THE STRUCTURAL STUDY OF PARKINSON'S DISEASE RELATED PROTEIN LRRK2 AND VACCINIA VIRUS PROTEIN A6

By

YUE HAN

Bachelor of Science in Life Science
University of Science and Technology of China
Hefei, Anhui
2007

Master of Science in Biochemistry and Molecular Biology
Oklahoma State University
Stillwater, OK
2013

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2014

THE STRUCTURAL STUDY OF PARKINSON'S DISEASE RELATED PROTEIN LRRK2 AND VACCINIA VIRUS PROTEIN A6

Dissertation Approved:

Dr. Junpeng Deng

Dissertation Adviser

Dr. Robert Matts

Dr. Andrew Mort

Dr. Haobo Jiang

Outside Committee Member

Name: YUE HAN

Date of Degree: MAY, 2014

Title of Study: THE STRUCTURAL STUDY OF PARKINSON'S DISEASE RELATED PROTEIN LRRK2 AND VACCINIA VIRUS PROTEIN A6.

Major Field: Biochemistry and Molecular Biology

Abstract:

LRRK2 is a multi-domain protein expressed widely in organs. The mutations of LRRK2 are related to Parkinson's disease. The Roc domain in LRRK2 is a Ras GTPase, and its GTP hydrolysis activity has been proved to regulate kinase activity of the kinase domain. With the structure of GDP bound Roc domain (inactive state) solved, we intend to determine the structure of GTP analog bound Roc domain (active state). The change between them should reveal the dynamic conformational changes in the ROC GTPase domain, and provide important clues on the mechanisms of GTP hydrolysis and signaling. We engineered Roc based on surface entropy reduction and used GDP, AlF_4^- and GppNp as GTP analogs. Mutants of Roc crystallized but only GDP was found in the binding pocket.

A6 is a vaccinia virus protein is expressed during late translation and is packed in the viron core. The absence of A6 dislocates some viron proteins, disrupts the recruitment of virus membrane and stops virus maturation. There is no homologue structure of A6, so we intend to solve the structure and compare it with other known functional domain structures. In this way, we can expand our understanding of the role A6 plays during vaccinia virus replication. A6 and its homologues are recalcitrant to crystallization, but surface entropy reduction and truncation based on limited proteolysis promoted crystallization. The mutant AC crystal diffracted to 4 Å and truncated mutant C diffracted to 8 Å. Antibodies including Fabs and SdAbs were also used to assist crystallization. But more combination of A6 and antibodies need to be tried.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER I

INTRODUCTION

A. LRRK2

Parkinson's disease (PD) is the second most common neurodegenerative disease affecting six million people around the world. Patients of PD suffer from motor abnormality caused by loss of dopamine. Leucine-rich Repeat Kinase 2 (LRRK2) is a multi-domain protein expressed broadly in organs including brain, heart and liver. The mutations of LRRK2 are mainly found in patients of PD, and a major part of pathological mutations are located in the central part of LRRK2, including a Ras of Complex (ROC) GTPase domain, a kinase domain and a domain in between called C-terminal of Roc (COR) domain. Previous research has found that the Roc domain regulates kinase activity as a binary switch through a GTP/GDP bound cycle. The structure of the GDP-binding Roc (inactivated state) has been solved previously and we intend to determine the structure of ROC in its activated state. The expected results would reveal the dynamic conformational changes in the ROC GTPase domain, providing important clues on the mechanisms of GTP hydrolysis and signaling. We engineered Roc based on surface entropy reduction and used GDP, AlF_4^- and GppNp as GTP analogs. Mutants of Roc crystallized but only GDP was found in binding pocket.

B. A6

Vaccinia viruses undergo a series of morphological changes during replication. A6 is a vaccinia virus protein essential for vaccinia virus assembly and it was proposed to be involved in vaccinia virus morphogenesis by recruiting membrane from the infected cell, as well as other important proteins to virion factories. Although the sequence of A6 is highly conserved throughout vertebrate poxvirus family, there are no known homologs outside the poxvirus family to suggest its function, or any homolog structure inside the family. We intended to solve the structure of A6 by x-ray crystallography to obtain insights into its function. By comparing the structure of A6 with other known functional domain structures, we can expand our understanding of the role A6 plays during vaccinia virus replication. A6 and its homologues are recalcitrant to crystallization, but surface entropy reduction and truncation based on limited proteolysis promoted crystallization. The mutant AC crystal diffracted to 4 Å and truncated mutant C diffracted to 8 Å. Antibodies including Fabs and SdAbs were also used to assist crystallization. But more combinations of A6 and antibodies need to be tried.

CHAPTER II

METHODOLOGY

A. Introduction

Protein x-ray crystallography is widely used to determine the 3-D structures of proteins or protein complexes. It involves several steps including protein expression, purification, crystallization, data collection and processing, and structure determination refinement and validation. In this chapter, I will discuss protein x-ray crystallography by reviewing each step on the basis of my experiments and understanding.

B. Protein Expression and Purification

The first step of protein x-ray crystallography is to prepare milligram scale of a protein with ≥95% purity. The protein can be either purified from a natural source or overexpressed by an expression system such as bacteria, yeast, insect or mammalian cells.

In bacterial expression systems, it is easy and fast to make the constructs and express proteins. But the lack of the eukaryotic post-translational modifications may result in insoluble or unfolded protein when the target protein is eukaryotic.

Yeasts grow comparably fast as bacteria and are also easy to culture. Yeast expression systems contains the eukaryotic post-translation modification. But the chitinaceous cell wall of yeasts is recalcitrant to lysis, so they are less commonly used to produce cytoplasmic proteins. A signal peptide sequence needs to be carefully selected for the expression of secreted proteins by yeasts.

In contrast, insect and mammalian cell expression systems are both time consuming and expensive for construction and protein expression, but they can express eukaryotic proteins with correct post-translational modification or folding.

For experiments carried out in this thesis, I mainly used *Escherichia coli* (*E. coli*) system for overexpression of the targeted proteins.

1. Protein Overexpression from *E. coli*

Throughout our crystallographic studies, *E. coli* BL21 Gold DE3 strains (Strategene) was used for all protein expression. (Thony, Neuheiser et al. 1998, Johnson, Bornemeier et al. 1999, Wagner, Betzenhauser et al. 2006, Betzenhauser, Wagner et al. 2008) A set of expression vectors were made in house based on the backbone of the pET vectors (Qiagen) (Gao, Cao et al. 2009) with the T7 promoter was also utilized. Gene encoding target protein was amplified by polymerase chain reaction (PCR) and then ligated to specific expression vector for overexpression as 6xHis tagged fusions. N-terminal TEV cleavable 6xHis tag fusion, C-terminal permanent 6xHis tag fusion, SUMO fusion and MBP fusion were used in my projects and will be described in following chapters. The plasmids were subsequently transformed into *E. coli* BL21 Gold DE3 strain by electroporation.

BL21 Gold DE3 cell has lambda DE3 lysogen carrying the T7 RNA polymerase under the control of *lac*UV5 promoter. Also, deletion of *lon* protease and mutation on OmpT protease in BL21 Gold DE3 cell prevent expressed target protein from degradation by these proteases. This furthermore strain has a major advantage of being used for direct cloning of protein expression

constructs, as the transformation efficiency is increased to >1x10^8 cfu/ug of pUC 18 DNA, by selecting the Hte phenotype. Moreover, *enA* gene encoding endonuclease I is disabled to avoid plasmid DNA degradation. Therefore, using this strain shortens the experimental time by up to 50%. (Stratagene competent cells manual)

The mechanism of pET expression vectors are as follows: (1) the *lacI* gene encodes the lac repressor; (2) the lac repressor binds to the *lacUV5* promoter and represses binding of T7 RNA polymerase to T7 promotor; (3) the presence of isopropyl β-D-1-thiogalactopyranoside (IPTG) induces the lac repressor to dissociate from the *lacUV5* promoter and allows the bindin of T7 RNA polymerase. Thus, the target gene ligated to pET expression vector would be transcribed for protein expression. (Gao, Cao et al. 2009)

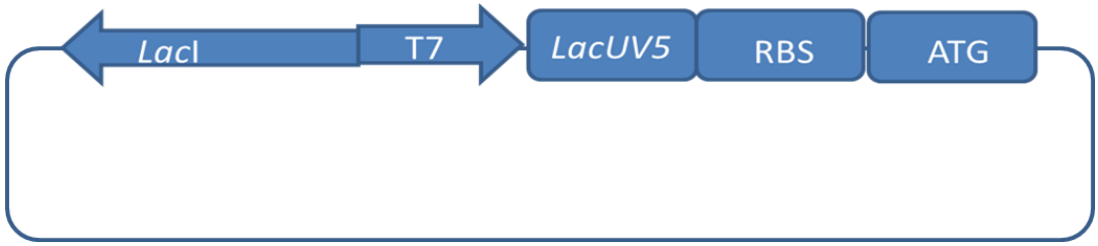


Figure 2-1: Schematic illustration of pET vector.

2. Protein Purification

To facilitate protein purification from *E. coli*, the target proteins are usually fused with tags, such as polyhistidine (6His) (Hengen 1995), Glutathione S-transferase (GST) (Benard and Bokoch 2002) and maltose binding protein (MBP) (Nallamsetty and Waugh 2007). Each tag binds to a specific resin allowing affinity purification.

A common protocol for purification of 6xHis fusion protein is described here. Frozen cell pellets were typically resuspended with buffer A (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 20 mM Imidazole, 10% glycerol). The cells were lysed either by lysozyme followed by sonication (550

Sonic dismembrator from Fisher Scientific), or high-pressure homogenizer (EmulsiFlex-C5). Protease inhibitors were used to prevent protein degradation and DNase was added to digest the bacterial genome DNA in order to reduce viscosity. The cell lysate was fractionated into insoluble and soluble components by high-speed centrifugation (Thermo scientific Sorvall RC 6+ centrifuge) at 15,000g for 0.5 hour at 4℃. The supernatant is subsequently mixed with proper amount of Nickel-Nitrilotriacetic Acid (Ni-NTA) resin (GE Healthcare), typically 1 ml resin for 10 mg of 6xHis tagged protein. The mixture is incubated on a platform rocker at 4℃ for 1 hour to allow efficient binding before transferred into a small gravity flow column. The unbound material is separated from Ni-NTA resin as flow through. 20 column volumes of Buffer A is used to wash off the nonspecifically bound contaminants, and the target protein is eluted with buffer B (20 mM Tris-HCl, pH8.0, 500 mM NaCl, 250 mM Imidazole, 10% glycerol). For crystallization purposes, the artificial 6xHis tag and the flexible linker need to be removed by specific proteases. Ubiquitin-like protein-processing protease *Ulp*1 is used to recognize SUMO and cleaves at its C- terminus after GG (Hanington, Barreda et al. 2006) and tobacco etch virus (TEV) protease is used to cleave between Q and S/G of the consensus sequence E-X_{aa}-X_{aa}-Y-X_{aa}-Q-S/G (Carrington and Dougherty 1988). After digestion, the mixture of tag-cleaved protein, the tagged protease and uncleaved protein is applied to Ni-NTA resin again. Only tag-cleaved protein passes through. For further purification, size exclusion chromatography (HiLoad 16/60 Superdex 200 column or Superdex 200 10/300 GL column from GE Healthcare) (Paul-Dauphin, Karaca et al. 2007) and ion-exchange chromatography (column Q and S) (Sakash and Kantrowitz 2000) are used. Size exclusion chromatography separates proteins depending on their molecular weight and ion-exchange chromatography purifies proteins by different surface charges.

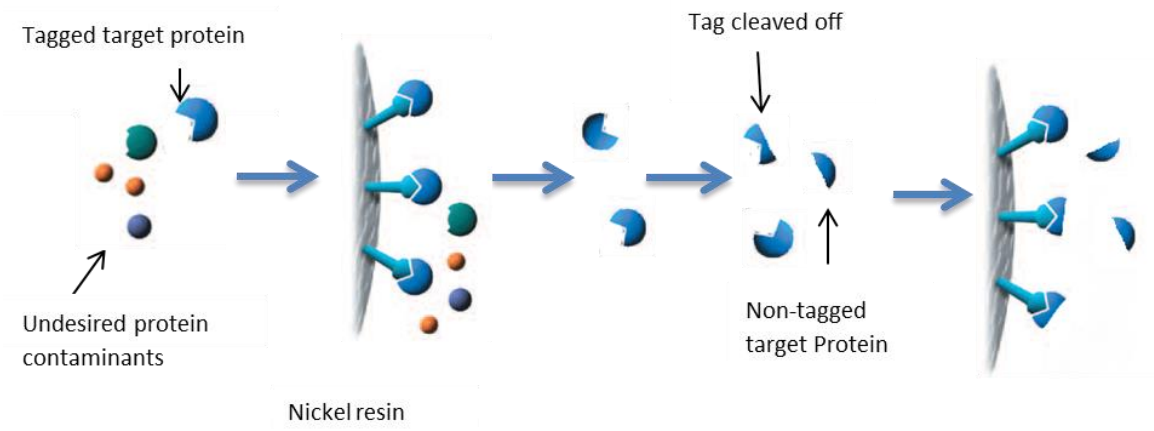


Figure 2-2: The scheme of two-step Nickel purification.

