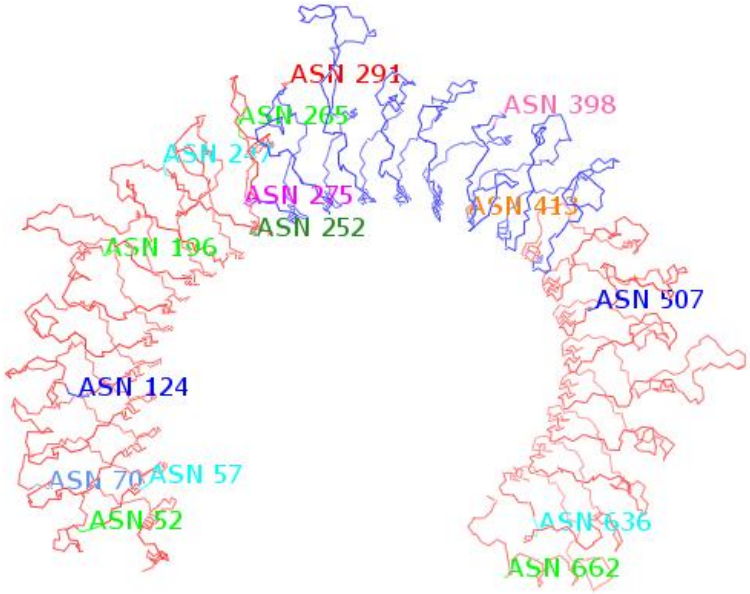

Figure 4-7. Model of TLR3 ECD highlighting the N-glycosylation sites.

The expression and activity of TLR3 SNPs

Cultured HEK293T cells were co-transfected with TLR3 and reporter plasmits. Stimulating TLR3 with substrate polyIC activates the transcription factor NF-κB, which increases the expression of the report enzyme (firefly luciferase). The expression of renilla luciferase is under the control of a constitutive activated promote, which is used as a control. The ratio of firefly luciferase activity to *Renilla* luciferase activity after the stimulation of wild-type TLR3 is four times higher than the control (Figure 4-8). This cell-based dual-luciferase activity assay was used to test the effect of TLR3 mutations.

There are many single nucleotide polymorphisms (SNPs) of TLR3 in human population (Lazarus et al., 2002); four of them are at coding region of TLR3 and cause amino acid substitutions. Two of the SNPs, Y307D and S737T, had the same level of activity as wild type TLR3 (~ 7.8). The L412F substitution reduced TLR3 activity to ~ 5.6, and N284I had even less activity (~ 2).

In addition to testing the TLR3 SNP function by measuring the induced luciferase activity, we also determined the expression level of the TLR3 SNPs. Western blot of the transfected cells with TLR3 antibody shows that the expression level of TLR3 SNP was similar (Figure 4-9). This result shows that the reduced activity of N284I and L412F substitutions is not caused by changes in expression level. The underline reason needs more studies. The mutations may influence the binding to substrate directly, or change the structure of the protein. Of note, N284 and L412 of TLR3 are conserved among all species from human to fish, whereas Y307 and S737 are only conserved in mammals (data not shown).

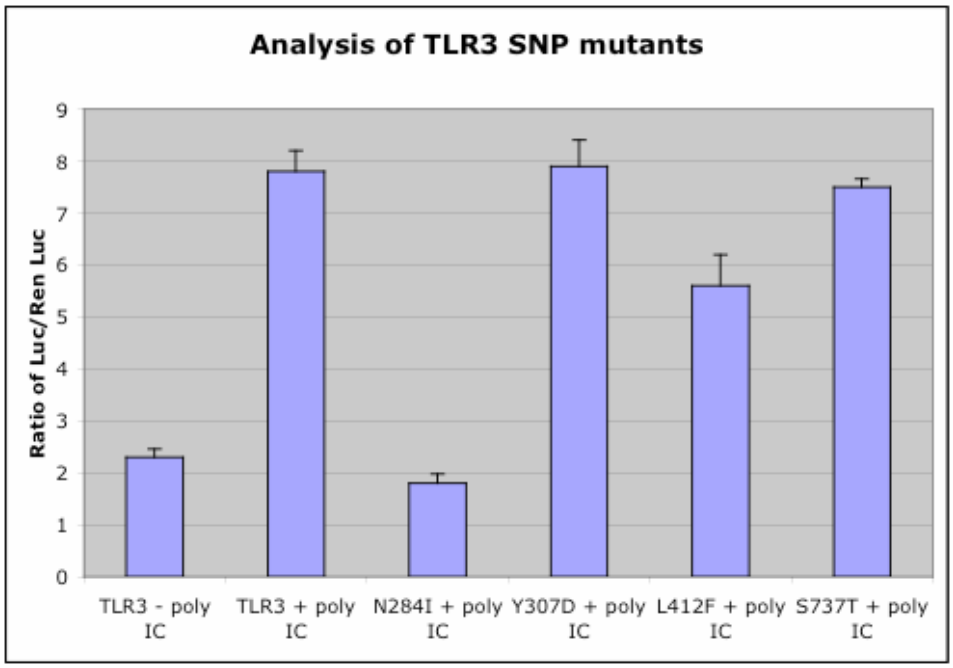

Figure 4-8. Cell-based activity of TLR3 SNPs. TLR3 activity was expressed as luciferase ratio with the stimulation of substrate polyIC. The first two lanes are wild type TLR3 with or without polyIC stimulation. The last four lanes are the activity of the single amino acid mutants of TLR3 (Ranjith Kumar, personal communication).

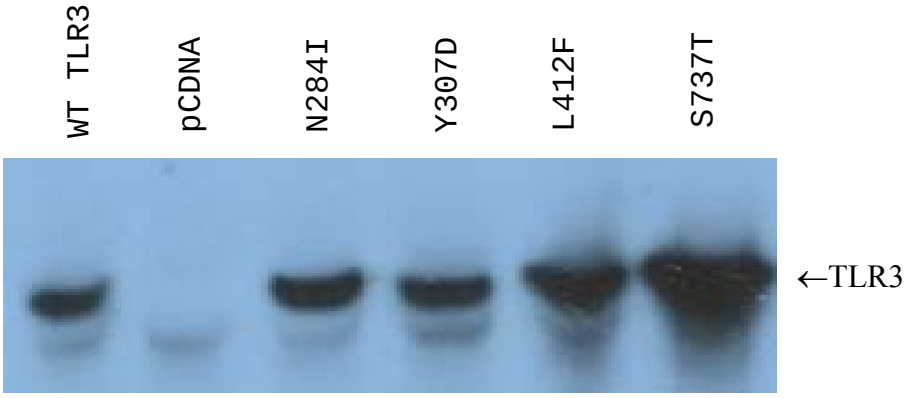

Figure 4-9. Western blot showing the expression level of TLR3 SNPs. HEK293T cells transfected with TLR3 SNP plasmids were probed with a TLR3 antibody. The first two lanes are wild type TLR3 and control plasmid without TLR3. The last four lanes are the four TLR3 mutants (Ranjith Kumar, personal communication).

Glycosylation of TLR3 ECD

To determine the importance of glycosylation on the function of the TLR3, we used tunicamycin to inhibit the addition of N-acetylglucosamine, the first sugar of N-linked glycosylation. Again, we stimulated the TLR3 expressed in HEK293T cells and measured the induced luciferase activity. The addition of tunicamycin inhibits TLR3 activity in a concentration-dependent manner (Figure 4-10A). Using fluorescence-activated cell sorting (FACS) with a monoclonal antibody to the TLR3 ECD, tunicamycin was also shown to affect the cell surface expression of TLR3 (data not shown). These results indicate that glycosylation affects the expression of TLR3.

There are 15 potential N-glycosylation sites in the TLR3 ECD with the consensus Asn-X-Ser/Thr sequence (Figure 4-7). Each asparagine residue in the proposed N-glycosylation sites was mutated to an alanine and the function of the TLR3 variants was tested using the cell-based assay. Most mutants did not influence the TLR3 signaling, only two of mutations (N247A and N413A) lowered the activity significantly (Figure 4-10B and C). Double mutations of N247A and N413A lower the activity even more. The reason for the decrease in TLR3 function of these two mutants needs to be studies in more detail, as the expression level of N247A and N413A mutants did not change from Western Blot (data not shown). It seems that each individual N-glycosylation site of TLR3 is not as important as each N-glycosylation of TLR4, which had been shown to be essential for the cell surface expression of TLR4 (da Silva and Ulevitch, 2002).

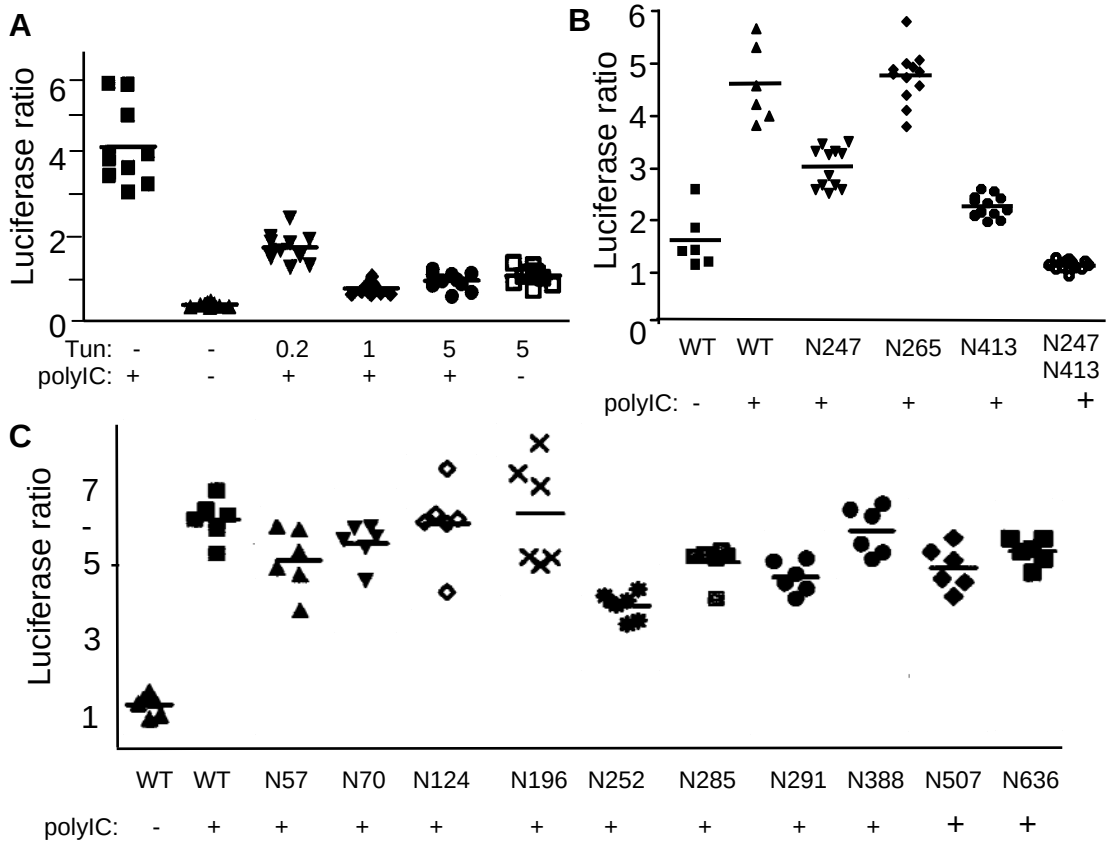

Figure 4-10. Mutational analysis of glycosylation sites of TLR3. (A) TLR3 activity (expressed as luciferase ratio) in the presence and absence of tunicamycin (Tun) and substrate (polyIC). (B and C) Activity after mutating asparagine residues to alanine. In (C) the double mutation (N247A and N413A) lowers the activity more than each of the individual mutation (Cheng Kao, personal communication).