C. Protein Crystallization

1. Mechanism

Protein crystals are formed by protein molecules packing together through weak intermolecular interactions, including dipole-dipole interactions, hydrogen bonds, salt bridges, and van der Waals (vdW) interactions. To achieve this process, a saturated protein concentration should be reached. Figure 2-3 shows the protein solubility phase diagram at a given temperature. It explains that protein reaches its saturated state by increasing either protein concentration or the precipitant concentration (Rupp 2010).

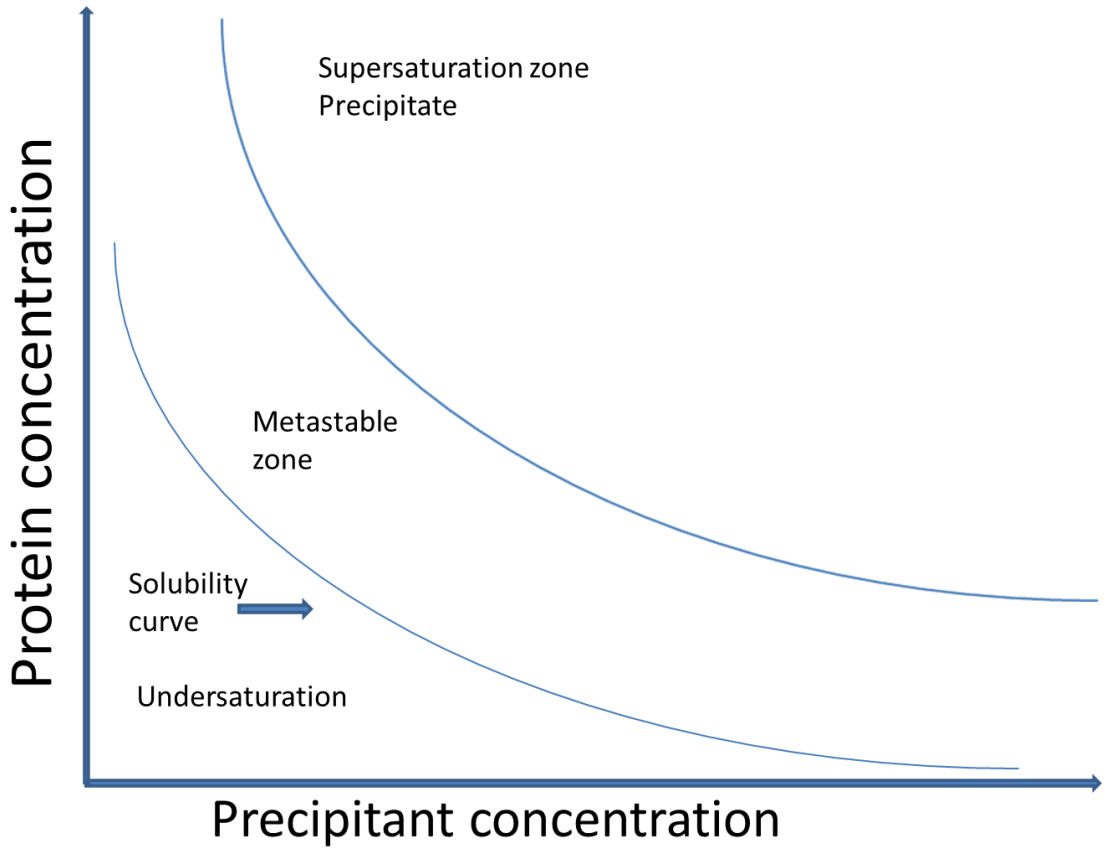


Figure 2-3: Protein solubility phase diagram. Adapted from (Rupp 2010).

Besides the availability of sufficient protein molecules, spontaneous crystal formation also requires a negative free energy. Free energy of crystallization ΔG is calculated from equation (2.1)

$$\Delta G = \Delta H - T(\Delta S_{protein} + \Delta S_{solvent}) \quad (2.1)$$

Given the weak interaction forming the protein crystal package, ΔH is only weakly negative during the protein crystallization. $\Delta S_{protein}$ is the entropy that the protein gains from the translation and rotation of protein molecules during crystallization and $\Delta S_{solvent}$ is the entropy that the solvent obtains from the release of water molecules from the protein surface. As the temperature during crystal forming is always positive, for the free energy to be negative, the entropy gain from

releasing ordered water molecules from hydrophobic residues has to be larger than the entropy lost from the forming of regular network of interactions (Derewenda and Vekilov 2006).

2. Protein Crystallization Methods

In order to obtain a crystal, protein solution is mixed with hundreds of crystallization solutions. A crystallization solution contains a buffer to keep a specific pH, a certain concentration precipitants and additives. With this crystallization solution, several techniques are used to drive a protein solution into a saturation state. Figure II-4 presents five widely used techniques in protein crystallization (Rupp 2010):

(1) Hanging-drop vapor diffusion;

(2) Sitting-drop vapor diffusion;

(3) Micro-dialysis;

(4) Micro-batch under oil;

(5) Free-interface diffusion.

In hanging-drop vapor diffusion, a mixture of protein solution and crystallization solution with a certain ratio is placed on a cover slide, which hangs above the crystallization solution in a sealed well.

In the sitting-drop vapor diffusion, the mixed protein solution and crystallization solution is placed on a shelf or a post. Compared with the hanging-drop vapor diffusion, the sitting-drop vapor diffusion is compatible for a robotic dispenser.

In micro-dialysis, protein is placed in a dialysis bag. The membrane of bag has specific size pores and allows water and salt, not the protein molecules, to exchange with crystallization solution

outside of dialysis bag. This method is not normally used in a large-scale initial screening, but would be suitable for growing large crystals under known conditions.

In micro-batch crystallization, the protein solution is mixed with the crystallization solution, and is placed under oil allowing the solvent to slowly evaporate from the mixture.

In the free-interface diffusion, protein solution and crystallization solution are placed in two chambers, and equilibrate against each other by channels.

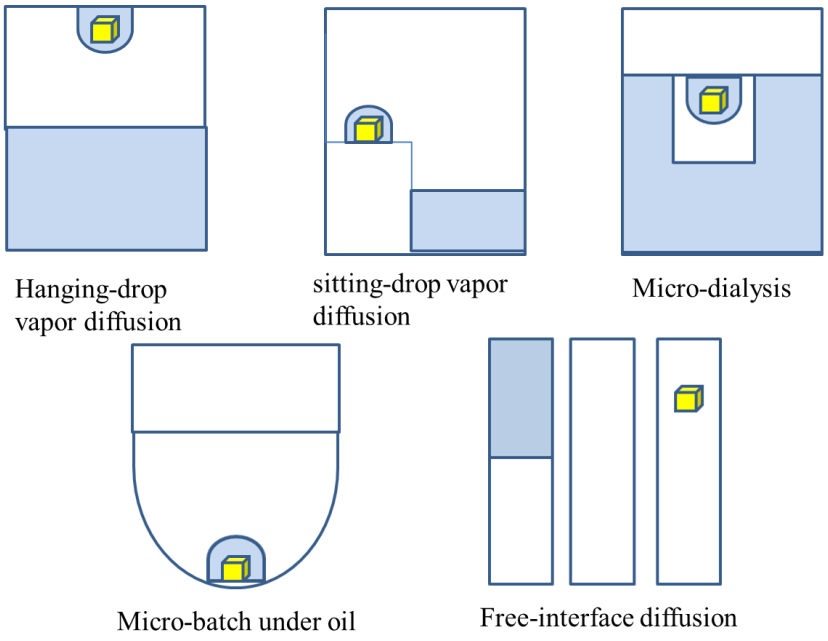


Figure 2-4: Five protein crystallization techniques. Adapted from (Rupp 2010).

D. X-ray Diffraction of Protein Crystals

A crystal is the periodic assembly of basic unit cells. So understanding the diffraction of crystal begins with understanding of the unit cell.

1. The Unit Cell and Bravais Lattices

The unit cell is the smallest unit in the protein crystal, and is defined by vectors a, b and c, and the angles α, β and γ (Figure 2-5). Depending on the values of these six parameters, protein crystals are divided into seven crystal systems, which are triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal and cubic (Figure 2-5).

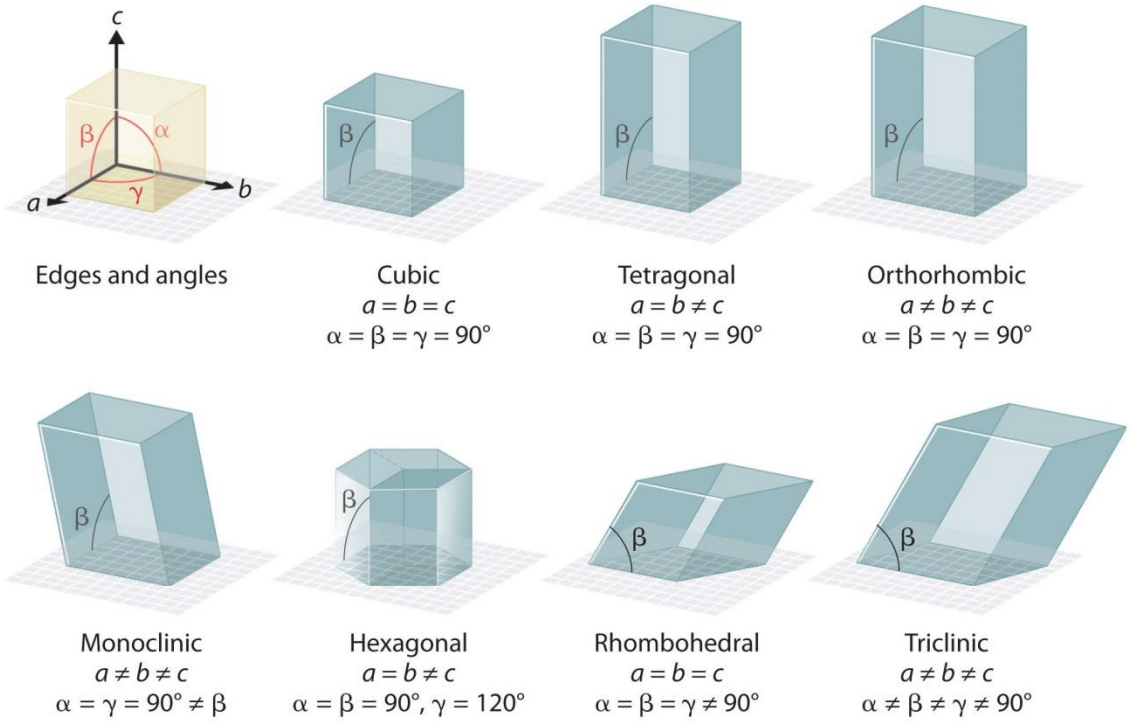


Figure 2-5: Unit cell and seven crystal systems. Copied from http://chemwiki.ucdavis.edu/Under_Construction/Lardbucket/Chapter_12/12.2_The_Arrangement_of_Atoms_in_Crystalline_Solids.

Besides these seven shape systems, unit cells are further defined by the positions of molecules located in the unit cell, into primitive, body centered and face centered (Figure 2-6).

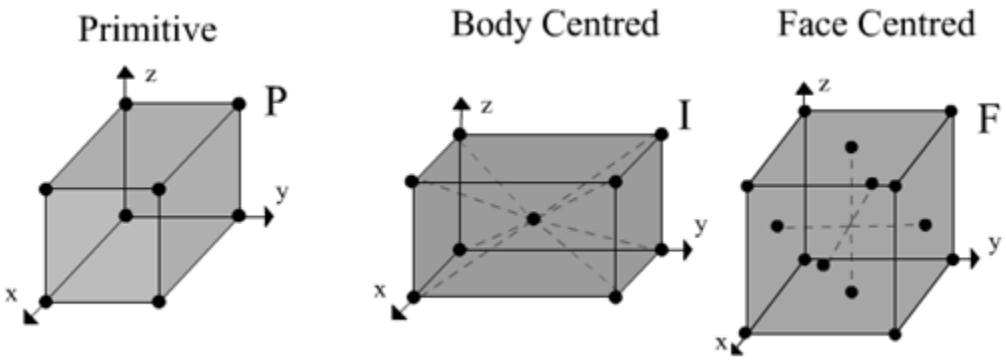


Figure 2-6: Primitive (P), body centred (I) and face centred (F) cells. Copied from http://www.doitpoms.ac.uk/tlplib/crystallography3/unit_cell.php

Combined seven crystal systems with three lattice types, it produces 14 Bravais lattices for protein crystals (Figure 2-7).(Blundell and Johnson 1976)

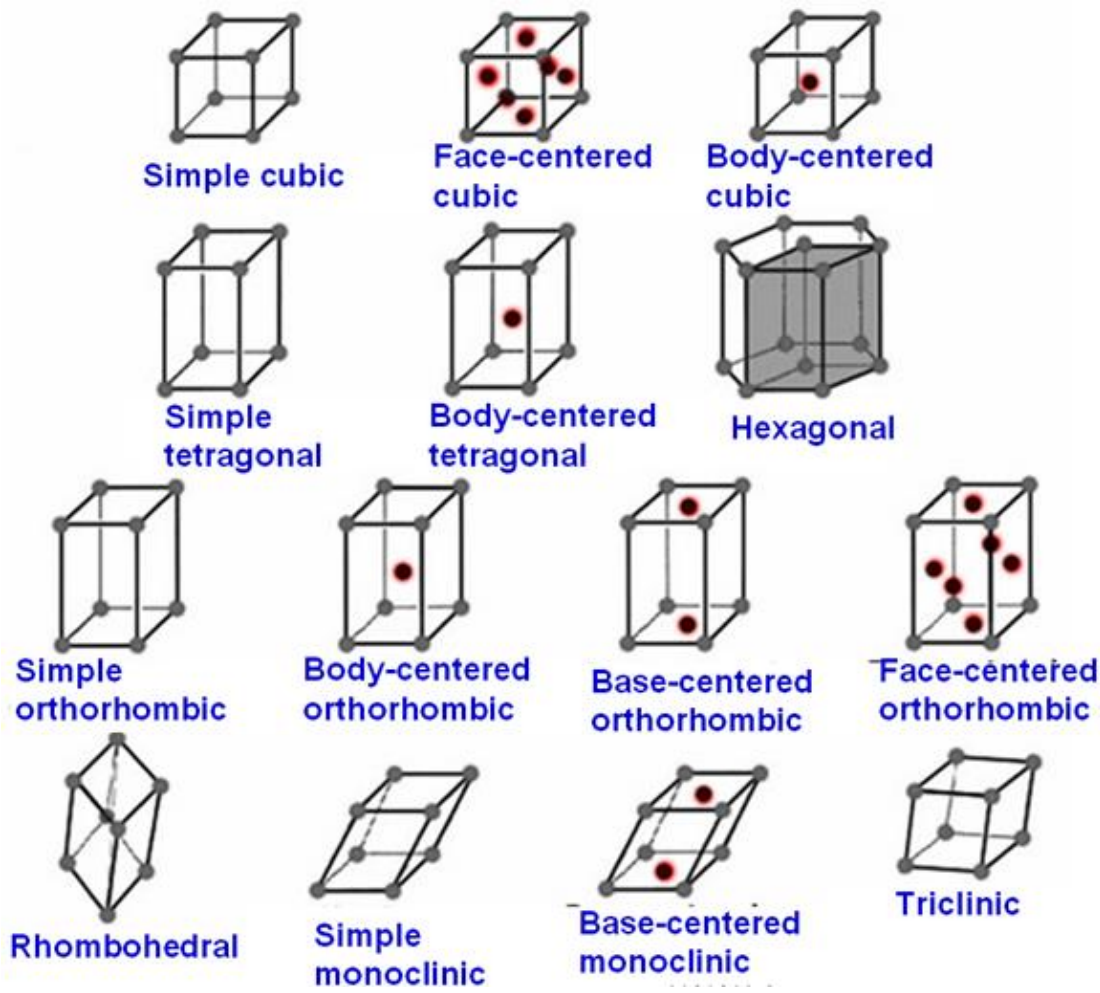


Figure 2-7: 14 Bravais lattices. Copied from http://www.seas.upenn.edu/~chem101/sschem/solidstatechem.html.

2. X-ray Scattering by Crystals

Successfully obtaining the crystal is the important step toward the eventual structure determination. The qualified single crystals will be used for data collection from x-ray diffraction.

When a beam of x-ray penetrates the protein crystals, most of them pass through the crystals, while a small amount of them will be scattered by the atoms. The scattering occurs from two processes. 1) Coherent scattering, also called Thomson scattering, happens when the electron is forced by the fluctuating electromagnetic field of the incident photon wave, to oscillate in the

same frequency, and gives out a second ray in the same wavelength and opposite phase. 2) Incoherent scattering, also known as Comton scattering, is random fluctuation of EM rays and is usually ignored in crystallography.

X-ray diffraction can only occur in certain directions with respect to the incident beam, which is defined as the Bragg law (Figure 2-8). In Bragg law, the crystal lattices are taken as a series of parallel planes with interplanar spacing distance d. The x-rays of wavelength λ impinge on these planes at an angle θ. Only when θ meets the condition 2dsin θ=n λ (n is an integer), the rays diffracted from successive planes in crystals are in phase with each other, producing a strong diffracted beam.(Blundell and Johnson 1976)

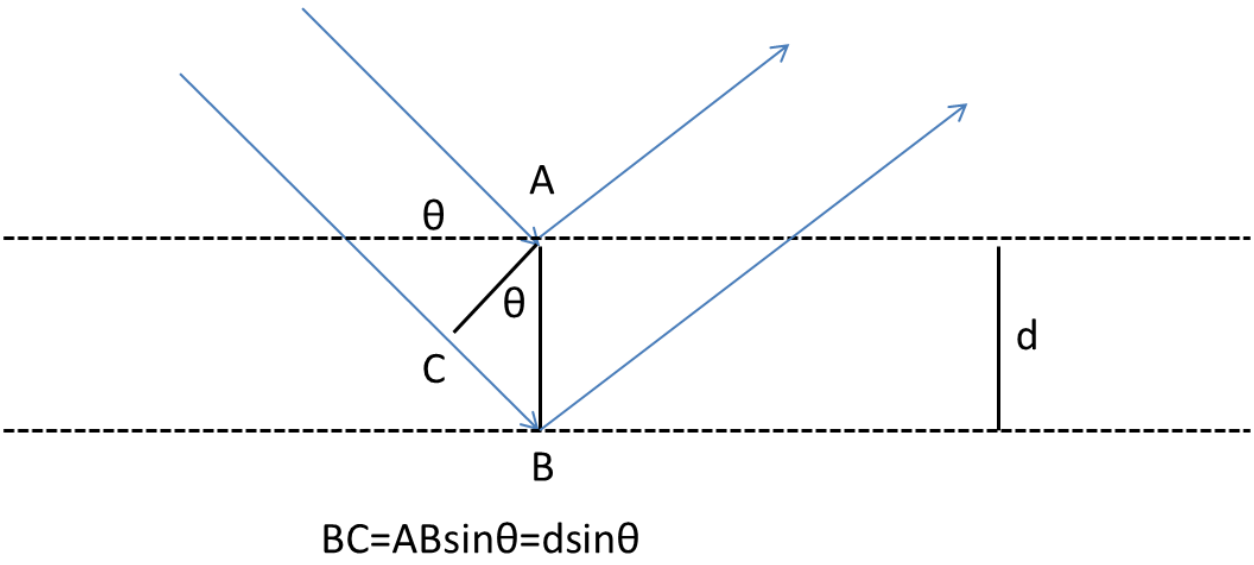


Figure 2-8: Bragg's law.

The scattering of x-ray by two adjacent particles is diagrammatized as in Figure 2-9, with S_0 being the incident vector and S_1 being scattered vector. The corresponding phase difference is

$$\Delta\varphi = 2\pi(s_1 - s_0)\mathrm{r} = 2\pi \mathrm{S} \cdot r \tag{2.2}$$

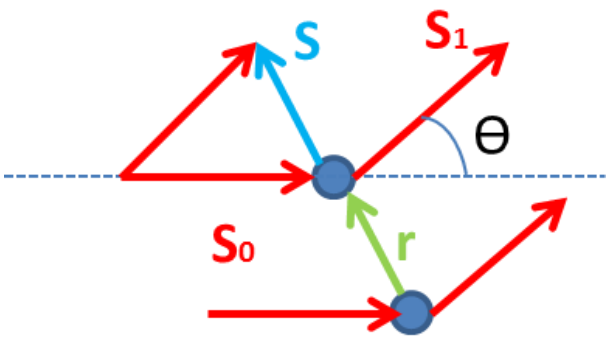


Figure 2-9: The scattering of x-ray by two atoms. Adopted from (Rupp 2010).

They can generate the maximal constructive interference when $\Delta\varphi_{max} = 2\pi \mathrm{S} \cdot r = n \cdot 2\pi$ and n being integer. When taking the three dimensions into consideration, we can obtain the Laue equation as (2.3):

$$\mathrm{S} \cdot \mathrm{a} = n_1, \mathrm{S} \cdot \mathrm{b} = n_2, \text{and } \mathrm{S} \cdot \mathrm{c} = n_3 \tag{2.3}$$

The a, b, c are unit cell vectors and the integers are related to Miller indices (*h, k, l*).

3. Reciprocal Lattice

The crystal lattices are treated as a series of parallel planes and the periodic repeat of the Bravais lattice. In order to visualize the spacing of reflections that can be collected, crystallographers construct an imaginary lattice, the reciprocal lattice, to represent the spacial organization of the Fourier transformation of the Bravais lattice. The Miller indices (*h, k, l*) are the smallest denominators in three-dimension lattice planes, which are related reciprocally to a, b, and c of the real crystal lattice.

One point in the real lattice is described as a position vector **r** and one point in the reciprocal lattice is described as a vector $\mathbf{r}^*$, and their axes are related as:

$$\begin{pmatrix} a^* \\ b^* \\ c^* \end{pmatrix} (a \quad b \quad c) = \begin{pmatrix} aa^* & ba^* & ca^* \\ ab^* & bb^* & cb^* \\ ac^* & bc^* & cc^* \end{pmatrix} = I \tag{2.4}$$

When the points in both real and reciprocal lattices are defined by planes, the corresponding interplanar spacing vectors $\mathbf{d}_{hkl}$ and $\mathbf{d}_{hkl}^*$ has the following relation:

$$\mathbf{d}_{hkl}{}^* = \frac{1}{\mathbf{d}_{hkl}} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* = (\mathbf{a}^* \quad \mathbf{b}^* \quad \mathbf{c}^*) \begin{pmatrix} h \\ k \\ l \end{pmatrix} = (\mathbf{A}^*)^{\mathrm{T}}\mathbf{h} \tag{2.5}$$

with **h** being the index(column) vector and $(A^*)^T$ being the matrix containing the reciprocal vectors in its column (Rupp 2010).

If we bring in Bragg's Law $2d\sin\theta = n\lambda$ (n is an integer) to determine when the diffraction happens, we shall find that for the set of lattice planes *h, k, l*, its reciprocal lattice plane vector $\mathbf{d}_{hkl}^*$ needs to meet the following requirement to diffract:

$$d_{hkl}{}^* = \frac{1}{d_{hkl}} = \frac{2\sin\theta}{n\lambda} \tag{2.6}$$

That is to say, when a point in reciprocal lattice meets the circle with origin at the crystal and a radius being $1/\lambda$ (wavelength), which is defined as Ewald sphere, this point satisfies Bragg's Law and a reflection can be observed if CCD detector is present.(Ayer 1989)

E. The Mathematics of Crystallography

In protein crystallography, each reflection is the x-ray diffracted from the crystal and thus can be described as x-ray wave. The x-ray wave is described by a periodic function:

$$f(x) = Fsin2\pi(hx + \alpha)\ or\ f(x) = Fcos2\pi(hx + \alpha) \tag{2.7}$$

In these functions, the constant *F* specifies the amplitude of the wave; the constant *h* specifies the frequency of the wave; the variable *x* implies the position of the wave from the origin; and the constant α is the phase of the wave.