Discussion

From the TEM studies, we have shown that TLR3 ECD is a large horseshoe-shaped molecule and exists as monomer in solution. The molecular mass of TLR3 ECD is 115 kDa with 35% mass belong to sugar. It is ordinarily impossible to perform single particle 3D reconstruction of such small proteins without any symmetry. Single particle method is thought to work only on protein complexes larger than 200 kDa. The

successful 3D reconstruction of TLR3 ECD proves that this method can be applied to small proteins with special shapes.

The horseshoe-shape of TLR3 ECD is consistent with other LRR proteins, but TLR3 ECD is much larger with 90 Å diameter. Proteins with LRRs are often participates in ligand binding or protein-protein interactions. This is consistent with the function of TLR3 ECD as the ligand-binding domain (Bell et al., 2003).

How substrate (dsRNA) binds TLR3 ECD is not clear from the protein structure. It was proposed that dsRNA might bind to the concave surface, like most other LRR proteins, or on the convex surface. Because there is no structural or biochemical evidence that TLR3 interacts dsRNA directly, it is also possible that dsRNA binds TLR3 indirectly through an adaptor protein. Our TEM structure shows that the inside of the horseshoe-shaped TLR3 ECD, which is fully glycosylated and functional, is large enough to accommodate a dsRNA. Attempt to visualize the binding directly in TEM did not success. This may implicate that the binding of dsRNA to TLR3 ECD is very weak, or an adaptor protein is needed for their interaction.

We have crystallized 6His-tagged TLR3 ECD using the lipid monolayer method with Ni-NTA-DOPS lipid. Several different interactions between the monomers are visualized from the electron crystallography analysis, implicating several possible dimer formations. Whether these dimers exist *in vivo* need other independent experiment to conform.

The mutation analysis of TLR3 SNP and N-glycosylation revealed several sites that are important for TLR3 signaling on the cell-based activity assay. The exact

mechanism how these mutants interferes TLR3 functional need more investigation. For the SNP N284I and L412F, comparative analysis of normal persons to those with the mutation will give some clues on the role of TLR3 *in vivo*.

The structure and function of glycosylation in TLR3 needs more studies. For the N247A and N413A mutation, the role of these asparagines residues in the structure and function of TLR3 are not clear. As these mutations did not influence the expression of TLR3, they may have more direct role in the TLR3 signaling.

The importance of the immune system for human health cannot be overstated. Considering the pivotal role of TLRs in sensing infection of pathogenic microbes and initiating immune response (Iwasaki and Medzhitov, 2004), it is a surprise that they were only discovered recently. In collaboration with Drs. Cheng Kao, Jin Xiong and Lamine Mbow's labs, we have studied some aspects of the TLR3 structure and function *in vitro*. The next step is to improve the crystals of the fully glycosylated TLR3 ECD and get high-resolution structure of the protein. How dsRNA interacts with TLR3 ECD is a major question waiting to answer. The same technique used in TLR3 studies could also be applied to the structural and functional analysis of other TLRs.

Materials and Methods

Preparation of holey-carbon grids

To prepare holey-carbon grids (Harris and Scheffler, 2002), 0.34 g Formvar powder was dissolved in 100 ml of chloroform. 100 drops of 50% glycerol/water solution were added and sonicated for several minutes. A clean glass slide was dipped

into the Formvar solution; after removal from the solution, excess fluid was drained off and the slide was dried in a dust-free environment. The glass used to make knives for ultra-thin sectioning was the most suitable for flowing off the Formvar film on water surface.

The edge of the glass slide was scratched with a knife and slowly the slide was immersed into the water to flow off the Formvar film on water surface. Cleaned grids were placed onto the Formvar film, and then white office paper was placed on top of the grids. After the paper got sufficiently wetted, the "sandwich" of "Formvar film-grids-paper" was picked up and placed in a covered petri dish, with the Formvar side up, and allowed to dry.

The paper with the coated grids was placed onto a methoanol-soaked filter paper in a covered petri dish for ~ 30 minutes in order to perforate *pseudo*-holes of the support film. After drying, the quality of the holey Formvar film was checked in a light microscope.

A thin carbon (~ 10 nm) was evaporated directly onto the Formvar-coated grids. Again, the paper together with the coated grids was placed on filter paper that had been soaked with ethylene dichloride and incubated for 30 minutes in a closed petri dish. This incubation removed the Formvar and left only carbon film.

Electron microscopy of TLR3 ECD

The cDNA of TLR3 (U88879) extracellular domain, corresponding to amino acids 1 – 703, was cloned into pcDNA3.1 in frame with a C-terminal 6His-tag.

HEK293T cells transfected with the TLR3 ECD plasmid expressed a truncated protein (N-terminal 26 amino acids was cleaved), which was secreted to the cell supernatant. Ni-NTA and gel filtration chromatography were used to purify the 6His-tagged TLR3 ECD.

For single particle imaging, 2.5 μl 50 μg/ml TLR3 ECD in Tris-buffered saline (pH 6.9) was applied to glow-discharged carbon grid, washed with distilled water and stained for 1 minute with an aqueous solution of 1% uranyl acetate. Alternatively, the protein solution was applied to holey carbon films and stained with 5% ammonium molybdate and 1% trehalose. To examine the extent of glycosylation in TLR3 ECD, *Ricinus commonis* lectin (L2390, Sigma) was mixed at a 2:1 ratio to the TLR3 ECD and stained with uranyl acetate.

2D crystal of TLR3 ECD was grown in a buffer containing 100 mM Na_3Citrate, pH 5.6, 100 mM Li_2SO_4 and 0.6 M sodium formate by the lipid monolayer method with Ni-NTA-DOPS (100 μg/ml) and egg PC (400 μg/ml) as descript before.

Single particle analysis with IMAGIC

Some image analysis was performed using the IMAGIC-5 software package (van Heel et al., 1996). The selected particles were filtered first with the IMAGIC-5 program, and an iterative reference-free alignment was used to center the particles. The image densities of all the particles were combined and rotational averaged to produce a symmetric ring. All the particles were aligned to this ring and centered by moving in the image plane. A new ring with the centered particles was produced and another round of

alignment was applied to center the particles. This process was repeated several times until movement of the particles became very small.

Multivariate-statistical analysis (MSA) was applied to separate the raw particles into defined numbers of classes, and the particles in each class were aligned and averaged. Several characteristic views from the class averages were selected and a multi-reference alignment (MRA) was used to realign the raw particles with the selected particles as references. Original filtered particles (no movement or rotation was applied) were used in the MRA to reduce any interpolation errs that may deteriorate the resolution. The MSA and MRA process was repeated many times to produce the final classification and averaging.

3D modeling of TLR3 ECD

Homology modeling of TLR3 ECD was performed with the CPH server (Lund, 1997) (http://www.cbs.dtu.dk/services/CPHmodels). The TLR3 ECD sequence (gi4507531) was compared to all the sequences in Protein Data Bank (PDB). After identifying a template that had a high level of similarity, an improved alignment between query and template sequence was constructed by profile-profile alignment using a Blosum62 matrix (Henikoff and Henikoff, 1992). The corresponding residues derived from the alignment were used as a starting point for homology modeling. Missing residues were added using the segmod program (Levitt, 1992) from the GeneMine package. The structure was then refined with the encad program (Levitt et al., 1995).

Docking atomic model into TEM density map

Chimera program was used to docking the atomic model into the TEM density map. The scan scale of TEM map was changed to 1 Å/pixel using 'proc3d' program of EMAN package with the 'scale' parameter. Alternatively, the Å/pixel value in the 'data set' of Chimera user interface was change to the real scan step (5.14 Å/pixel). In the 'model' panel, the TEM structure was inactivated or fixed. The atomic model was moved and rotated to manually docking into the TEM density map, which was set to be transparent so the ribbon diagram was visible in the TEM map.

Mutational analysis of TLR3

N-glycosylation sites in TLR3 with consensus Asn-X-Ser/Thr motif were predicted from web-based software (http://www.cbs.dtu.dk/services/NetNGlyc). Each asparagines residue at the predicated N-glycosylation sites was substituted to alanine with the QuickChange kit of Stratagene Inc.

136 single nucleotide polymorphisms (SNPs) of TLR3 in human population were found in a genomic wide sequencing of several ethnic groups (Lazaurus et al., 2002) (http://innateimmunity.net/IIPGA2/PGAs/InnateImmunity/TLR3/). Four of the SNPs (N284I, Y307D, L412F, and S737T) are at the coding exons of TLR3 and cause amino acid changes. QuickChange kit was used to create these SNPs for *in vitro* assay.

Cell-based activity assay of TLR3

The Dual-Luciferase Reporter Assay System (Promega) was used to exam TLR3 activity. Passive Lysis Buffer (PLB), Luciferase Assay Reagent, Stop & Glo Reagent were prepared according the manufacturer's manual. Cell growth medium was prepared by adding 10% heat-inactivated fetal bovine serum, 2 mM L-glutamine to Dulbecco's Modified Eagles Medium (DMEM). Dulbecco's phosphate buffered saline (DPBS) was obtained from Cambrex. Lipofectamine 2000 (LF2K) (#11668-070) and Opti-MEM I Reduced Serum Medium (#31985-070) were from Invitrogen.

The full-length human TLR3 gene was cloned into the pcDNA 3.1 vector (pcDNA-TLR3, MolBio group). A NF-κB *cis*-reporting vector (pNF-κB-Luciferase from Stratagene #219077) was used to detect the intracellular signaling upon stimulation of TLR3 by substrate polyIC (Amersham #27-4732-01). Vector phRL-TK with *Renilla* luciferase gene (Promega #E6241) was used as a control.

HEK293T cells were grown to 90% confluency, the medium was removed, and the cells were washed with DPBS buffer. Trypsin/EDTA solution (Life Technologies #25300-054) was added and incubated 2 minutes at room temperature for the cells to detach from the flask. Cells were resuspended and collected by centrifugation at 3000g for 2 minutes with a desktop centrifuge, washed two times and resuspended in DMEM medium.

The cells were counted with a VWR counting chamber (#15170-208). 10 μl of cells were added to the chamber and the total number of cells was counted (100 cells in 16 wells corresponded to 1×10^6 cells/ml). Cells were diluted to 2.2×10^5/ml in DMEM,

and 200 µl cells were added to each well in a 96-well assay plate and incubated for 24 hours at 37 °C.

Cells were grown to 85-95% confluency. Three plasmids (15 ng pcDNA-TLR3, 30 ng pNF-κB-Luciferase, and 5 ng phRL-TK) were added to 5 µl OptiMEM medium. 18 µl LF2K was diluted in 300 µl OptiMEM media in a separate tube (5 µl/well). After 5 minutes at room temperature, the plasmids and LF2K were mixed and incubated 20 minutes at room temperature. The mixture was added to the cells and incubated for 24 hours at 37 °C.

The growth medium was aspirated from the wells, substrate polyIC was added to final concentration of 1 µg/ml in 75 µl serum-free DMEM and incubated for 24 hours at 37 °C.

The growth medium was aspirated from wells; 100 µl DPBS buffer and 20 µl PLB were added to each well and incubated 10 minutes at room temperature. The luciferase activity was measured with Luciferase Assay Reagent and Stop & Glo Reagent in the first and second injector of the BMG FLUOStar Optima luminometer. The ratio of the luciferase activities was calculated for each well.

CHAPTER V

DISCUSSION AND CONCLUSIONS

The Choice of Structure Study Method

There are three techniques to determine the three-dimensional (3D) structure of biological macromolecules: X-ray crystallography, NMR, and transmission electron microscopy (TEM). This thesis focuses on the TEM method.