For the complicated wave, it can be described as the sum of a lot of simple waves, which is called a Fourier series, and written in the following form.

$$f(x) = \sum_{h=0}^{n} F_h\, cos2\pi(hx + \alpha_h)\ or\ f(x) = \sum_{h=0}^{n} F_h\, sin2\pi(hx + \alpha_h) \tag{2.8}$$

In these two equations, the waveforms of cosine and sine can be combined as $[\cos 2\alpha\pi(hx)+i\sin 2\pi(hx)]$, and the equation for the description of reflection is described as:

$$f(x) = \sum_{h=0}^{n} F_h\, [cos2\pi(hx) + isin2\pi(hx)] \tag{2.9}$$

Further, $\cos\theta+i\sin\theta=e^{i\theta}$, so the description of reflection is described as:

$$f(x) = \sum_{h=0}^{n} F_h\, e^{2\pi i(hx)} \quad \text{or simply} \quad f(x) = \sum_{h} F_h\, e^{2\pi i(hx)} \tag{2.10}$$

For the three-dimensional wave with the variable *h, k, l*, the *f(x,y,z)* is written in the form:

$$f(x,y,z) = \sum_{h} \sum_{k} \sum_{l} F_{hkl}\, e^{2\pi i(hx+ky+lz)} \tag{2.11}$$

The Fourier transformation (FT) of *f(x), F(h)* is written in the form:

$$F(h) = \int_{-\infty}^{+\infty} f(x)\, e^{2\pi i(hx)} dx \tag{2.12}$$

In this equation *F(h)* is called the Fourier transformation of *f(x)*, the units of the variable h are reciprocals of the units of x. For the three-dimensional wave, fourier transform of *f(x,y,z)*, *F(h, k, l)* is described as follows:

$$F(h,k,l) = \iiint_{x\,y\,z} f(x,y,z)\, e^{2\pi i(hx+ky+lz)} dxdydz \tag{2.13}$$

In this equation, *F(h, k, l)* is the structure factor that describes reflection *hkl*; *f(x, y, z)* is the electron density at position *(x, y, z)*; and *dxdydz* is equal to the volume V. This equation also is written as:

$$F(h,k,l) = \iiint_{x\,y\,z} \rho(x,y,z)\, e^{2\pi i(hx+ky+lz)} dV \tag{2.14}$$

As the Fourier transformation operation is reversible, the electron density *ρ(x, y, z)* is the transformation of the structure factors, (Rupp 2010)

$$\rho(x,y\,z) = \frac{1}{v} \sum_{h} \sum_{k} \sum_{l} F_{hkl}\; e^{-2\pi i(hx+ky+lz)} \tag{2.15}$$

F. Data Collection, Processing and Analysis

1. Data Collection

Data collection in protein crystallography is the recording of the intensity of all reflections diffracted (or as much as possible) from protein crystals. Several factors affect the data quality, such as x-ray source, cryo-condition, exposure time and the quality of crystal.

Protein crystallographers use both the in-house and synchrotron x-ray source. The in-house x-ray source produces x-ray through bombarding the metal target (most commonly copper) with accelerated electrons, which displace electrons from a low-lying orbital in target metal atom. .For Cu anode x-ray generator, electrons dropping from the L shell of Cu to replace displaced K shell electrons emit X rays of 1.54 Å wavelength.

At the synchrotron x-ray source, electrons in the storage rings accelerate to velocities near the speed of light. These electrons emit synchrotron radiation in a wide range of wavelengths when forced into curved motion by powerful magnets. X-rays at selectable wavelengths are provided by focusing mirrors and monochromators system. In US, synchrotron x-ray sources are available at Advanced Photon Source at Argonne National Laboratory, Cornell High-energy Synchrotron Source, National Synchrotron Light Source at Brookhaven National Laboratory, and Advanced Light source at Lawrence Berkeley National Laboratory.

The synchrotron x-ray source is much more powerful than the in-house source, and thus can be utilized to collect high resolution data sets, or derivative data sets for phasing calculation, which will be described later.

Data collection is generally performed at 100 K to minimize the radiation damage and extend the survival time of crystals during data collection. However, this low temperature could also induce ice formation within the crystals that may disrupt protein lattices and cause either loss of diffraction or lead to anisotropic diffraction. To prevent this commonly found problem associated with cryogenic condition, an optimal cryoprotectant at the correct concentration is necessary. Therefore, a cryo-solution is routinely screened experimentally. Glycerol, PEG, ethylene glycol,

sucrose, or salt at various concentrations are widely chosen as cryo-solutions, whose concentrations have to be precisely determined empirically.

Before starting the full data collection, several images are collected to optimize the exposure time, the distance between crystal and detector, and strategy for data collection.

Longer exposure time will surely increase the diffraction intensity and enhance weak diffraction, but it will also boost up radiation damage to the crystal. The distance between crystal and detector should be far enough for the detector to collect all diffraction spots, especially those high resolution spots at the edge, but not too far so spots can be distinctly separated. Depending on the symmetry of the crystal (described in C.1), the angle the crystal needs to rotate so a full set of data can be collected differs. We certainly want as much data as possible, but the longer the crystal is exposed to x-rays, the bigger the radiation damage is. So pros and cons are carefully considered during strategy optimization.

2. Data Processing

After the data collection, raw data set is processed to assign the unique indices (h, k, l) with their observed intensities. Data processing involves several steps, including indexing, integration, and scaling. Indexing is the process to determine the unit cell parameters, crystal orientation and decide the actual Bravais lattice symmetry of the given crystal. Integration is to identify the background and spot region and measure the intensity of the spot. Scaling is the process to correct the diffraction intensities. Various physical factors lead to observed intensities being on different scales. These factors include: 1) the incident beam and the camera; 2) the crystal and the diffracted beam; and 3) the detector. After the scaling, the multiple measurements of equivalent reflections are merged to produce unique reflection.

3. Data Analysis

The quality of the collected dataset is analyzed by several statistical parameters: the highest resolution, completeness, redundancy, the number of unique reflections, the mean signal to noise ratio of the reflections (I/σI) and R_{merge}. The data quality indicator R_{merge} is the linear merging R factor from given amount (N) of reflections **h** within a resolution range.

$$R_{merge} = \frac{\sum_{\mathbf{h}} \sum_{i=1}^{N} \left| I_{(\mathbf{h})i} - \overline{I_{(\mathbf{h})}} \right|}{\sum_{\mathbf{h}} \sum_{i=1}^{N} I_{(\mathbf{h})i}} \tag{2.16}$$

$\overline{I_{(\mathbf{h})}}$ is the averaged intensity of every reflection. R_{merge} is computed for entire data set or for each resolution shells. When relative error increases and reflections are weak, the R_{merge} increases rapidly in high resolution.

A typical 'good' data set meets the following criteria. It has high resolution. The ideal completeness is 100%. The redundancy is reasonable according to symmetry. Every unique reflections should be collected at least once. The lowest mean signal to noise ratio of the reflections (I/σI) presenting at the highest resolution shell has a cutoff value 2.0. R_{merge} is low.

G. Structure Determination

1. Crystallographic Phase Problem

When reconstructing electron density using the Fourier back-transformation, we can reverse the equation (2.15) we described in chapter D as below,

$$\sum_{\mathbf{h}=-\infty}^{+\infty} |F(\mathbf{h})| \cdot \exp[-2\pi i(\mathbf{h} \cdot \mathbf{r}) + i\varphi(\mathbf{h})] = \rho(\mathbf{r}) \tag{2.17}$$

$|F(\mathbf{h})|$ is proportional to the square root of I, thus it is measurable, leaving the phase information φ(h) lost. (Taylor 2003)

This is the well-known 'Phase Problem' in crystallography. Phase problem can be solved by ab initio determination via experimental procedures, such as from anomalous scattering, or by mathematic calculations from available phases obtained from an existing model, using molecular replacement.

2. Molecular Replacement

If a model from an already determined homolog structure is available, one can obtain the phases of an unknown protein rapidly from its diffraction intensities.

If the unknown protein and the model are isomorphous, they not only share the same space group and unit cell dimensions, but also have almost the same atoms at the same position. In this case, the phases of crystal of unknown protein can be directly computed from its intensity.

If they are not isomorphous, as in most cases, we can determine the model position and orientation by comparing the calculated structure factors amplitudes from the model and measured amplitudes from the unknown protein. Before digging into details, we need to first understand the concept of the Patterson Function.

a. Patterson Function

The Patterson function arises from the autocorrelation of molecular structure. The value is directly calculated from the diffraction intensity by putting each atom at the origin, and plotted into Patterson maps.

$$P(uvw) = \frac{2}{V}\sum_{h=0}^{+\infty}\sum_{k=-\infty}^{+\infty}\sum_{l=-\infty}^{+\infty} F_{\mathrm{h}}^2 \cos 2\pi(hu + kv + lw) \tag{2.17}$$

The peaks generated indicate fitted interatomic vectors in a unit cell and thus reveal information about distances between atoms.

Patterson maps cannot be used to solve the phases of protein molecule, for the amount of atoms is large in proteins, but it can be applied to identify substructural position of heavy atoms, locate anomalous scatterers and orientate the molecular model in molecular replacement. Software like SOLVE (Terwilliger and Berendzen 1999) (http://solve.lanl.gov/) is based on Patterson function.(Rupp 2010)

Since molecules with the same orientation and different positions in the unit cell have the same Patterson map, a trial-and-error method can help to determine best orientation of model for the new protein (Tickle and Driessen 1996).

After the relative orientation of unknown protein was determined, this information is used to obtain the position of noncrystallographic operators related to crystallographic symmetry in the unit cell. The Patterson functions of intermolecular vectors are calculated to find the translation vectors of unknown structure. Finally, the initial phase is improved by extension from low resolution to high resolution (Rossmann 1990).

3. Experimental Phasing

Experimental phasing is to obtain phase information from the intensity information collected during x-ray diffraction experiment.

a. Marker Atom Substructure Methods

The first step of solving protein structure without any prior structure information involves determination of marker atoms substructure positions. Marker atoms covalently bond to protein molecule produce electronic difference map and thus generate initial protein phases. It is normally carried out by the Patterson function and/or the direct method as discussed above.

b. Isomorphous Replacement

Isomorphous replacement is a method to simplify the structure solving by introducing a set of strong diffractors called heavy atoms, and locating them in a unit cell before solving the structure of entire protein. Two types of crystals are used. One is the native crystal and the other is the derivative crystal. A derivative crystal contains heavy atoms introduced by heavy metal ions soaking. The reflections from derivative crystal are the sum of those from native crystal and those from heavy atoms.

Assuming the crystal lattice remains the same after heavy atom soaking, the reflection from derivative crystal $\mathbf{F}_{HP}$ is made up by the reflection from heavy atom $\mathbf{F}_{H}$ and that from native crystal $\mathbf{F}_{P}$.

$$\mathbf{F}_{HP} = \mathbf{F}_{H} + \mathbf{F}_{P} \tag{2.20}$$

The intensity difference between native crystal and derivative crystal is generated by subtraction of those two (Figure 2-10), and used to compute both the amplitudes and phases of structure factor of heavy atoms F_H via difference Patterson function or direct method (Rupp 2010).

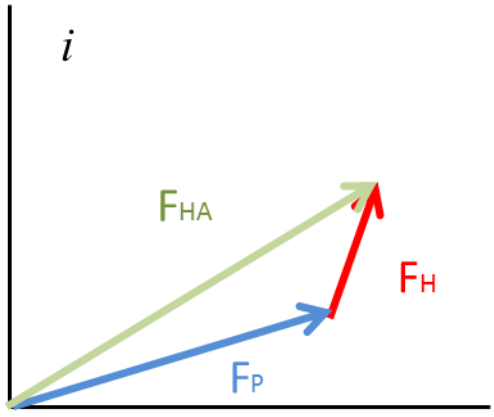


Figure 2-10: The complex structure factors of native crystal and derivative crystal with length being amplitude and angle representing phase.

F_H provides phase information to solve F_P. However, single isomorphous replacement (SIR) using only one heavy atom has ambiguity of obtaining two phase angles. If we use a Harker diagram to graphically represent the phase equation as in Figure 2-11, the two solutions are indicated by orange arrows.