The TEM method has three approaches: electron crystallography, single particle 3D reconstruction, and electron tomography. Electron crystallography works on crystallized macromolecules, whereas the single particle method works on purified proteins directly without the need for crystals. Both methods depend on merging data from large numbers of molecules, using either an experimental approach (diffraction of crystals) or a computational approach (alignment and merging of single particles). An assumption of both methods is that all the protein molecules in a purified sample are the same.

Electron tomography is the only way to determine the 3D structure of unique specimens. Even thought the resolution is relatively low, it is the only way to determine the structure as large as a whole cell. Because each specimen has a unique structure, we cannot merge data from different specimens. To perform 3D reconstruction, all the data is collected from the same specimen in different orientations.

There are also several ways to prepare samples for electron crystallography and single particle analysis, mainly negative stain and cryo-EM. Each method has its own

advantages and pitfalls. Some comparisons of the available methods are given in the following sections.

Crystallography and non-crystallography methods

Crystallization is a pre-requirement for determining the 3D structure of a biological macromolecule by X-ray and electron crystallography. Finding the crystallization condition for most proteins is difficult. A freshly purified protein sample is needed for each crystallization experiment, as many proteins aggregate, denature, or become heterogeneous during storage. It may take years to prepare high quality crystals for each protein. If no crystals can be obtained, or if the crystals do not diffract well, no structural information can be extracted. However, when good crystals can be made, the reward is great. Dr. Hartmut Michel was awarded the 1988 Nobel Prize for determining the crystal structure of the photosynthesis reaction center from *Rhodopseudomonas viridis* (Deisenhofer et al., 1985). Dr. John Walker was awarded the 1997 Nobel Prize for solving the crystal structure of F1-ATPase (Abrahams et al., 1994). It took many years' effort for Dr. Michel to crystallize the photosynthesis reaction center. To prepare F1-ATPase crystals, over 5 g of F1-ATPase had been purified from bovine heart mitochondria in Dr. Walker's lab.

On the other hand, NMR and the TEM single particle technique can determine the 3D structures of biological macromolecules in solution without the need for crystals. But these two methods also have limitations: NMR can only work on small proteins and needs very high protein concentrations. The TEM single particle technique can only

work on large proteins and the resolution is not as high as with the crystallographic method.

If good protein crystals have been obtained, the crystallographic method is the most powerful approach to determine the 3D structure at high-resolution. For proteins that cannot form good crystals despite many years' effort, we must either just wait for the miracle that perfect crystals suddenly appear, or use the non-crystallographic methods to obtain useful structural information.

2D crystal or 3D crystal

X-ray crystallography needs 3D crystals, and electron crystallography needs 2D crystals. Which method is preferred when trying to crystallize a protein? Although both approaches can build an atomic model if the crystals are of good quality, X-ray crystallography is more efficient in data collection and structural determination. Most steps in the X-ray method can be automated now, including crystal mounting, aligning the crystal in the synchrotron beam, collecting diffraction data, and structure refinement. If the crystals are perfect, it takes only several months, or even weeks, to solve the structure. On the other hand, most steps in electron crystallography can only be performed manually, and experience is needed to record high-resolution images and electron diffractions, especially at high tilt angles. Most of the programs used in electron crystallography are not well documented and experience is needed to use them properly. It may take several years to solve a structure by electron crystallography even with high-quality 2D crystals.

X-ray crystallography is the first choice for protein structural determination if the protein can form good 3D crystals. For those proteins that cannot form 3D crystals, 2D crystallization provides an alternative way to solve the 3D structure using electron crystallography. For membrane proteins, electron crystallography may have a more important role in their structure determination.

Membrane proteins exist within the lipid bilayer of a cell membrane and a large portion of the protein surface is hydrophobic. Only detergent can remove membrane proteins from this lipid environment and keep them soluble in aqueous solution. Crystallizing membrane proteins in detergents, which are usually heterogeneous, is extremely difficult. Even the pioneer of this method, Dr. Hartmut Michel, took seven years to crystallize the yeast cytochrome bc1 complex (Hunt et al., 2000). There are thousands of important membrane proteins that need to be studied. 2D crystallization of membrane proteins in a lipid bilayer is an alternative way to solve their structures by electron crystallography and membrane proteins are close to their native conformation in a lipid bilayer. Atomic resolution has been reached for 2D crystals of several membrane proteins, including bacteriorhodopsin (Henderson et al., 1990) and light-harvesting complex II (LHCII) (Kuhlbrandt et al., 1994).

Some soluble proteins are also resistant to forming 3D crystals, but can form 2D crystals. A good example is tubulin, which forms large and well-ordered 2D crystals. An atomic model of tubulin was built by electron crystallography (Nogales et al., 1998). The lipid monolayer technique provides an alternative way to prepare 2D crystals of proteins that are difficult to crystallize by other methods. Several examples in this thesis show

that valuable structural information can be obtained even from 2D crystals that are small and not perfect. This method is especially good for studying protein-protein interaction and ligand binding, as the protein molecules in 2D crystals are still accessible from bulk solution.

2D crystallization on a lipid monolayer

The lipid monolayer technique is a promising approach to crystallizing proteins. Although not every protein can be crystallized by this method, the success rate is much higher than for other crystallization methods. Taking advantage of the high affinity binding of 6His-tagged proteins to Ni-NTA lipid, crystallization by the lipid monolayer method may become as easy as running a Ni-NTA affinity chromatography. As more proteins are expressed and purified with a 6His-tag, we can expect that more proteins will be crystallized by this approach.

The requirement for only a small amount and low concentration of protein sample is a great advantage of the lipid monolayer approach. Many proteins cannot be expressed to high levels, or are difficult to purify in large amounts. At high protein and salt concentration, which is needed for 3D crystallization, some proteins are easily denatured. Some protein complexes can only be maintained at close to physiological conditions. In such cases, only the TEM single particle and lipid monolayer methods can be applied.

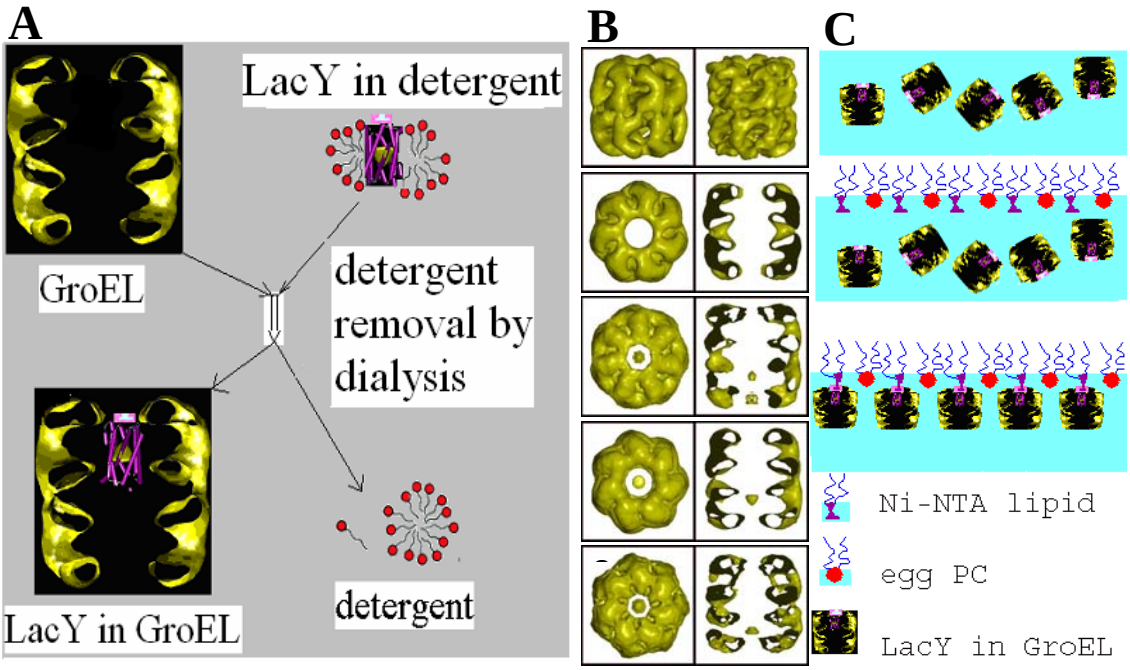

Figure 5-1. Co-crystallization of GroEL-membrane protein complex. (A) Diagram showing membrane protein LacY bound to GroEL after detergent removal. (B) 3D reconstruction of empty GroEL (top two) and GroEL with several membrane proteins (bottom three) bacteriorhdopsin, LacY, and S105 protein (Sun et al., 2005). An additional density can be readily discerned in the complexes when compared to empty GroEL. (C) Diagram of the lipid monolayer technique using the Ni-NTA lipid for the crystallization of 6His-tagged membrane protein and GroEL complexes.

Although the lipid monolayer method is designed for 2D crystallization of soluble proteins, membrane proteins can also be crystallized by this approach. In cooperation with Drs. Ryland Young and Ronald Kaback's labs, we found that membrane proteins could form a complex with GroEL after detergent was removed (Figure 5-1A). Single particle 3D reconstruction of membrane protein-GroEL complexes shows that the membrane proteins are located in the central cavity of GroEL (Figure 5-1B) (Sun et al., 2005). As two of the membrane proteins, LacY and S105, were expressed and purified with 6His-tags, the lipid monolayer technique was used to co-crystallize them with GroEL. With Ni-NTA lipid, the membrane proteins in the central

cavity of GroEL can bind to the Ni-NTA lipid and concentrate at the lipid surface, just like soluble proteins (Figure 5-1C). 2D crystals were indeed observed (Figure 5-2). This approach opens a new avenue for membrane protein crystallization.

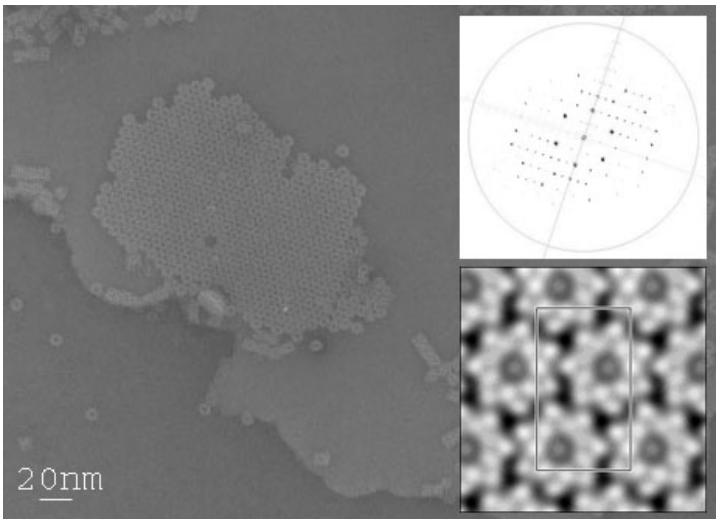

Figure 5-2. 2D crystal of 6His-tagged LacY-GroEL complexes. The inserts depict the FFT (top) and the corresponding projection map (bottom).

Perspective of lipid monolayer method

More effort is needed to make the lipid monolayer technique more fruitful. We need to understand the mechanism and find the optimal conditions for crystal formation. Automation at some steps will be a great help. Commercial support will make the researcher's work easier and more productive. The time-limiting step for the lipid monolayer technique is to prepare and examine 2D crystals in the TEM. Either automatic TEM operation or other alternative ways to prepare crystals will greatly increase the efficiency of this approach.

2D crystallization of the highly glycosylated and asymmetric human TLR3 ECD is a good example of how the lipid monolayer technique can deal with difficult cases. One of the most amazing facts is that the initial protein concentration was very low, only 20 μg/ml. The amount of protein used for crystallization was less than the amount needed for SDS-PAGE, but much more information was obtained. We not only could measure its size and mass, but also visualize its shape, oligomerization state, and substrate binding or protein-protein interaction.

From my experience, if a protein is not too flexible or too unstable, there is a good chance for it to be crystallized by the lipid monolayer technique after trying about 50 experimental conditions. Compared to 500 conditions typically tested for 3D crystallization, 2D crystallization has a higher success rate. If the protein forms a trimer, tetramer, or hexamer, the success rate is even higher. I am optimistic that more proteins will be crystallized by the lipid monolayer method in the future.

NMR and TEM single particle reconstruction

Both NMR and TEM single particle reconstruction can determine the 3D structure of biological macromolecules in solution without the need for crystallization, but they work on proteins of different sizes. The NMR method is best for small proteins and peptides, whereas TEM single particle reconstruction is more fit for large protein complexes. A good example is the structural study of circadian clock proteins KaiA and KaiC of cyanobacteria. The molecular mass of KaiA is only ~ 30 kDa, but it is still too large for the NMR method. As KaiA is composed of two structurally and functionally

independent domains, only one domain was studied at a time. The C-terminal domain of KaiA (180-283 aa), which interacts with KaiC, was cloned, expressed, purified, and its 3D structure was determined by the NMR method (Vakonakis et al., 2004).

The molecular mass of KaiC is ~ 60 kDa, nearly double that of KaiA, but the monomer form of KaiC is still too small for TEM single particle reconstruction. However, in the presence of ATP, KaiC forms a hexamer (Mori et al., 2002), which is large enough to build a 3D model at ~ 20 Å resolution by TEM single particle reconstruction. Furthermore, the even larger complex of KaiC and KaiA was visualized by the TEM method. To study the protein-protein interaction between KaiC and KaiA, only a small peptide of KaiC (488-518 aa) that binds the C-terminal KaiA was studied by the NMR method (Vakonakis and LiWang, 2004).

Negative stain and cryo-EM

Most projects described here used negative staining, but cryo-EM was also attempted on Nsp15. The advantages and disadvantages of negative staining and cryo-EM are summarized below.

The benefits of negative staining are that it is quick and easy, and does not require expensive equipment. The contrast is excellent, and the images are clear. Several thousand single molecule images are enough for a 3D reconstruction by the TEM single particle approach. The disadvantages of negative staining are the low resolution and lack of reproducibility. It is hard to get consistent results even when all the operations are kept the same. Different areas in the same grid may stain differently. Local heterogeneity

on the support film, in smoothness, hydrophobicity, or charge, may cause these differences.

Cryo-EM holds the promise to reach higher resolution than negative staining. A thin layer of protein solution on holey carbon film is frozen so fast that there is no time for ice crystals to form. Protein structure in the vitrified ice is essentially the same as in solution. The disadvantages of cryo-EM are that it needs expensive equipment, requires user experience, and consumes more time. More serious problems are the increased sensitivity of the protein specimen to radiation damage and the extremely low contrast of the images. Many more single molecule images are needed for cryo-EM than negative staining to increase the signal-to-noise ratio. It is very difficult to select particles from cryo-EM images. False positive and false negative selections cannot be avoided when you have to pick 30,000 particles with very low contrast.

Contrast transfer function (CTF) correction, which is essential at high resolution, is difficult to automate. Most steps still have to be performed by hand and require experience. Computer time and memory also become limiting when you work on large data sets. With a 50-unit Linux cluster, a data set of 30,000 particles needs weeks to perform the structural refinement.

Even with all these difficulties, cryo-EM single particle reconstruction may become more popular in research labs in the near future with faster and less expensive computers, more efficient software, and better TEM. As has been stated before, many protein complexes are difficult to crystallize, and they are also too large for the NMR method. Cryo-EM may be the only choice to solve their structure at high resolution. Life

is not simple. If we want to see the beauty and complexity of living organisms in great detail, we need sophisticated tools.

Combining different methods for protein complexes

For those protein complexes that are larger than 200 kDa, single particle 3D reconstruction is the first choice to determine their 3D structure at low resolution. A fruitful approach to determining the structure of large protein complexes at high resolution is to combine different techniques. The TEM method provides an overview of a large complex: the size, shape, and architecture. X-ray crystallography and NMR are preferred to determine the high-resolution structure of individual subunits. Docking the high resolution structure of each subunit into the low resolution structure of the large complex can provide a biological context for the subunits and also give more insight into how the large complex works.

Finally, all the available structural information is integrated with genetic, biochemical, and biophysical data along with bioinformatic analysis to provide a complete picture. One good example of this process is the study of ribosomes. After the first successful cryo-EM reconstruction, different conformations of ribosomes in different functional states during translation have been visualized, including the binding of tRNA, initiation factors, elongation factors, and termination factors (Frank, 2001). Combined with X-ray crystallography, modeling, and other tools, a clear and complete picture of the dynamic process of the translational machinery was revealed (Frank, 2001).

Structural-based drug design

Characterizing the structure and function of critical protein complexes is not only important for our understanding of the basic processes of living organisms, but also is the foundation for drug discovery. One good example is the invention of anti-HIV drugs (Tomasselli and Heinrikson, 2000). Although the human immune system is very efficient at defending against invading microbes, some bacteria and viruses have evolved mechanisms to evade the immune system. HIV viruses mainly infect and kill CD4+ T-cells. After a prolonged period, CD4+ T-cells are finally reduced to very low levels. When the immune system cannot mount an efficient defense, the HIV patient will be overwhelmed by many opportunistic infections. Understanding the biology of HIV infection and determining the structure of several critical proteins, including reverse transcriptase and proteinase, is the key to designing inhibitors of the virus. An inhibitor cocktail has shown to reduce the morbidity and mortality substantially in HIV patients (Tozser and Ososzlan, 2003).

Characterizing Biologically Important Macromolecules

Several biomedically important macromolecules were studied in this thesis. New structural information for these proteins was obtained from TEM and other analytic methods, and new techniques were established through these studies. Integrating different approaches, including genetic, biochemical, biophysical, and bioinformatic methods, is important for characterizing these macromolecules. At the start of a new project, bioinformatic analysis can provide valuable information to direct experimental

procedures. The time spent searching the literature and perusing bioinformatics databases is worth the effort. The experimental results obtained from one approach may be useful in other studies. Many ongoing projects in our lab prove the importance of a multi-disciplinary approach.

Gel filtration, analytical ultracentrifugation, native gel analysis, and UV crosslinking/SDS-PAGE all showed that the purified Nsp15 could exist as a hexamer. Without such biochemical experiments, we may not even have thought of trying the TEM method to visualize the hexamer. Even if I had tried, I may have given up after several failed attempts because no ring-shaped particles could be observed. The final decision to use D3 symmetry for the refinement was based on all the information available from biochemical, biophysical and structural analysis.

Biochemical studies showed that TLR3 is activated by substrate binding only at low pH (de Bouteiller et al., 2005). After initial experiments at different conditions, I screened more conditions at low pH and finally crystallized TLR3 ECD in a buffer of pH 5.6 by the lipid monolayer method.

Structure and function of IniA

Since the finding of IniA as a potential drug target, there has been great interest in crystallizing it and determining its structure. I had purified IniA more than 20 times and set up more than a thousand conditions for 3D crystallization. None of those conditions led to crystal formation. Therefore, I tried the lipid monolayer method and was able to acquire 2D crystals after trying about 100 conditions. Although the crystals

could not be reproduced, 2D crystallization seems relatively easier than 3D crystallization.

New approaches are needed for structural study of this kind of unstable protein. One idea is to fuse a "big tag" to the target protein. The fusion partner may help the target protein to become more soluble or more stable in solution. There are reports that fusions to maltose-binding protein (MBP) help to crystallize the target proteins (Smyth et al., 2003). Most fusion partners are monomeric proteins, which do not interfere with the oligomeric state of the target protein. A fusion partner with a similar oligomeric state as target protein may be better for structural study. The cyanobacterial clock protein KaiC will be a good candidate to fuse with IniA to make it more soluble and stable. KaiC has two similar domains (KaiC1 and KaiC2) and each domain can form a hexameric ring independently. If we can fuse IniA in the middle of KaiC, the hybrid protein may also form a hexamer. The hexameric KaiC1-IniA-KaiC2 would be a perfect specimen for TEM single particle reconstruction and the complex structure could be determined at a higher resolution than the KaiC hexamer itself. Crystallization of the hybrid protein may also be easier.

Genetic evidence indicates that IniA may function as a pump (Colangeli et al., 2005). As IniA is predicted to have only one transmembrane helix, it may form a complex with other proteins, including IniB and IniC, to form the pump. IniB is predicted to have four transmembrane helixes and may form the transmembrane channel. A long-term goal should be to co-express IniA, IniB, and IniC, purify them as a complex, reconstitute the pump activity, and determine the complex's structure.

Identification of other components in the pump is also important to understand the *in vivo* role of IniA.

Oligomeric state of Nsp15

Nsp15 is an interesting specimen to study protein folding and quaternary structure, as one single amino acid substitution can change the equilibrium of its oligomeric state. This thesis focused on the K289A mutant, which is predominately hexameric. Another mutant protein, H234A, resembles the wild-type protein and exists as both a monomer and a hexamer. Wild-type Nsp15 is toxic to *E. coli*, so its expression level is very low, rendering it difficult to purify for structural studies.

Many different crystal forms of Nsp15 H234A were observed with the lipid monolayer method (Sun, unpublished observations). One crystal was composed of small triangular-shaped particles, and another crystal had both small and big triangular forms. It is possible that the small triangular-shaped particle is an intermediate state between the monomeric and hexameric forms.

Single particle 3D reconstruction of the H234A mutant resembled K289A with minor differences. Unfortunately, at this resolution level (~ 20 Å) these small differences cannot be assigned to individual residues. Higher resolution is needed to explain the detailed changes in conformation at the amino acid level.

From sequence analysis, several other putative RNA processing enzymes were predicted from the SARS genome, including Nsp9, an RNA binding protein; Nsp 14, an RNA exonuclease; and Nsp16, an RNA methytransferase (Snijder et al., 2003). Nsp9 has

been shown to form a complex with Nsp15 (Satheeth, unpublished observation), and Nsp14 forms triangular-shape molecules larger than Nsp15 (Sun, unpublished result). It is possible that Nsp15 provides a "scaffold" for binding of other RNA processing enzymes. An interesting project will be to determine the 3D structure of the multi-enzyme complex.

Another example of a multi-protein complex of RNA processing enzymes is the "RNA degradosome" (Carpousis, 2002). RNase E of *E.coli* is an endonuclease involved in the processing of rRNA and the degradation of mRNA (Carpousis, 2002). During the purification of RNase E, two other enzymes, a 3' exoribonuclease and a DEAD-box RNA helicase, were identified. RNase E is shown to provide a "scaffold" that can bind PNPase, enolase and RhlB.

Ligand binding to TLR3 ECD

Although many experiments have shown that dsRNA can stimulate the TLR3 signaling pathway (Salio and Cerundolo, 2005), there is no biochemical or structural evidence that they interact with each other directly. It is possible that TLR3 recognizes dsRNA through an adaptor protein, much like the way that TLR4 recognizes LPS through an adaptor protein (Visintin et al., 2005). Clarifying the mode of dsRNA and TLR3 interaction will require more studies. Our 2D crystal analysis and single particle reconstruction both show that the inner diameter of TLR3 ECD is large enough to accommodate a dsRNA. Detailed structural information of TLR3 binding to ligand or an

adaptor protein will be the next major goal of our studies. Lipid monolayer and single particle reconstruction method will be powerful tools to solve this problem.