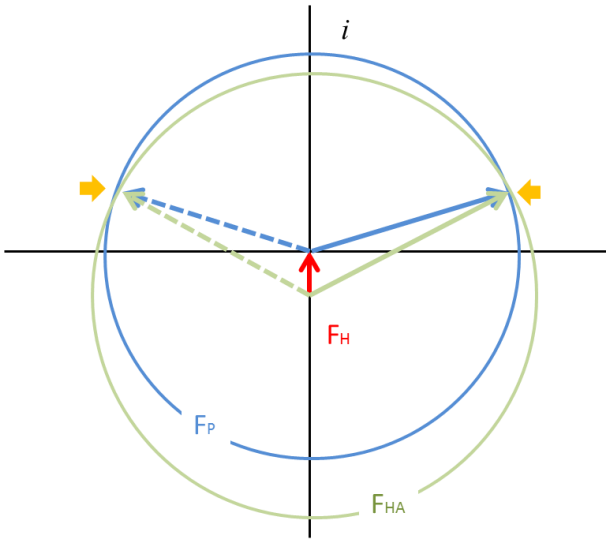


Figure 2-11: Harker diagram of SIR.

In this case, we need to introduce a second derivative for multiple isomorphous replacement (MIR). Shown in Figure 2-12, the addition of F_{HA2} resolves the ambiguity.

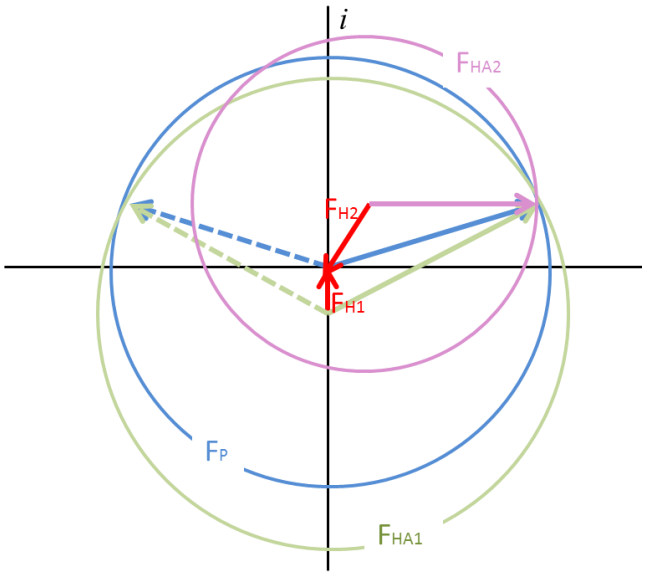


Figure 2-12: Harker diagram of MIR.

c. Anomalous Scattering

Anomalous scattering is the phenomenon that heavy atoms absorb x-ray anisotropically. It results in unsymmetrical reflection.

Friedel pairs are centrosymmetric structure pairs with the same magnitude and opposed phase angle.

$$\mathbf{F_h} = |F_h|(\cos\varphi_h + i\sin\varphi_h) = A_h + iB_h \tag{2.21}$$

$$\mathbf{F_{-h}} = |F_h|(\cos(-\varphi_h) + i\sin(-\varphi_h)) = |F_h|(\cos\varphi_h - i\sin\varphi_h) = A_h - iB_h$$

However, in the presence of heavy elements, which exerts a strong attraction to inner shell electrons, the ratio of inelastic scattering increases. The intensity obtains a real component f' while the phase gains an imaginary component f''.

The entire scattering is described with the following equation (Rupp 2010)

$$f = f^0 + f' + if'' \tag{2.21}$$

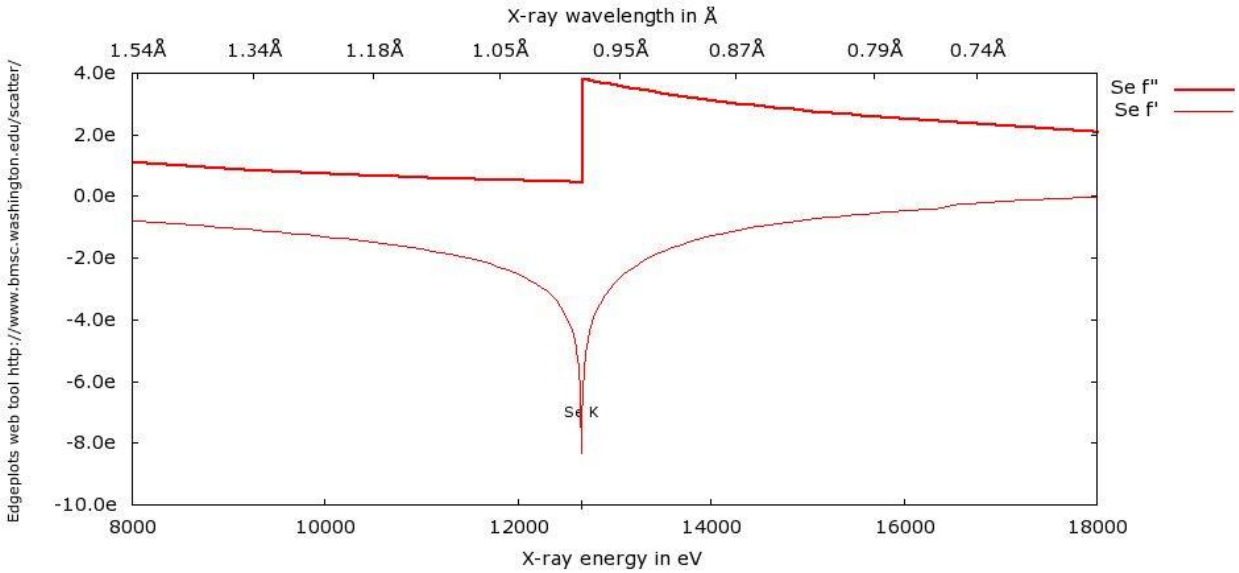


Figure 2-11 Example of Se f' and f'' curves. f' is the real anomalous scattering contribution and f'' is the imaginary anomalous scattering contribution. The sharp jump in the f'' curve indicates absorption edge of Se element. Copied from http://www.bmsc.washington.edu/scatter/.

f' is the Karmers-Kronig transformation of normalized x-ray edge scan and f'' is proportional to atomic absorption coefficient μ. Both can be obtained directly from the experiment.

This generated anomalous difference, consisting the introduced real component Fa($f^0 + f'$) and imaginary component Fa″(f''), breaks the internal centrosymmetry of Friedel pairs (Figure 2-12).

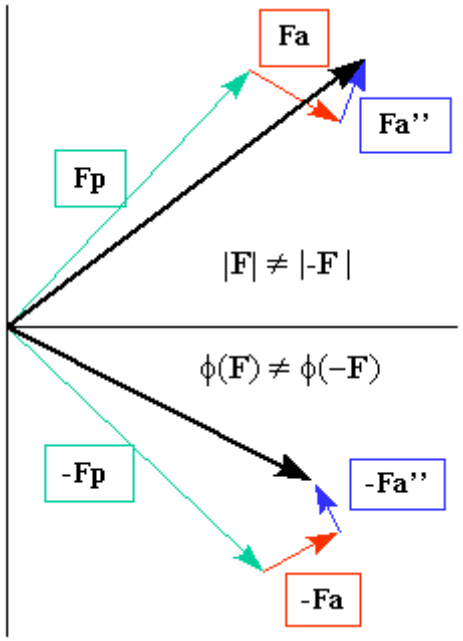


Figure 2-12: The scattering contribution from an anomalous scatterer in a matrix of normal scatterers. Copied from http://www.xtal.iqfr.csic.es/Cristalografia/parte_07-en.html.

The magnitudes of Fa and Fa″ are dependent on the scattering element and the phases of them are relied on the position of the heavy atom. After obtaining the heavy atom positions by Patterson method and calculating phases, the Fa and Fa″ are used to solve the structure factors of protein.

d. MAD/SAD

Multiwavelength anomalous diffraction (MAD) is the method to collect data at multiple wavelengths to obtain maximum information from the same heavy atom derivative crystal.

Absorption edge is a function of an atom in which its x-ray absorption ability drops significantly at the wavelength just below its characteristic emission wavelength K_{β}, caused by excited electron departure.

If the anomalous differences are strong enough when the data set is collected at the wavelength of maximum f'', phases can be solved just by the peak data set using single-wavelength anomalous

diffraction (SAD). In other cases, we collect three sets of data adjusting the x-ray to different wavelengths (Figure 2-13): one peak data set at the *f″* peak has the maximum anomalous contribution thus create an intensity difference; one inflection data set at the *f′* negative peak to generate maximum dispersive difference when comparing to the remote data; and one remote data set at a high energy wavelength around 200 eV above the absorption edge. This multi-wavelength anomalous diffraction (MAD) is used to obtain as much information as possible (Giacovazzo and Siliqi 2004).

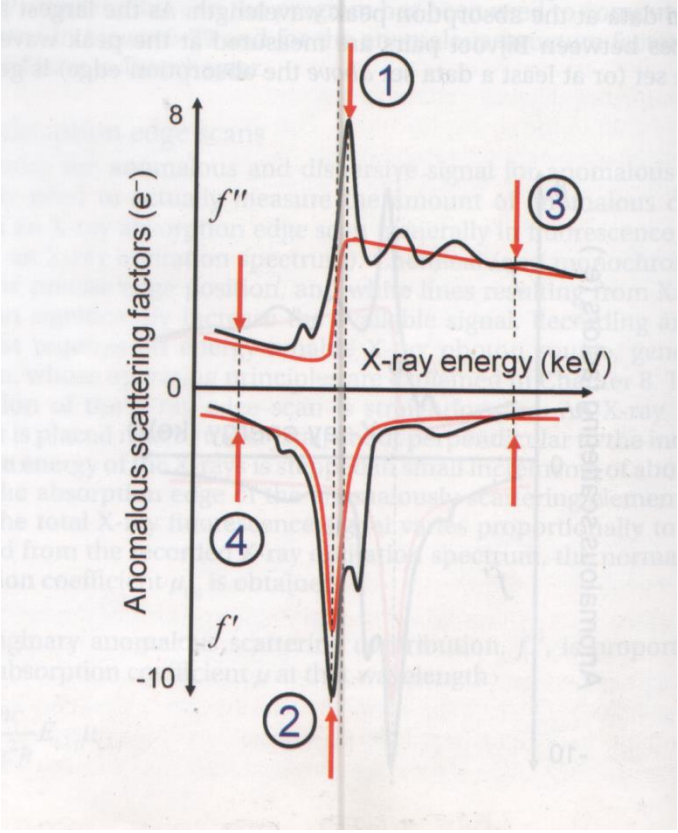


Figure 2-13: Choice of wavelengths for anomalous data collection. (Rupp 2010)

e. Direct Method

The Direct method is based on the theory of triple relation that three phases are related by the following formula.

$$\alpha_{-h} + \alpha_{h'} + \alpha_{h-h'} \cong 0 \tag{2.18}$$

Thus, the tangent formula

$$\tan \alpha_{\mathrm{h}} = \frac{< E_{\mathrm{h}'} E_{\mathrm{h}-\mathrm{h}'} \sin(\alpha_{\mathrm{h}'} + \alpha_{\mathrm{h}-\mathrm{h}'}) >_{\mathrm{h}'}}{< E_{\mathrm{h}'} E_{\mathrm{h}-\mathrm{h}'} \cos(\alpha_{\mathrm{h}'} + \alpha_{\mathrm{h}-\mathrm{h}'}) >_{\mathrm{h}'}} \tag{2.19}$$

shows once the phases of some reflections are known, the phases of other reflections can be deduced by bootstrapping (Taylor 2010). The Direct method is usually used to determine the structure of small molecules at atomic resolution (<1.2 Å), or find the heavy atoms in derivative crystals using programs like SHELXD (Schneider and Sheldrick 2002) (http://shelx.uni-ac.gwdg.de/SHELX).

H. Model Building and Refinement

Model building is to construct an atomic structure model that fits into the electron density. Refinement is to match observed and calculated intensities by shifting atomic parameters. They are performed tightly and repeatedly if necessary.

1. Model Building

After phasing calculation, an interpretable electron density map is produced. Next step is the model building, which is building the backbone of amino acid and fitting it to the electron density map.

When resolution is high, model building is easy for the side chains are so distinctly present. However, building model from low resolution data can be problematic.

Several softwares, such as ARP/wARP (http://www.embl-hamburg.de/ARP/) (Langer, Cohen et al. 2008) and RESOLVE (https://solve.lanl.gov/) (Terwilliger and Berendzen 1999, Terwilliger 2000), could serve the auto-building if the resolution of electron density map is good enough.

During manual model building, Cα-atoms are put at the side chain branching point as a start, followed by placing distinct amino acids like tryptophan and phenylalanine as well as di-sulfade bonds into the electron density map. Secondary structure of the protein is normally predicted to assist building and typical helixes or sheets backbone is easily discerned and located in the electron density map. As for these unconnected secondary structures, heavy atoms are used as markers to identify the exact sequences of the helix or sheet. In addition, amino acids with large side chains, such as trypophan, tyrosine or lysine are also used as markers.