Glycosylation of TLR3

Many secreted proteins or cell surface proteins in mammalian cells are glycosylated. In general, glycosylation is important for correct folding of the glycoproteins (Parodi, 2000). Without proper glycosylation, some glycoproteins cannot be expressed normally or correctly localized.

TLR2 and TLR4 are the most well studied TLRs. There are four N-glycosylation sites in TLR2 and TLR4. All of the sites are important for expression of the proteins on the cell surface (Da Silva and Ulvitch, 2002; Weber et al., 2004). This led us to investigate the role of glycosylation in TLR3. TLR3 is highly glycosylated when expressed in human cells; about 35% of its mass is sugar. Treatment with tunicamycin, which inhibits N-linked glycosylation, prevented cell-surface expression of TLR3 and abolished the cell-based activity. There are 15 potential N-glycosylation sites on the extracelluar domain of TLR3. Mutational analysis showed that not all of these sites are necessary for TLR3 activity. Only two of the mutations, N247A and N431A, lowered activity by the luciferase reporter assay. Glycosylation of TLR3 seems not as important as glycosylation of TLR2 and TLR4.

Considering the extent of glycosylation of TLR3, it is a surprise that most substitutions at the predicated glycoslylation sites have little influence on TLR3 function. The *in vivo* roles of glycosylation in TLR3 need more study. Maybe mutation

at multiple sites at the same time will have a more profound effect on TLR3 expression and function, as N247A, N431A double mutants have shown.

When docking the 3D model of TLR3 ECD produced by homology modeling into the 3D structure obtained from TEM single particle reconstruction, some differences in the horseshoe-shaped structure were observed, which may due to glycosylation of the protein. Single particle analysis of TLR3 ECD bound with lectin demonstrates that there are multiple glycosylation sites around the outside of the horseshoe-shaped molecule. A higher resolution structure is needed to localize individual glycosylation sites and to visualize the extent of glycosylation.

Structural base of TLR3 SNPs

The *in vivo* function of TLR3 is to sense the invasion of pathogenic viruses and stimulate the production and release of inflammatory cytokines to alert the immune system (Schulz et al., 2001). The importance of TLR3 in immune responses is shown by the mechanisms used by several viruses to evade the TLR3 pathway (Schroder and Bowie, 2005). Vaccinia virus expresses two proteins to inhibit TLR3 signaling. A protease from hepatitis C virus also targets the TLR3 pathway.

Polymorphisms of TLR2 and TLR4 are correlated with susceptibility to many human diseases (Cook et al., 2004), but less is known about the polymorphisms of TLR3. The *in vitro* studies of TLR3 SNP can provide valuable information about their roles *in vivo*.

TLR3 is reported at different subcellular locations in different cell lines (Matsumoto et al., 2003). Human fibroblasts express TLR3 on the cell surface, whereas in immature dendritic cells, TLR3 predominantly resides inside the cells. In a more recent report, Nishiya et al. (2005) demonstrated by domain swapping and mutational analysis that the intracellular localization of TLR3 in bone marrow-derived macrophages is achieved by a 23 amino acid sequence (Glu^{727} - Asp^{749}) present in the linker region between the transmembrane domain and intracellular TIR. In HEK293T cells, SNPs Y307D and S737T, like wild-type TLR3, are expressed on the cell surface. Although the S737T mutation occurs in the middle of the linker sequence, the expression pattern and activity of TLR3 S737T does not change.

The TLR3 ECD has 23 leucine-rich repeats (LRRs) with conserved leucine or other hydrophobic residues at defined position. The leucine residues in the conserved positions mainly function through hydrophobic interactions to stabilize the solenoid assembly, and can sometimes be substituted by phenylalanine and other hydrophobic amino acids (Bell et al., 2003). It is unexpected that the conserved single amino acid substitution L412F reduced activity. One possible explanation is that this SNP interrupts glycosylation at N413, which has been shown to have N-acetylglucosamines attached in the crystal structure of TLR3 ECD (Choe et al., 2005; Bell et al., 2005). The significance of this SNP deserves more studies because this SNP occurs at a high frequency in human population (22.5%).

N284 is at the end of the type VI β-turn region (consensus motif position N^{10}) of LRR10 (Choe et al., 2005). The side chain of this asparagine forms extensive hydrogen

bonds to amide and carbonyl groups of the same and neighboring LRR motifs. Asparagine at this position is conserved among all organisms we examined (Homo, Pan, Canis, Bos, Rattus, Mus, Takifugu, Tetraodon, Paralicht, Carassius, Branchios, Oncorhync, Danio, and Ictalurus). The N284I mutation disrupts hydrogen bonds and destabilizes the structure of TLR3. It will be interesting to examine persons with the N284I mutation, as this TLR3 SNP does not have any activity *in vitro*.

More studies are needed to understand the structural and mechanistic changes that are caused by the SNPs. It will also be interesting to find the phenotypes associated with the SNPs. People with different TLR SNPs may have different susceptibility to some inflammatory diseases (Lazarus et al., 2002; Tantisira et al., 2004; Hammann et al., 2005). Patients with different genetic backgrounds may need different treatment. So-called molecular medicine, including structure-based drug discovery and gene therapy, will have important roles in the future for the treatment of human diseases. The characterization of TLR3 and its mutations, the structural determination of TLR3 ECD and other biomedically-critical macromolecules are important steps toward understanding how the human immune system attempts to combat complex human diseases.

LRR-containing proteins and pathogen recognition

When comparing the receptors of the innate immune system of plants, insects, and animals, one remarkable conserved feature is that the majority of the receptors contain multiple Leucine-rich repeat (LRR) motifs (Kufer et al., 2005). The ECD of

human TLRs have a large number of LRRs, ranging from 19 to 25 LRR in each protein. The majority of plant disease resistance (R) proteins encode nucleotide-binding site (NBS)-LRR proteins, which are similar to the mammalian nucleotide-binding oligomerization domain (NOD)-LRR family (Inohara and Nunez, 2003). Mammalian NOD proteins are cytoplasmic sensors for PAMPs and regulate inflammatory processes and apoptosis (Eckmann and Karin, 2005). Polymorphisms of NOD2 have been associated with increased susceptibility to several human inflammatory diseases, including Crohn's disease (Eckmann and Karin, 2005).

One interesting question is why evolution selects LRRs to detect pathogens. Unlike the adaptive immune system, which can produce nearly unlimited types of T- and B-cell receptors to recognize unlimited numbers of antigens, the innate immune system only has a limited number of germline-encoded receptors to sense nearly unlimited kinds of pathogens (Takeda et al., 2003). To make the job even more difficult, pathogenic microbes are evolving very quickly, but the germline-encoded receptors of multi-cellular hosts evolve much slower. This seemingly impossible task for the innate immune system is aided by two approaches. The first approach is that only the common structural patterns of pathogens (PAMPs) are recognized by the PAMP receptors (PRRs) (Werling and Jungi, 2003). There are limited PAMPs, mainly acylated lipoproteins, LPS, flagellin, dsRNA, and unmethylated CpG DNA. These PAMPs are usually essential for the pathogenic microbes' survival, or important for their virulence. Mutations that change the PAMPs are not beneficial for the pathogens themselves in most cases. Thus, the

conserved and limited kinds of PAMPs make the receptors' job easier, only a limited number of PRRs are rquired.

The TLR3 ECD is composed of 23 LRRs, with each LRR harboring ~ 24 amino acids with the consensus motif $xL^{2}xxL^{5}xL^{7}xxN^{10}xL^{12}xxL^{15}xxxxF^{20}xxL^{23}x$ (Choe et al., 2005). Each LRR forms a loop with the hydrophobic residues at the conserved positions facing inward to make a hydrophobic core (Bell et al., 2003). Neighboring LRR loops stack together to form a solenoid-like structure. The hydrophilic residues at non-conserved positions are facing outward and are exposed to solution, either at the convex face or at the concave face. The horseshoe-shaped solenoid structure provides an architecture that has the maximum possible surface area for a given sequence length. The large surface provides a large space for protein-protein interactions and ligand binding. Mutations at the non-conserved positions can change the specificity/affinity for a ligand, but do not change the architecture or folding. This ability to acquire changes in the protein surface without disrupting protein folding provides a way for the hosts to create new proteins to recognize newly evolved microbes with minimal cost.

The genomes of different organisms were compared to trace the origin and evolution of TLRs. There is only one TLR1 gene in the Tetraodon fish genome, but two tandem TLR1 paralogs in chicken and opossum (Roach et al., 2005). There are three tandems TLR1 genes in *Xenopus*, they correspond to TLR10, TLR1, and TLR6 in human and mouse. It is possible that TLR10, TLR1, and TLR6 in the human genome were produced by gene duplication, and then the genes diversified by mutations to have different functions. This model of copy and diversification of TLR proteins is further

supported by comparing the sequences of human TLRs, where TLR1, TLR2, TLR6, and TLR10 are more similar to each than to the other TLRs (Roach et al., 2005). The PAMPs of this TLR2 subfamily are bacterial lipoproteins, but each TLR has a different specificity for different lipoproteins. Triacylated lipoproteins are recognized by the TLR2/TLR1 heterodimer, whereas diacylated lipoproteins are recognized by the TLR2/TLR6 heterodimer (Wetzler, 2003). It is possible that at one time, only one TLR2 gene existed and that this ancestral TLR2 formed homodimers upon detection of certain bacterial lipoproteins. Over evolution, the genes encoding lipoproteins of some bacteria mutated, and the hosts adapted to evolve multiple TLR2 subfamily members, including TLR1 and TLR6, to better sense the various bacterial lipoproteins.

It should be noted that the gene duplications and diversifications must have occurred at a very early time. Vertebrate TLRs can be classified into six major subfamilies: (1) TLR1, 2, 6, 10 and 14; (2) TLR3; (3) TLR 4; (4) TLR5; (5) TLR7, 8 and 9; and (6) TLR11-13, 21-23 (Roach et al., 2005). Selective pressure to recognize the major PAMPs (lipoproteins, LPS, flagellin, nucleic acids of bacteria and virus) maintained the six TLR subfamilies in all vertebrates, with few exceptions. The PAMPs of the TLR11 family are not clear, but it has been shown that mouse TLR11 responds to uropathogenic bacteria (Zhang et al., 2004). Human have only a *pseudo*-gene of TLR11, which may explain why human beings are highly susceptible to infections of the urogenital system (Zhang et al., 2004).

It is a surprise that the irreducibly complex immune system of human beings uses the same simple detectors that have been used for billions of years by fish, flies, and

plants. But simple is the best. The special folding of LRRs provides the host with the ability to use only one type of weapon to detect and respond to unlimited and constantly changing pathogenic microbes. Determining the structure of the PAMP receptors is critical for us to understand how our innate immune system works.

Structural studies of other biological macromolecules

2D crystallization and single particle 3D reconstruction has been applied successfully to several other important biological macromolecules. Circadian clock protein KaiC from the cyanobacterium *Synechococcus elongatus* forms 2D crystals of hexameric form on a lipid monolayer with Ni-NTA lipid (Figure 5-3). TEM single particle 3D reconstruction revealed that each KaiC monomer has two similar domains, and each domain forms an independent hexameric ring with a narrow linker between the two domains.

KaiA has been shown to interact with KaiC by the yeast two-hybrid assay and each KaiC has two KaiA-binding domains (Taniguchi et al., 2001). KaiA forms dimers in solution (Vakonakis et al., 2004) and KaiA can be co-purified with 6His-tagged KaiC by Ni-NTA chromatography (Sun, unpublished result). However, the ratio of KaiA to KaiC cannot be deduced from this biochemical experiment.

To determine how KaiA interacts with KaiC, a KaiA-KaiC complex was co-crystallized by the lipid monolayer method. From electron crystallography analysis, it is clear that two KaiC hexamers bind one KaiA dimer (Figure 5-3). The binding position is likely at the C-terminus of KaiC. This result is consistent with a recent proposal that the

Kai proteins form a rotary clock with the KaiA dimer rotating inside the central channel of the KaiC hexamer and causing the phosphorylation state of each KaiC monomer to oscillate (Wang, 2005). Two kinds of complexes are proposed in that paper: one KaiA dimer binds the C-terminus of one KaiC hexamer, or two KaiA dimmers are located in the middle of two KaiC hexamers (Figure 5-4). From our crystal data, it is more likely that one KaiA dimer is sandwiched in the middle of two KaiC hexamers. Instead of KaiA rotating inside KaiC, it is more likely that KaiC rotates around the axis of KaiA. The exact mechanisms how the rotary clock of Kai proteins controls the circadian behavior of cyanobacteria needs more studies in light of the new structural information.

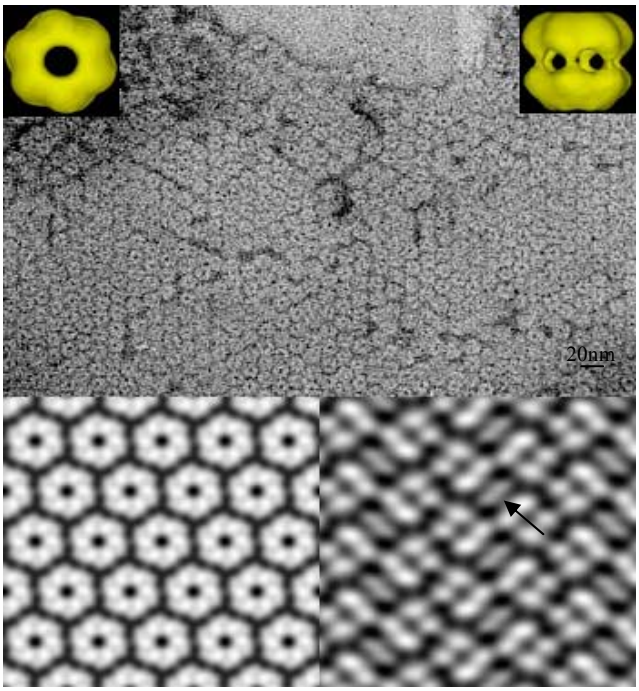

Figure 5-3. Electron crystallography of KaiC and KaiC-KaiA complex. Top: Electron micrograph of a negatively stained 2D crystal of KaiC. The inserts show the single particle 3D reconstruction of KaiC viewed face-on and side-on. Bottom: Fourier projection maps of KaiC (left) and the KaiC-KaiA complex (right). Arrow indicates one KaiA dimer between two KaiC hexamers.

The many successful examples in our lab have proved that TEM is a very powerful tool to study protein structure and protein-protein interaction. I am optimistic that TEM will have a greater impact on structural biology in the future.

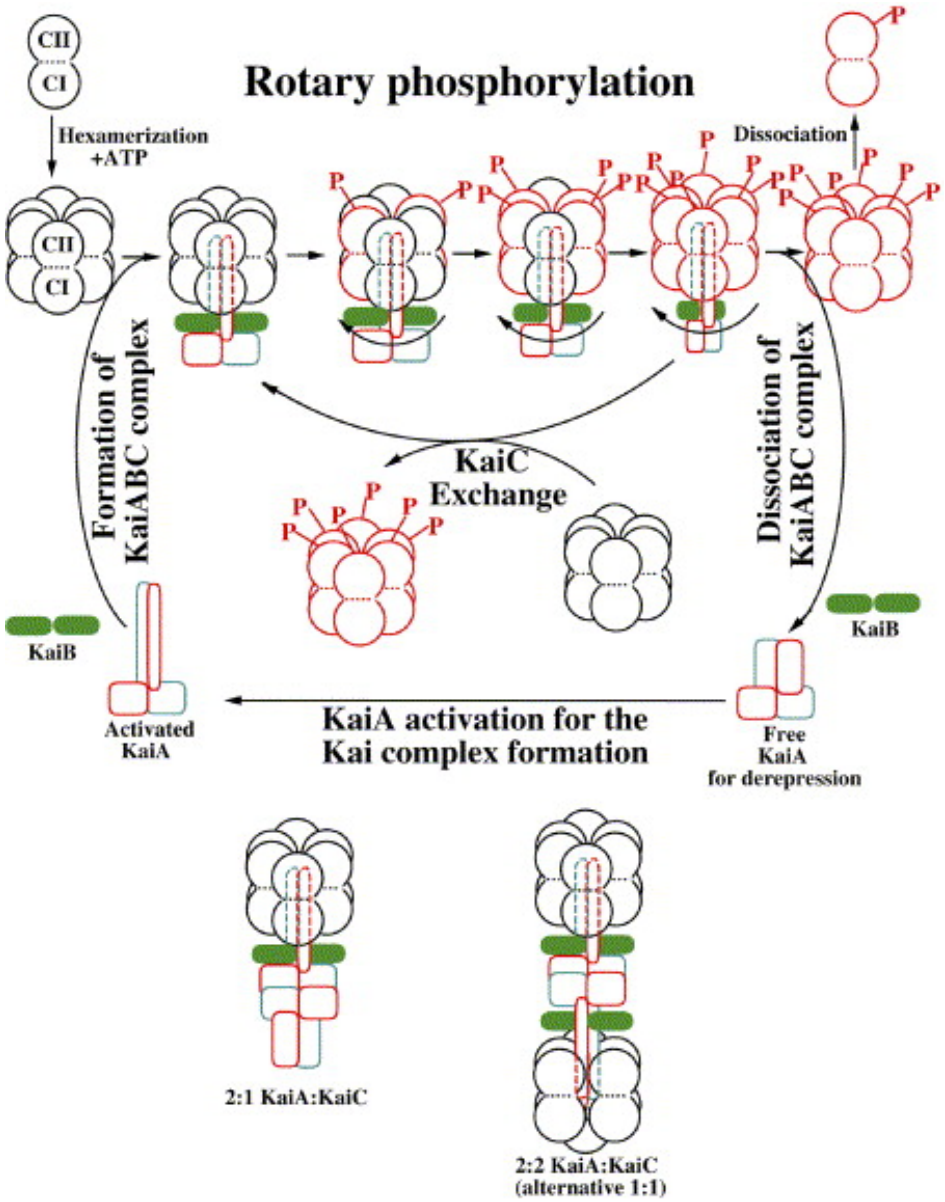

Figure 5-4. Model of the Kai proteins rotary clock. The KaiA dimer is thought to rotate inside the central channel of the KaiC hexamer (Wang, 2005).

REFERENCES

Abrahams, J.P., Leslie, A.G., Lutter. R., and Walker, J.E. (1994). Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. Nature *370*, 621-628.

Akira, S., Takeda, K., and Kaisho, T. (2003). Toll-like receptors: critical proteins linking innate and acquired immunity. Nat. Immunol. *2*, 675-680.

Alland, D., Kramnik, I., Wesbrod, T.R., Otsubo, L., Cerny, R., Miller, L.P., Jacobs, W.R., and Bloom, B.R. (1998). Identification of differentially expressed mRNA in prokaryotic organisms by customized amplification libraries (DECAL): the effect of isoniazid on gene expression in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. USA. *95*, 13227-13232.

Alland, D., Steyn, A.J., Weisbrod, T., Aldrich, K., and Jacobs, W.R. (2000). Characterization of the *Mycobacterium tuberculosis* iniBAC promoter, a promoter that responds to cell wall biosynthesis inhibition. J. Bacteriol. *182*, 1802-1811.

Amos, L.A. (2000). Focusing-in on microtubules. Curr. Opin. Struct. Biol. *10*, 236-241.

Amos, L.A, Henderson, R., and Unwin, P.N.T. (1982). Three-dimersional structure determination by electron microscopy of two-dimensional crystals. Prog. Biophys. Mol. Biol. *39*, 183-231.

Anderson, K.V., Jurgens, G., and Nusslein-Volhard, C. (1985). Establishment of dorsal-ventral polarity in the *Drosophila* embryo: genetic studies on the role of the Toll gene product. Cell *42*, 779-789.

Baker, T.S., Olson, N.H., and Fuller, S.D. (1999). Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. Microbiol. Mol. Biol. Reviews. *63*, 862-922.

Baumeister, W., and Steven, A.C. (2000). Macromolecular electron microscopy in the era of structural genomics. Trends. BioChem. Sci. *25*, 624-631.

Bell, J.K., Botos, I., Hall, P.R., Askins, J., Shiloach, J., Segal, D.M., and Davies, D.R. (2005). The molecular structure of the Toll-like receptor 3 ligand-binding domain. Proc. Natl. Acad. Sci. USA. *102*, 10976-10980.

Bell, J.K., Mullen, G.E., Leifer, C.A., Mazzoni, A., Davies, D.R., and Segal, D.M. (2003). Leucine-rich repeats and pathogen recognition in Toll-like receptors. Trends Immunol. *24*, 528-533.

Bhardwaj, K., Guarino, L., and Kao, C. C. (2004). The severe acute respiratory syndrome coronavirus Nsp15 protein is an endoribonuclease that prefers manganese as cofactor. J. Virol. *78*, 12218–12224

Booth, C., Jiang, W., Baker, M., Zhou, Z., Ludtke, S., and Chiu, W. (2004). A 9 Å single particle reconstruction from CCD captured imageds on a 200 kV electron cryomicroscope. J. Struct. Biol. *147*, 116-127.

Bowie, A.G., and Haga, I.R. (2005). The role of Toll-like receptors in the host response to viruses. Mol. Immunol. *42*, 859-867.

Boyer, P.D. (2002). A research journey with ATP Synthase. J. Biol. Chem. *277*, 39045-39061.

Carpousis, A.J. (2002). The *Escherichia coli* RNA degradosome: structure, function and relationship in other ribonucleolytic multienzyme complexes. Biochem. Soc. Trans. *30*, 150-155.

Castro, K.G. (1995). Nosocomial tuberculosis: new progress in control and prevention. Clin. Infect. Dis. *21*, 489-505.

Chen, N., Walsh, M.a., Liu, Y., Parker, R., and song, H. (2005). Crystal structures of human DcpS in ligand-free and m7GDP-bound forms suggest a dynamic mechanism for scavenger mRNA decapping. J. Mol. Biol. *347*, 707-718.

Chang, W., and Kornberg, R.D. 2000. Electron crystal structure of transcription factor and DNA repair complex, core TFIIH. Cell *102*:609-613.

Chiu, W., Avila-Sakar, A.J. and Schmid, M.F. (1997). Electron crystallography of macromolecular periodic arrays on phospholipid monolayers. Adv. Biophys. *34*, 161-172.

Chiu, W., Baker, M.L., Jiang, W., Dougherty, M., and Schmid, M.F. (2005). Electron cryomicroscopy of biological machines at subnamometer resolution. Structure. *13*, 363-372.

Choe, J., Kelker, M.S., and Wilson, I.A. (2005). Crystal structure of human Toll-like receptor 3 (TLR3) ectodomain. Science *309*, 581-585.

Clemons, W.M., Menetre, J.F., Akey, C.W., and Rapoport, T.A. (2004). Structural insight into the protein translocation channel. Curr. Opin. Struct. Biol. *14*, 390-396.

Colangeli, R., Helb, D., Sridharan, S., Sun, J., Vama-basil, M., Hazbon, M.H., Harbacheuski, R., Megjugorac, N.J., Jacobs, W.R., Holzenburg, A., Sacchettini, J.C., and Alland, D. (2005). The *Mycobacterium tuberculosis* iniA gene is essential for activity of an efflux pump that confers drug tolerance to both isoniazid and ethambutol. Mol. Microbiol. *55*, 1829-1840.