At this stage, those amino acids with unclear side chain are designated as alanines. The basic principle during the first crude model building is to only assign amino acids based on electron density map, without introducing error to the model. Next, this crude model is used in combination with original native intensity to calculate structure factors and the phases. This step will generate a clearer map, which is used for further map improvement.

2. Refinement

Refinement is to adjust the parameters describing a model in order to fit them with experimental observations. But we have more experimental data points (n) available than model parameters (p).

In order to increase data-to-parameter ratio, besides reflections data, the stereochemical restraints are included such as bond lengths, bond angles, planarity and torsion angles to increase the number of data. Also, constraints are applied to reduce the number of parameters.

The R factor is used to describe difference between the calculated structure from model and observed native structure factor. It indicates how well the model is superimposed with the real protein molecule.(Rupp 2010)

$$R = \frac{\sum\left||\mathrm{F_{obs}}| - |\mathrm{F_{calc}}|\right|}{\sum|\mathrm{F_{obs}}|} \tag{2.22}$$

After iterative improvement of maps and models, the R factor could be lowered to less than 30%. After water molecules around the protein are supplied to the model, the R factor is further reduced.

I. Validation

Before we certify a structure as 'a decent structure', several validations need to be taken into consideration.

R_{free} is the R factor of a set of reflections not used in refinement and R_{work} is the R factor of the set of reflections used in refinement. If a refinement improves R_{free} as well as R_{work}, it qualifies a good refinement.

A commonly allowable difference between R_{work} and R_{free} is less than 5%, and R_{work} being 1/10 of the resolution. For example, a 2.2A resolution dataset may have a R_{work} around 22%.

The Ramachandran plot is a diagram displaying the torsion angles ψ and ϕ of peptide bond between non-glycine amino acids (Figure 2-14). It defines three regions of possible backbone conformations that amino acids forming α helix, β sheet and turns would fall into, and a reasonable structure shouldn't contain any outliners.

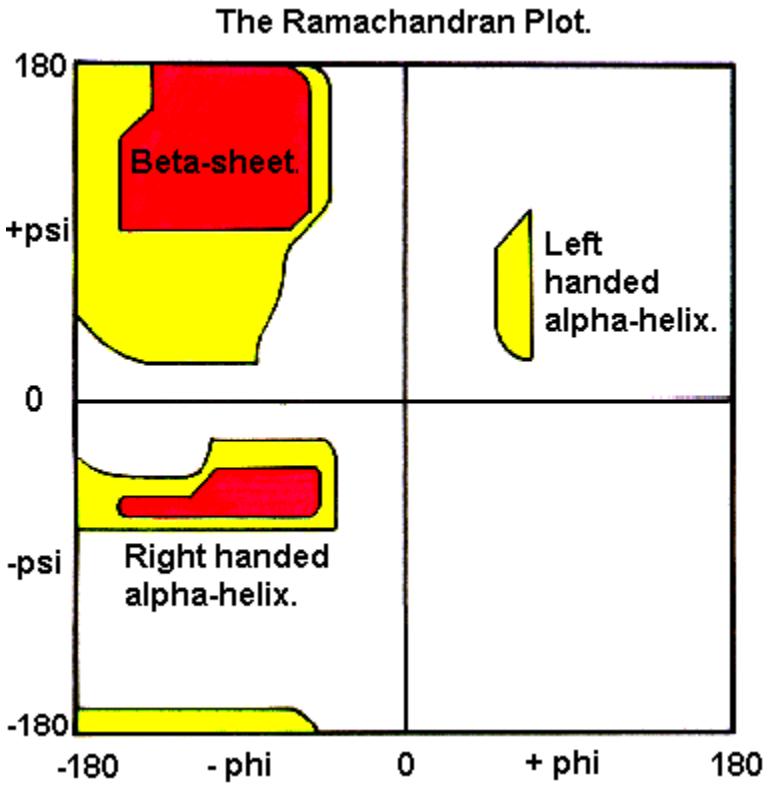


Figure 2-14: A Ramachandran plot showing possible ψ and ϕ values of α helix and β sheet. Copied from http://swissmodel.expasy.org/course/text/chapter1.htm.

Geometry root mean squate deviation (RMSD) of bond length and bond angle is computed and should be less than 0.01 and 1, accordingly.

Amino acid side chain rotamer conformation is taken into consideration when multiple positions are present with individual occupancy.

Structure is examined at any close contacts to identify whether the contact is rational or not.

Servers like Molprobity (http://molprobity.biochem.duke.edu/)(Chen, Arendall et al. 2010) are used to check validation of structure.

CHAPTER III

LRRK2

A. Introduction

Leucine-rich repeat kinase 2 (LRRK2) is a multi-domain protein expressed widely in organs including brain, heart and liver. Mutations of LRRK2 are largely found in patients of Parkinson's disease, an age-related neurodegenerative disease affecting six million people around the world. A major part of pathological mutations are located in the central part of LRRK2, containing a Ras of Complex (ROC) GTPase domain, a kinase domain and a domain between them called C-terminal of Roc (COR) domain. Previous research has provided evidence that the Roc domain regulates kinase activity as a binary switch through a GTP/GDP bound cycle. The structure of the inactivated Roc has been solved previously in our lab and we propose to determine the structure of ROC in the activated state. The expected results would reveal the dynamic conformational changes in the ROC GTPase domain, providing important clues on the mechanisms of GTP hydrolysis and signaling.

1. Parkinson's disease

Parkinson's disease (PD) is a common age-related neurodegenerative disease.

PD is the most commonly observed movement disorder. The physical symptoms of PD include muscle rigidity, bradykinesia, resting tremor and postural instability. It affects six million people

over the world (Goedert 2001). Its demographics vary among age groups, increasing significantly along with age. PD affects 17.4 out of 100 000 person from 50 to 59 years old and 93.1 out of 100, 000 person years from 70 to 79 years old (Bower, Maraganore et al. 1999).

PD is a progressive bradykinetic disorder. The pathological phenotype of PD is severe loss of dopaminergic neuromelanin-containing neuronal cells in the pars compacta of substantia nigra in the midbrain, as well as protein aggregation enriched with α-synuclein in three intraneuronal inclusions in the surviving neurons: the Lewy body, the pale body and the Lewy neurite (Hassler 1938). The loss of dopamine causes motor abnormality in PD patients.

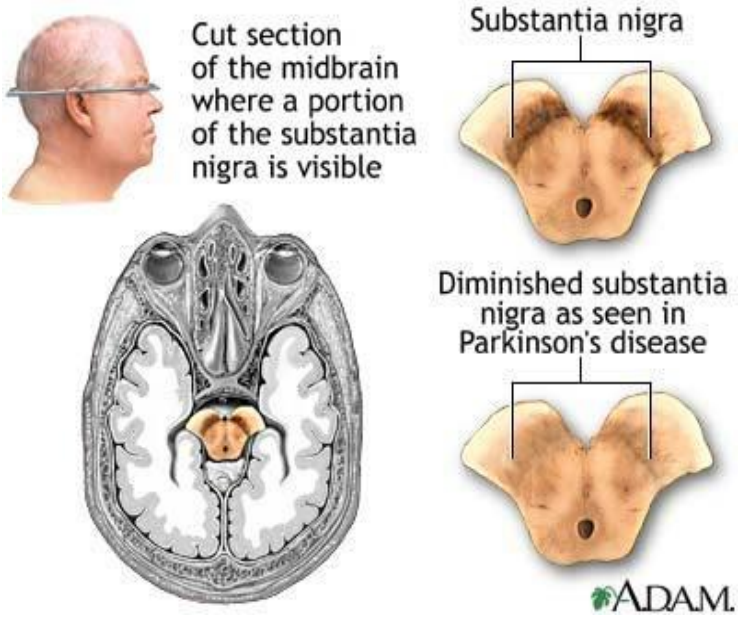


Figure 3- 2: Normal and diminished substantia nigra in PD. Copied from A.D.A.M. 2011.

Most PD cases are sporadic and idiopathic, while few are possibly implicated by environmental factors. For example, researchers found out toxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MTPT) induces parkinsonism (Langston, Ballard et al. 1983). Genetic factors were found to also play a role, since mutations in LRRK2 and α-synuclein genes were found in both familial and sporadic PD (Polymeropoulos, Lavedan et al. 1997) (Krüger 1998) (Horowitz and Greenamyre 2010).

Although therapeutic treatments targeting individual psychological symptoms of PD have been developed through clinical study (Lees, Hardy et al. 2009), none of them is able to stop the undergoing neurodegenerative process.

2. LRRK2

Recent genetic research has revealed some connection between the leucine rich repeat kinase 2 (LRRK2) and PD. Mutations of LRRK2 being largely identified in both familial and sporadic PD made the protein a promising pharmaceutical target against PD (Aasly, Toft et al. 2005).

LRRK2 is a multidomain protein which forms a functional dimer.
Through genome-wide association study, the gene *LRRK2* encoding a 2527 residues. protein was identified as one of the most likely susceptibility causative genes for autosomal dominant Parkinsonism (Satake, Nakabayashi et al. 2009).

LRRK2 belongs to the ROCO protein family and contains six domains. As seen in Figure 3-2, LRRK2 contains an ankyrin repeats domain, a leucine-rich repeats (LRR) domain, a Ras of complex (Roc) domain, a carboxy-terminal of Roc (COR) domain, a (serine/threonine) kinase domain and a WD40 domain. Most pathogenic mutations are located in the three central domains: Roc, COR and kinase domain, for example, R1441C, R1441G, R1441H in the Roc domain, Y1699C in the COR domain and G2019S and I2020T in the kinase domain. (Cookson 2010) LRRK2 requires dimerization through several interfaces across the protein to be kinase-active (Greggio, Zambrano et al. 2008), indicating it regulates its own function. The portion of dimerization dependents on kinase activity, which might further be related to pathology (Sen, Webber et al. 2009).

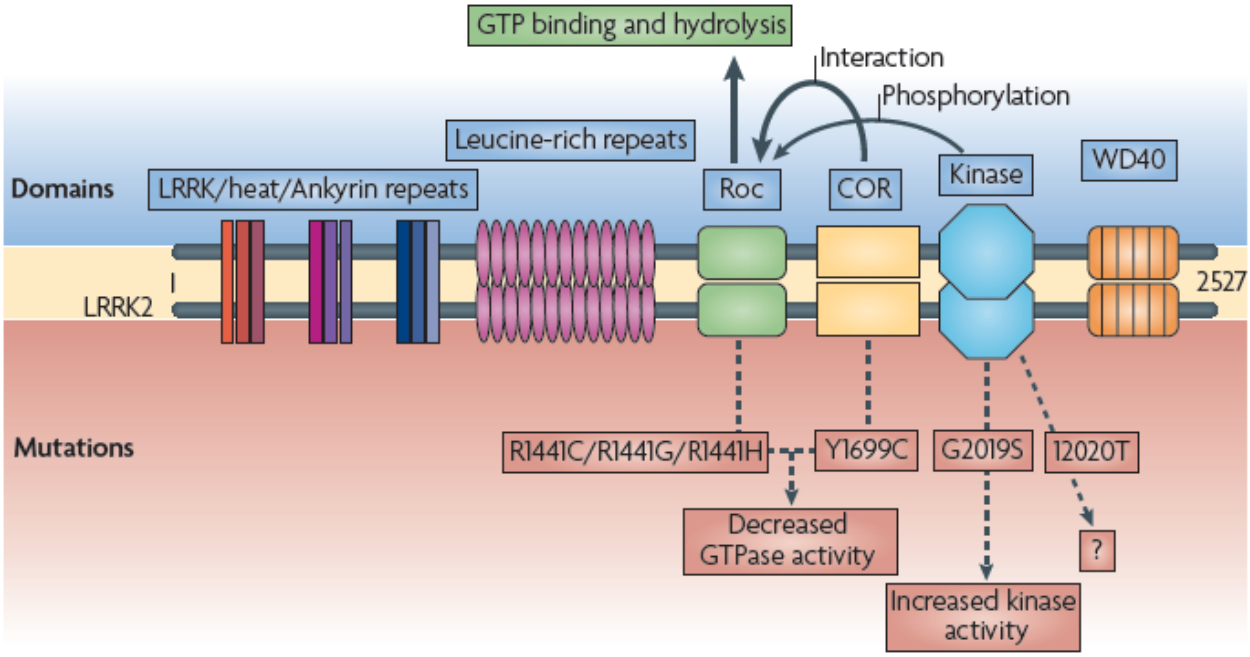


Figure 3-2: Scheme of the domain architecture of LRRK2. LRRK2 dimerizes through several interface. Functions and some important pathogenic mutations are indicated (Cookson 2010).