Cook, D.N., Pisetsky, D.S., and Schwartz, D.A. (2004). Toll-like receptors in the pathogenesis of human disease. Nat. Immunol. *5*, 975-979.

Dang, T.X., Milligan R.A., Tweten, R.K., and Wilson-Kubalek, E.M. (2005). Helical crystallization on nickel-lipid nanotubes: perfringolysin O as a model protein. J. Struct. Biol. *152*, 129-139.

Da Silva C. J., and Ulevitch, R.J. (2002). MD-2 and TLR4 N-linked glycosylations are important for a functional lipopolysaccharide receptor. J. Biol. Chem. *277*, 1845-1854.

De Bouteiller, O., Merck, E., Hasan, U.A., Hubac, S., Benguigui, B., Trinchieri, G., Bates, E.E., and Caux, C. (2005). Recognitioin of double stranded RNA by human Toll like receptor 3 and downstream receptor signaling requires multimerisation and an adidic pH. J. Biol. Chem. *280*, 38133-38145.

Deisenhofer, J., Epp, O., Miki, K., and Michel, H. (1985). Structure of the protein subunits in the photosynthetic reaction center of *Rhodospeudomonas viridis* at 3 Å resolution. Nature *318*, 618-624.

DeRosier, D., and Klug, A. (1968). Reconstruction of 3-dimensional structures from electron micrographs. Nature *217*, 130-134.

Diamond, M.S., Shrestha, B., Mehlhop, E., Sitati, E., and Engle, M. (2003). Innate and adaptive immune responses determine protection against disseminated infection by West Nile encephalitis virus. Viral Immunol. *16*, 259-278.

Doi, K. (2005). Current status and future potential of computer-aided diagnosis in medical imaging. Br. J. Radiol. *78*, S3-19.

Eckmann, L., and Karin, M. (2005). NOD2 and Crohn's disease: loss or gain of function? Immunity *22*, 661-667.

Frank, J. (1975). Averaging of low exposure electron micrographs of nonperiodic objects. Ultramicroscopy *1*, 159-162.

Frank, J. (1996). Three-dimensional electron microscopy of macromolecular assemblies (New York: Academic Press).

Frank, J. (2001). Cryo-electron microscopy as an investigative tool: the ribosome as an example. Bioessays. *23*, 725-732.

Frank, J. (2002). Single-particle imaging of macromolecules by cryo-electron microscopy. Annu. Rev. Biophys. Biomol. Struct. *31*, 303-319.

Fujiyoshi, Y. (1998). The structural study of membrane proteins by electron crystallography. Adv. Biophys. *35*, 25-80.

Gabashvili, I.S., Agrawal, R.K., Spahn, C., grassucci, R., Svergun, D., Frank, J., and Penczek, P. (2000). Solution structure of the *E. coli* 70s ribosome at 11.5Å resolution. Cell *100*, 537-549.

Gonen, T., Gheng, Y., Sliz, P., Hiroaki, Y., Fujiyoshi, Y., Harrison, S.C., and Walz, T. (2005). Lipid-protein interations in double-layered two-dimensional AQP0 crystals. Nature *438*, 633-638.

Guarino, L., Bhardwaj, K., Dong, W., Sun, J., Holzenburg, A., and Kao, C. (2005). Mutational analysis of the SARS virus Nsp15 endoribonuclease: identification of residues affecting hexamer formation. J. Mol. Biol. 353, 1106-1117.

Guenebaut, V., Vincentelli, R., Mills, D., Weiss, H., and Leonard, K.R. (1997). Three-dimensional structure of NADH-dehydrogenase from *Neurospora crassa* by electron microscopy and conical tilt reconstruction. J. Mol. Biol. *265*, 409-418.

Hammann, L., Gomma, A., Schroder, N.W., Stamme, C., glaeser, C., Schulz, S., Gross, M., Anker, S.D., Fox, K., and Schummann, R.R. (2005). A frequent Toll-like receptor (TLR)-2 polymorphism is a risk factor for coronary restenosis. J. Mol. Med. *83*, 475-485.

Harris, J.R., and Scheffler, D. (2002). Routine preparation of air-dried negatively stained and unstained specimens on holey carbon support films: a review of application. Micron *33*, 461-480.

Hemmi, H., and Akira, S. (2005). TLR signaling and the function of dendritic cells. Chem. Immunol. Allergy. *86*, 120-135.

Henderson R., Baldwin J.M., Ceska T.A., Zemlin F., Beckmann E. and Downing K.H. 1990. Model for the structure of bacteriorhodopsin based on high-resolution electron cryomicroscopy. J.Mol.Biol. *213*, 899-929.

Henikoff, S., and Henikoff., J.G. (1992). Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA. *89*, 10915-10919.

Hidari, K.I., Horie, N., Murata, T., Miyamoto, D., Suzuki, T., Usui, T., and Suzuki, Y. (2005). Purification and characterization of a soluble recombinant human ST66Gal I functionally expressed in *Escherichia coli*. Glycoconj. J. *22*, 1-11.

Hoppe, W., Gassmann, J., Hunsmann, N., Schramm, H.J., and Sturm, M. (1974). Three-dimensional reconstruction of individual negatively stained fatty-acid synthetase molecules from tilt series in the electron microscopy. Hopp-Seyler Z. Physiol. Chem. *355*, 1483-1487.

Hovmoller, S. (1992). CRISP –crystallographic image processing on a personal computer. Ultramicroscopy *41*, 121-136.

Hunt, C., Koepke, J., Lange, C., Rossmanith, T., and Michel, H. (2000). Structure at 2.3 Å resolution of the cytochrome bc1 complex from the yeast *Saccharomyces cerevisiae* co-crystallized with an antibody Fv fragment. Structure *8*, 669-684.

Inohara, N., and Nunez, G. (2005). NODs: intracellular proteins involved in inflammation and apoptosis. Nat. Rev. Immunol. *3*, 371-382.

Inohara, N., Ogura, Y., and Nunez, G. (2002). Nods: a family of cytosolic proteins that regulate the host response to pathogens. Curr. Opin. Microbiol. *5*, 76-80.

Isaacs, A., Klemperer, H.G., and Hitchcock, G. (1961). Studies on the mechanism of action of interferon. Virology *13*, 191-199.

Ishihara, S., Rumi, M.a., Kadowaki, Y., Ortega-Cava, C.F., Yuki, T., Yoshino, N., Miyaoka, Y., Kazumori, H., Ishimura, N., Amano, Y., and Kinoshita, Y. (2004). Essential role of MD-2 in TLR4-dependent signaling during *Helicobacter pyloriassociated* gastritis. J. Immunol. *173*, 1406-1415.

Ivanov, K.A., Hertzig, T., Rozanov, M., Bayer, S., Thiel, V., Gorbalenya, A.E., and Ziebuhr, J. (2004). Major genetic marker of nidoviruses encodes a replicative endoribonuclease. Proc. Natl. Acad. Sci. USA. *101*, 12694-12699.

Iwasaki, A., and Medzhitov, R. (2004). Toll-like receptor control of the adaptive immune response. Nat. Immunol. *5*, 987-995.

Janssens, S., and Beyaert, R. (2003). Role of Toll-like receptors in pathogen recognition. Clin. Microbiol. Rev. *16*, 637-646.

Klebe, G. (2000). Recent developments in structure-based drug design. J. Mol. Med. *78*, 269-281.

Kufer, T.A., Fritz, J.H., and Philpott, D.J. (2005). NACHT-LRR proteins (NLRs) in bacterial infection and immunity. Trends. Microbiol. *13*, 381-388.

Kuhlbrandt, W., Wang, D.N., and Fujiyoshi, Y. (1994). Atomic model of plant light-harvesting complex by electron crystallography. Nature *367*, 614-621.

Lazarus, R., Vercelli, D., Palmer, L.J., Klimecki, W.J., Silverman, E.K., Richter, B., Riva, A., Ramoni, M., Martinez, F.D., Weiss, S.T., and Kwiatkowski, D.J. (2002). Single nucleotide polymorphisms in innate immunity genes: abundant variation and potential role in complex human disease. Immunol. Rev. *190*, 9-25.

Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.M., and Hoffmann, J.A. (1996). The dorsoventral regulatory gene cassette spatzle/Toll/cactus control the potent antifungal response in *Drosophila* adults. Cell *86*, 973-983.

Levitt, M. (1992). Accurate modeling of protein conformation by automatic segmant matching . J. Mol. Biol. *226*, 507-533.

Levitt, M., Hirshberg, M., Sharon, R., and Daggett, V. (1995). Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. Computer Physics. Comm. *91*, 215-231.

Livak, F., Burtrum, D.B., Rowen, L., Schatz, D.G., and Petrie, H.T. (2000). Genetic modulation of T cell receptor gene segment usage during somatic recombination. J. Exp. Med. *192*, 1191-1196.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. *128*, 82-97.

Ludtke, S.J., Chen, D.H., Song, J.L., Chuang, D.T., and Chiu, W. (2004). Seeing GroEL at 6 Å resolution by single particle electron cryomicroscopy. Structure *12*, 1129-1136.

Lund, O., Frimand, K., Gorodkin, J., Bohr, H., Bohr, J., Hansen, J., and Brunak, S. (1997). Protein distance constraints predicted by neural networks and probability density functions. Protein Engineering *10*, 1241-1248.

Matsumoto, M., Funami, K., Oshiumi, H., and Tsukasa, S. (2004). Toll-like receptor 3: a link between Toll-like receptor, interferon and viruses. Microbiol. Immunol. *48*, 147-154.

Matsumoto, M., Funami, K., Tanabe, M., Oshiumi, H., Shingai, M., Seto, Y., Yamamoto, A., and Seya, T. (2003). Subcellular localization of Toll-like receptor 3 in human derdritic cells. J. Jmmunol. *171*, 3154-3162.

Medzhitov, R., Preston-Hurlburt, P., and Janeway, C.A. (1997). A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. Nature *388*, 323-324.

Misell, D.L., (1978). Image analysis and interpretation. (New York: Oxford).

Mitsuoka, K., Murata, K., Kimura, Y., Namba, K., and Fujiyoshi, Y. (1997). Examination of the LeafScan 45, a line-illuminating micro-densitometer, for its use in electron crystallography. Ultramicroscopy *68*, 109-121.

Miyake, K. (2004). Endotoxin recognition molecules MD-2 and Toll-like receptor 4 as potential targets for therapeutic intervention of endotoxin shock. Curr. Drug Targets Inflamm. Allergy. *3*, 291-297.

Mori, T., Saveliev, S.V., Xu, Y., Stafford, W.F., Cox, M.M., Inman, R.B., and Johnson, C.H. (2002). Circadian clock protein KaiC forms ATP-dependent hexameric rings and binds DNA. Proc. Natl., Acad. Sci. USA. *99*, 17203-17208.

Nishiya, T., Kajita, E., Miwa, S., and Defranco, A.L. (2005). TLR3 and TLR7 are targeted to the same intracellular compartment by distinct regulatory elements. J. Biol. Chem. *280*, 37107-37117.

Nogales, E., Wolf, S.G., and Downing, K.H. (1998). Structure of the αβ tubulin dimmer by electron crystallography. Nature *391*, 199-203.

O'Neill, L.A. (2003). Therapeutic targeting of Toll-like receptors for inflammatory and infectious diseases. Curr. Opin. Pharmacol. *3*, 396-403.

Padron, R., Almo, L., Murgich, J., and Craig, R. (1998). Towards an atomic model of the thick filaments of muscle. J. Mol. Biol. *275*, 35-41.

Palsson-McDermott, E.M., and O'Neill, L.A. (2004). Signal transduction by the lipopolysaccharide receptor, Toll-like receptor-4. Immunology. *113*, 153-162.

Parodi, A.J. (2000). Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. Biochem. J. *348*, 1-13.

Pasare, C., and Medzhitov, R. (2003). Toll-like receptors: balancing host resistance with immune tolerance. Curr. Opin. Immunol. *15*, 677-682.

Pasare, C., and Medzhitov, R. (2004). Toll-like receptors: linking innate and adaptive immunity. Microbes. Infect. *6*, 1382-1387.

Penczek, P.A., grassucci, R.A., and Frank, J. (1994). The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. Ultramicroscopy. *53*, 251-270.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera – A visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605-1612.

Pinner, R.W., Teutsch, S.M., Simonsen, L., Klug, L.A., Graber, J.M., Clarke, M.J., and Berkelman, R.L. (1996). Trends in infectious diseases mortality in the United States. J. Am. Med. Assoc. *275*, 189-193.

Radermacher, M., Wagerknecht, T., Verschoor, A., and Frank, J. (1987). Three-dimersional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit. J. Microsc. *146*, 113-136.

Ranson, N.A., Farr, G.W., Roseman, A.M., Gower, B., Fenton, W.A., Horwich, A.L., and Saibil, H.R. (2001). ATP-bound states of GroEL captured by cryo-electron microscopy. Cell. *107*, 869-879.

Rifkin, I.R., Leadbetter, E.A., Busconi, L, viglianti, G., and Marshak-Rothstein, A. (2005). Toll-like receptors, endogenous ligands, and systemic autoimmune disease. Immunol. Rev. *204*, 27-42.

Roach, J.C., Glusman, G., Rowen, L., Kaur, A., Purcell, M.K., Smith, K.D., Hood, LE., and Aderem, A. (2005). The evolution of vertebrate Toll-like receptors. Proc. Natl. Acad. Sci. USA. *102*, 9577-9582.

Salio, M., and Cerundolo, V. (2005). Viral immunity: cross-priming with the help of TLR3. Curr. Biol. *15*, R336-339.

Sali, A., Glaeser, R., Earnest, T., and Baumeister, W. (2003). From words to literature in structural proteomics. Nature *422*, 216-225.

Schulz, O., Diebold, S.S., Chen, M., Naslund, T.I., Nolte, M.A., Alexopoulou, L., Azuma, Y.T., Flavell, R.A., Liljestrom, P., and Reis e Sousa, C. (2001). Toll-like receptor 3 promotes cross-priming to virus-infected cells. Nature *433*, 887-892.

Schroder, M., and Bowie, A.G. (2005). TLR3 in antiviral immunity: key player or bystander? Trends. Immunol. *26*, 462-468.

Schroder, N., and Schumann, D. (2005). Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. Infectious Diseases. *5*, 156-164.

Smith, C.V., and Sacchettini, J.C. (2003). *Mycobacterium tuberculosis*: a model system for structural genomics. Curr. Opin. Struct. Biol. *13*, 658-664.

Smyth, D.R., Mrozkiewicz, M.K., McGrath, W.J., Listwan, P., and Kobe, B. (2003). Crystal structures of fusion proteins with large-affinity tags. Protein Sci. *12*, 1313-1322.

Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J., and Gorbalenya. A.E. (2003). Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavius group 2 lineage. J. Mol. Biol. *331*, 991-1004.

Stark, C.J., and Atreya, C.D. (2005). Molecular advances in the cell biology of SARS-CoV and current disease prevention strategies. Virology J. *2*, 35-43.

Stock, D., Leslie, A., and Walker, J.E. (1999). Molecular architecture of the rotary motor in ATP synthase. Science *286*, 1700-1705.

Sun, E., and Cohen, F.E. (1993). Computer-assisted drug discovery—a review. Gene. *137*, 127-32.

Sun, J., Savva., C.G., Deaton., J., Kaback., H.R., Svrakic., M., Young. R., and Holzenburg., A. (2005). Asymmetric binding of membrane proteins of GroEL. Arch. Biochem. Biophys. *434*, 352-357.

Takeda, K., and Akira, S. (2005). Toll-like receptors in innate immunity. Int. Immunol. *17*, 1-14.

Takeda, K., Kaisho, T., and Akira, S. (2003). Toll-like receptors. Annu. Rev. Immunol. *21*, 335-376.

Taniguchi, Y., Yamaguchi, A., Hijikata, A., Iwasaki, H., Kamagata, K., Ishiura, M., Go., M., and Kondo, T. (2001). Two KaiA-binding domains of cyanobacterial circadian clock protein KaiC. FEBS Lett. *496*, 86-90.

Tantisira, K., Klimecki, W.T., Lazarus, R., Palmer, L.J., raby, B.A., Kwiatkowski, D.J., Silverman, E., Vercelli, D., Martinez, F.D., Weiss, S.T. (2004). Toll-like receptor 6 gene (TLR6): single-nucleotide polymorphism frequencies and preliminary association with the diagnosis of asthma. Genes Immun. *6*, 343-346.

Tomasselli, A.G., and Heinrikson, R.L. (2000). Targeting the HIV-protease in AIDS therapy: acurrent clinical perspective. Biochim. Biophys. Acta. *1477*, 189-214.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature *302*, 575-581.

Tozser, J., and Oroszlan, S. (2003). Proteolytic events of HIV-1 replication as targets for therapeutic intervention. Curr. Pharm. Des. *9*, 1803-1815.

Tsang, K., and Seto, W.H. (2004). Severe acute respiratory syndrom: scientific and anecdotal evidence for drug treatment. Curr. Opin. Investig. Drugs. *5*, 179-185.

Yamamoto, M., Takeda, K., and Akira, S. (2004). TIR domin-containing adaptors define the specificity of TLR signaling. Mol. Immunol. *40*, 861-868.

Unwin, N. (1998). The nicotinic acetylcholine receptor of the Torpedo electric ray. J. Struct. Biol. *121*, 181-190.

Unwin, N. (2005). Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. J. Mol. Biol. *346*, 967-989.

Ureta-Vidal, A., Ettwiller., L., and Birney, E. (2003). Comparative genomics: genome-wide analysis in metazoan eukaryotes. Nat. Rev. Genet. *4*, 251-262.

Vakonakis, I., and LiWang, A.C. (2004). Structure of the C-terminal domain of the clock protein KaiA in complex with a KaiC-derived peptide: implications for KaiC regulation. Proc. Natl. Acad. Sci. USA. *101*, 10925-10930.

Vakonakis, I., Sun., J., Wu, T., Holzenburg, A., Golden, SS. and LiWang, AC. (2004). NMR structure of the Kaic-interacting C-terminal domain of KaiA, a circadian clock protein: implications for KaiA-KaiC interaction. Proc. Natl. Acad. Sci. USA. *101*, 1479-1484.

Valentine, R.C., Shapiro, B.M., and Stadtman E.R. (1968). Regulation of glutamine synthetase. XII. electron microscopy of the enzyme from *Escherichia coli*. Biochemistry. *7*, 2143-2152.

van Heel, M. (1987). Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction. Ultramicroscopy *21*, 111-124.

van Heel, M., Gatadeen, R., Orlova, E.V., Finn, R., Pape, R., Cohen, D., Stark, H., Schmidt, R., Schatz, M., and Patwardhan, A. (2000). Single-particle electron cryo-microscsopy: towards atomic resolution. Q. Rev. Biophys. *33*, 307-369.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. J. Struct. Biol. *116*, 17-24.

Visintin, A., Halmen, K.A., Latz, E., Monks, G., and golenbock, D.T. (2005). Pharmacological inhibition of endotoxin responses is achieved by targeting the TLR4 coreceptor, MD-2. J. Immunol. *175*, 6465-6472.

Wade, M.M, and Zhang, Y. (2004). Mechanisms of drug resistance in *Mycobacterium tuberculosis*. Front. Biosci. *9*, 975-994.

Walz, T., and Grigorieff, N. (1998). Electron crystallography of two-dimensional crystals of membrane proteins. J. Struct. Biol. *121*, 142-161.

Wang, J., (2005). Recent cyanobacterial Kai protein structures suggest a rotary clock. Structure. *13*, 735-741.

Weber, Q.N., More, M.A., and Gay, N.J. (2004). Four N-linked glycosylation sites in human TLR2 cooperate to direct efficient biosynthesis and secretion. J. Biol. Chem. *279*, 34589-34594.

Wetzler, L.M. (2003). The role of Toll-like receptor 2 in microbial disease and immunity. Vaccine. *21*, S55-60.

Wenzel, R.P., Bearman, G., and Edmond, M.B. (2005). Lessons from Severe Acute Respiratory Syndrom (SARS): implications for infection control. Arch. Med. Res. *36*, 610-616.

Werling, D., and Jungi, T.W., (2003). Toll-like receptors linking innate and adaptive immune response. Vet. Immunol. Immunopathol. *91*, 1-12.

Yonekura, K., Maki-Yonekura, S., and Namba, K. (2005). Building the atomic model for the bacterial flagellar filament by electron cryomicroscopy and image analysis. Structure. *13*, 407-412.

Zhang, D., Zhang, G., Hayden, M.S., Greenblatt, M.B., Bussey, C., Flavell, R.A., and Ghosh, S. (2004). A Toll-like receptor that prevents infection by uropathogenic bacteria. Science *303*, 1522-1526.

Zhang, Y. (2005). The magic bullets and tuberculosis drug targets. Annu. Rev. Pharmacol. Toxicol. *45*, 529-564.

VITA

Name: Jingchuan Sun

Address: Microscopy & Imaging Center, BSBW 119, Texas A&M University
College Station, Texas 77843-2257

Email Address: jingchuansun@neo.tamu.edu

Education: Ph.D., Microbiology, Texas A&M University, 2005,
M.S., Biophysics, Peking University, Beijing, China, 1990,
B. S., Biology, Peking University, Beijing, China, 1987,

Awards: First Place at Student Research Week 2005, Texas A&M University
2004 Presidential Student Award, Microscopy Society of America

Publications:

Colangeli, R., Helb, D., Sridharan, S., Sun, J., Vama-basil, M., Hazbon, M.H., Harbacheuski, R., Megjugorac, N.J., Jacobs, W.R., Holzenburg, A., Sacchettini, J.C., and Alland, D. (2005). The *Mycobacterium tuberculosis* iniA gene is essential for activity of an efflux pump that confers drug tolerance to both isoniazid and ethambutol. Mol. Microbiol. *55*, 1829-1840.

Deaton, J., Sun, J. Holzenburg, A., Struck, DK, Berry,J., and Young, R. (2004). Functional bacteriorhodopsin is efficiently solubilized and delivered to membranes by the chaperonin GroEL. Proc. Natl. Acad. Sci. USA. *101*, 2281-2286.

Vakonakis, I., Sun., J., Wu, T., Holzenburg, A., Golden, S.S., and LiWang, A.C. (2004). NMR structure of the Kaic-interacting C-terminal domain of KaiA, a circadian clock protein: implications for KaiA-KaiC interaction. Proc. Natl. Acad. Sci. USA. *101*, 1479-1484.

Sun, J., Duffy, K., Xiong. J., Ranjith-Kumar, C.T., Masarapu, H., Lamb, R., Santos, J., Cunningham, M., Holzenburg, A., Sarisky, R.T., Mbow, M.L., and Kao, C. (2006). Structural and functional analyses of the human Toll-like receptor. J. Biol. Chem. (accepted, JBC proofs: M510442200 article 4826).

Sun, J., Savva., C.G., Deaton., J., Kaback., H.R., Svrakic., M., Young. R., and Holzenburg., A. (2005). Asymmetric binding of membrane proteins of GroEL. Arch. Biochem. Biophys. *434*, 352-357.