3. Roc domain

Roc GTPase domain regulates kinase activity and LRRK2 toxicity.

The Roc domain is an authentic GTPase. It binds and hydrolyzes GTP (Deng, Lewis et al. 2008). Mutation in the ROC domain abolished its GTP binding ability, and also incapacitated the kinase activity of LRRK2, indicating that the GTPase activity of ROC domain regulates the kinase activity (Ito, Okai et al. 2007). GTP hydrolysis is also required for normal kinase activity, and furthermore controls the LRRK2-induced inhibition of neurite outgrowth in neuron cell (Biosa, Trancikova et al. 2013).

The COR domain is a constitutive tandem domain to ROC in all ROCO superfamily members (Bosgraaf and Van Haastert 2003) and interacts with the Roc domain. The kinase domain is a

mitogen-activated protein kinase kinase kinase (MAPKKK) (Marin 2006), which is proved to autophosphorylates LRRK2 in the Roc domain (T1343 and T1491) (Greggio, Taymans et al. 2009).

Some mutations in Roc or COR decreasing GTPase activity and mutations in kinase domain increasing kinase activity cause the same Parkinson symptoms, probably because they evoke similar downstream signaling changes (Cookson 2010). Evidence showed that the kinase activity of LRRK2 is stimulated upon ROC binding GTP (Ito, Okai et al. 2007) and mutations preventing GTP binding decreases kinase activity (Ito, Okai et al. 2007, Lewis, Greggio et al. 2007) . Mutated LRRK2 with reduced kinase activity also reduces neuron toxicity (Smith, Pei et al. 2006).

Overall, the Roc domain and the kinase domain work together with autoregulatory interdependence. Although the mechanism how these domains of LRRK2 regulate each other and effect downstream signal pathway and why it becomes pathogenic remains unclear, a hypothesis is that LRRK2 actively mediates downstream signaling pathway while binding GTP, perhaps by increasing the affinity to some unknown partner (Cookson 2010). Roc GTPase activity might be decreased upon phosphorylation. Thus, Roc-bound GTP might stimulate downstream pathways in a feed-forward loop, and the GDP- or GTP-bound states of Roc might indicate whether LRRK2 is pathogenic. This hypothesis is supported by some studies showing kinase-dead or GTP-binding-deficient mutants are less toxic than wild type (Greggio, Jain et al. 2006).

The Structure of ROC domain

The high resolution structure of Roc domain in complex with GDP and Mg^{2+} was determined previously in our lab (Deng, Lewis et al. 2008). Roc appeared to be a dimer in the crystal as a functional unit with extensive domain swapping in a head-to-tail fashion. The GDP Mg^{2+} binding site involves key residues from both protomers. Two important residues, R1441 and I1371, whose

mutations are pathologically related and increase kinase activity, are located at the dimer interface. It is possible that the mutations (R1441C and I1371V) would destabilize the dimer as a functional unit and further decrease the kinase activity. There are five α-helices and six β-strands in a monomer, and the loop between the second α-helix and the second β-strand (G2 loop, also known as switch I) is disordered (indicated in Figure 3-3), as well as the one between the second β-strand and the third β-strand. The switch I loop was shown to have significant conformational change upon binding GTP and GDP throughout GTPase families (Paduch, Jelen et al. 2001).

Switch I and switch II regions are conserved loops throughout the Ras GTPase family. They locate around the active site and facilitate GTP hydrolysis (Vetter and Wittinghofer 2001). It is possible that in ROC, switch I and switch II undergo conformational changes upon binding GTP, and during the GTP/GDP binding cycle, the conformational changes might be delivered to the kinase domain through the COR domain.

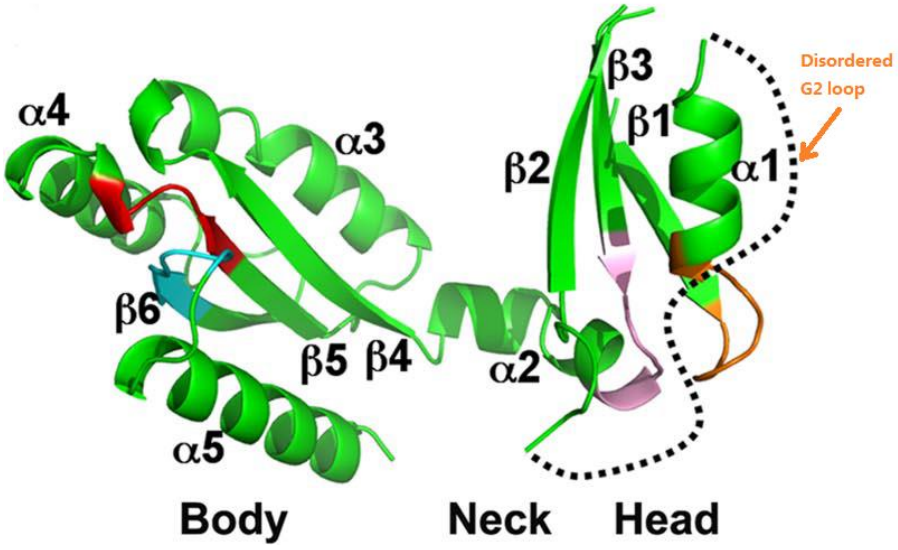


Figure 3-3: The crystal structure of a monomer of Roc. The 5 α-helices and 6 β-strands are marked. The disordered G2 loop (switch I region) is shown by a black dash line.

In summary, LRRK2 is a protein that might serve as a crucial target for the treatment of Parkinson's disease, and the GDP- or the GTP-bound states of the Roc domain might regulate the kinase domain. The structure of the inactivated state (GDP-bound) has already been solved, and we are interested in solving the activated state, for the comparison of inactivated state and activated state provides information regarding the mechanism of Roc GTP hydrolysis and signaling.

4. Prokaryotic Roc-Cor Domain

Studies on *Clorobium tepidum* LRRK2 showed that ligand binding induced significant change in the Roc-Cor tandem. The result of trypsin digestion of *C. tepidum* LRRK2 indicated that the GppNp (a nonhydrolysable GTP analog) binding Roc-Cor tandem was quite resistant to digestion, whereas the nucleotide-free Roc-Cor tandem was brokendown to the Cor domain.

The crystal structure of the Cor domain was solved. It contains two subdomains connected by a long linker. The N-terminal subdomain was mainly made up of α-helices, while the C-terminal subdomain contains seven-strand antiparallel β-sheets and a β-hairpin motif. Mutations on the dimer interface and gel filtration analysis demonstrated that two Cor molecules dimerize through their C terminal subdomain.

The nucleotide-free *C. tepidum* Roc-Cor mutant crystallized and the structure is shown in Figure III-4. The Cor domains dimerize through its C-terminal subdomain, while the two attached Roc domains are positioned apart. Only one Roc was visible in an asymmetric unit (RocA). The reason why the other one was absent might be that Roc domain is very mobile in the crystal.

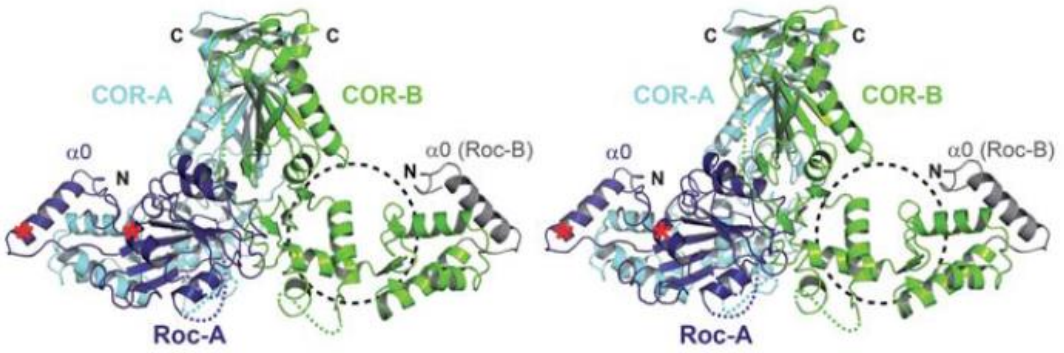


Figure 3-4: The crystal structure of *C. tepidum* Roc-Cor tandem.

Human LRRK2 and *C. tepidum* LRRK2 appear to be different in many aspects. The *C.tepidum* Roco has a much higher affinity for GTP than GDP (Gotthardt, Weyand et al. 2008), but human Roc has higher affinity for GDP (Li, Dunn et al. 2009). Also, *C. tepidum* Roc-COR dimerizes through its C terminal subdomain of Cor (Gotthardt, Weyand et al. 2008), while human Roc dimerizes through domain swapping (Deng, Lewis et al. 2008). Furthermore, the Roc domain in the *C. tepidum* Roco is a monomer displaying canonical Ras GTPase (Gotthardt, Weyand et al. 2008), but the human Roc domain is dimer adopting head-to-tail fashion, with catalytic site contributed by both protomers of dimer (Deng, Lewis et al. 2008).So our goal is to solve the structures of domains in human LRRK2.

With a lot of structure and function of LRRK2 remaining unknown, we propose to determine the structures of activated human Roc and Roc-Cor tandem domains. The superimposed GDP and GTP bound Roc domain will provide some detailed understanding of the structural change in the Roc, and the Roc-Cor tandem structure, and will provide some information about how GTPase activity change passes to kinase domain through Cor domain as a molecular hinge. Together they may contribute to the understanding of the structure and function of LRRK2 and its role in the development of PD.

B. Results

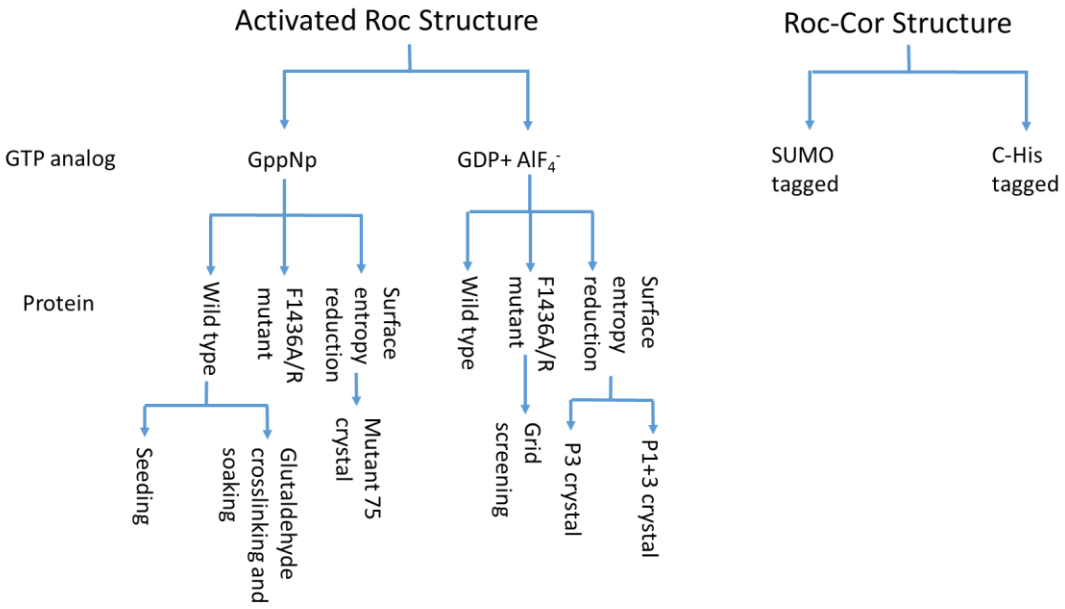


Figure 3-5: Projects flow chart.

1. Roc Plus GTP Analog Crystallization

a. Cloning and Purification

Cloning and purification of Roc or its mutants were carried out as described in II.B. GDP or GTP analog was supplied from the 1st nickel elution. The samples collected along the purification of wild type Roc in complex with GDP were analyzed with SDS-PAGE and shown in Figure III-6.

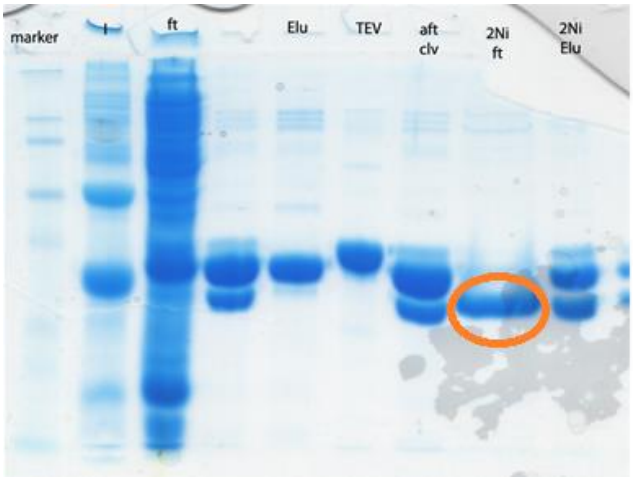


Figure 3-6: SDS-PAGE gel of samples along the Roc purification. The lanes from left to right indicate the marker, induced whole cell, cell lysate flown through Ni resin, TEV cleaved sample, Roc with N-terminal His tag eluted from Ni resin, TEV, TEV cleaved Roc, final Roc without His tag, uncleaved material eluted from Ni resin.

b. Stability of Roc in complex with ligands

Thermo-denaturation assay (thermo-melt) measures the temperature required for the folded protein to unfold (T_m) and thus indicates the stability of protein. The method is based on the character of fluorescence dye, such as SYPRO Orange fluorescence dye (Invitrogen) we used. SYPRO Orange is excited by light around wavelength 490 nm and emits fluorescence of 610 nm. When protein unfolds and hydrophobic regions of the protein are exposed to solvent, the fluorescence dye binds to them and becomes fluorescent, and Real Time PCR instrument was used to record fluorescence counts. Finally, the first derivative of fluorescence counts is calculated in Excel, and the temperature corresponding to the fastest change is the melting temperature (T_m).

The differential scanning fluorimetry is widely used to screen compound versus protein, for the stability of complex indicates the binding affinity between the two.

The result of Thermo melt of Roc with various ligands is shown in the figure below (Figure III-7). Roc is not stable without any ligand and different ligands have different affinity to it, all of which increases the stability of Roc.

In a recent paper, the affinity of Roc binding to GDP was measured with the K_D being 0.47 μM. A nonhydrolysable GTP analog GTPγS was measured to estimate the approximate binding affinity between Roc and GTP. The K_D is 7.85 μM (Liao, Wu et al. 2014). In combination with our T_m data, the binding affinity of all the ligands should be within the μM range.

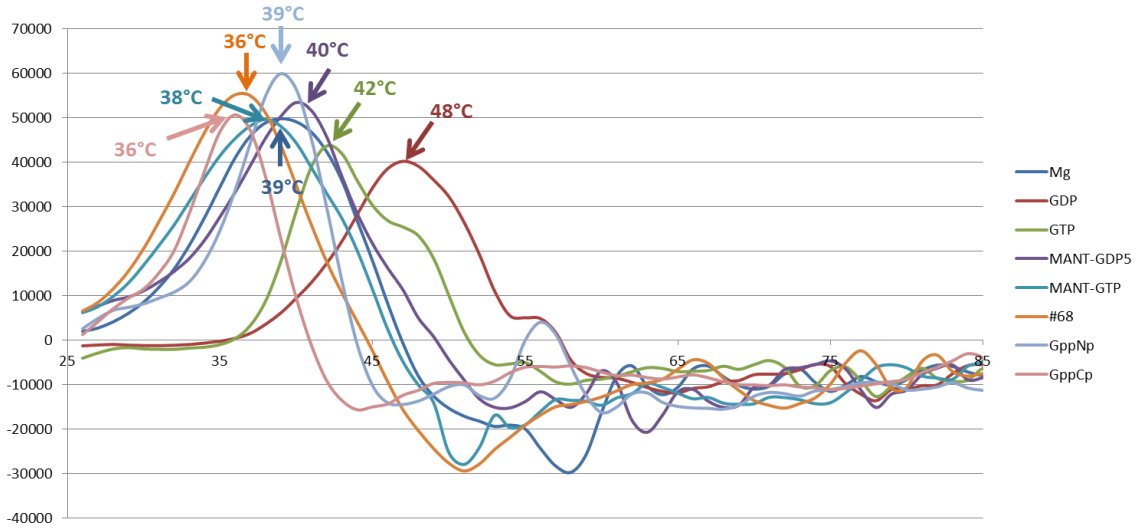


Figure 3-7: The first derivative of fluorescence reading of Roc in complex with various ligands. Melting temperatures are indicated above.

Ligand	T_m
Mg	39 ℃
GDP	48 ℃
GTP	42 ℃
MANT-GDP	40 ℃
MANT-GTP	38 ℃

68	36 ℃
GppNp	39 ℃
GppCp	36 ℃

Table 1: Melting Temperature of Roc in complex with various ligands.

c. Selecting proper ligand for crystallization

Roc is an authentic GTPase and hydrolyzes GTP. When GTP was used in Roc purification and crystallization, GDP was observed in the structure (Deng, Lewis et al. 2008). So instead of GTP, we used two types of analogues: GDP plus aluminum tetrafluoride (AlF_4^-) and non-hydrolysable GTP.

AlF_4^- has similar atomic size with PO_4^{3-}, and it appears to activate G protein by binding the Gα subunit of heterotrimeric G protein in its GTP binding state, mimicking the ɤ phosphate of GTP (Bigay, Deterre et al. 1985). In the similar way, AlF_4^- is also commonly used in complex with GDP as a GTP analogue in crystallography (Tesmer, Berman et al.). Here we apply AlF_4 and GDP together to Roc protein in order to simulate the transition state of Roc hydrolyzing GTP.

We used two types of non-hydrolysable GTP analogs, GppNp (Sigma) and GppCp (Jena Bioscience).

d. Roc in complex with GppNp

Roc in complex with GppNp was purified as described in II.1.2. The complex was concentrated to 8.4 mg/ml and some needle cluster shaped crystals were observed after broad screening.

i. Seeding

Seeding is a powerful method for crystal growth optimization. Crystals unqualified for data collection, such as small or twinned crystals, can be grinded finely in reservoir buffer and used as micro seeds. These seeds are introduced into pre-equilibrated supersaturated protein and reservoir buffer solution. In this way, we can control not only the number of nucleation sites in solution, but also the equilibration time before seed introduction (Bergfors 2007). Seeding has been proven to increase size of crystal reproducibility in many cases, but Roc-GppNp failed to be improved.

e. Glutaldehyde cross-linking and soaking

Roc with GDP crystalized in 2.1 M malic acid, but the structure showed that protein did not have ligand inside. It is understandable because the crystal grew in high salt condition. Strong ion strength would compete with the ligand binding and extrude it from the binding site. However, the binding pocket was open and empty, providing us the opportunity to soak the ligand back into the pocket.

During soaking, we stepwisely exchanged the crystal containing reservoir buffer from 2.1 M malic acid to lower ion strength, for ligand cannot soak in under the high salt condition. The potential differences between inside and outside of crystal caused the crystal melting or crashing, so we used glutaraldehyde to crosslink crystal.

Glutaraldehyde undertakes polymerization in alkaline condition and predominantly links the ε-amino groups of lysine residues. It fixes the crystal lattice and facilitates the soaking of ligand when the potential in and out of crystal lattice is different. Crystals were transferred to reaction buffer containing 0.1 M HEPES, 0.12 M NaCl, 25% ethylene glycol and different concentrations of glutaraldehyde for cross-linking for various length of time, and the reaction was stopped by washing crystal twice with reservoir buffer to remove residual glutaraldehyde. (Wine, Cohen-Hadar et al. 2007)

The crystal was then mounted in an in-house x-ray system to determine its diffraction ability. The original crystal diffracted 2.8 Å in synchrotron x-ray source. Different concentration of glutaldehyde and various length of reaction time was tested. Using glutaraldehyde concentration higher than 1% or reaction time longer than 5 minutes caused the crystal to loose diffraction, so we put crystal in 1% glutaldehyde for 5 minutes and then transferred the crystal to buffer containing GDP plus AlF_4^- or GppNp, while lowering the malic acid concentration and increasing the PEG 8000 concentration stepwisely. The soaking lasted for 10 to 30 minutes to allow the ligand entering into crystal.

This manipulated crystal diffracted 3.8 Å in synchrotron x-ray source, showed a $P2_12_12_1$ space group. There was no density in the GTP binding site.

f. Site-direct Mutagenesis

i. F1436 mutation

According to the known structure, phenylalanine 1436 occupies the space of gamma phosphate, so we mutated this phenylalanine to alanine.

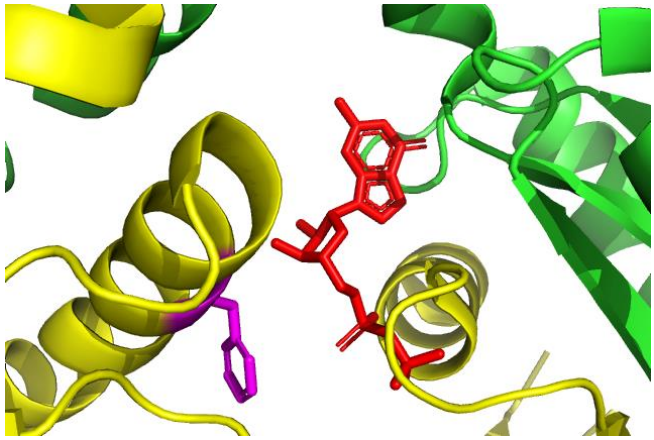


Figure 3-8: The illustration of the F1436 (purple) and GDP (red) in Roc GDP structure.
In order to vacate enough space for ɤ phosphate of GTP, F1436 was mutated to Alanine or Arginine.

Site-directed mutagenesis was performed as described in IV.2.3. Forward and reverse primers containing desired mutation (F1436A or F1436 R) were ordered from Invitrogen. Two pairs of primers, 1333 forward and 1436 reverse, 1436 forward and 1516 reverse, were extended using original plasmid as template by PCR, generating 1333-1436 and 1436-1516 fragments accordingly.

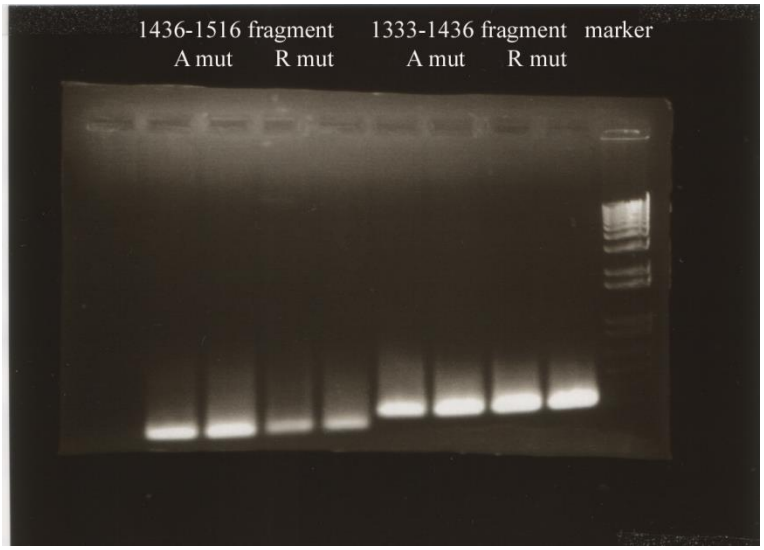


Figure 3-9: The DNA electrophoresis gel of mutated fragments (1333-1436 and 1436-1516).

The two overlapping fragments were elongated by PCR using 1333 forward primer and 1516 reverse primer. Then mutation containing fragments were digested with restriction enzymes and ligated to vector.

Insertion was confirmed via colony PCR and sequencing as described in III.B.2.

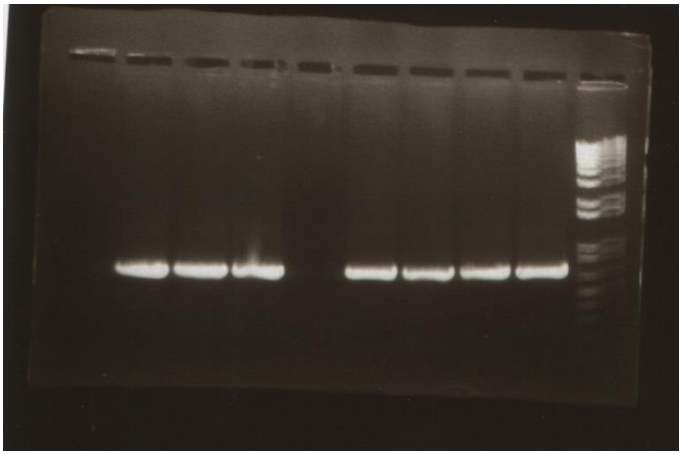


Figure 3-10: The DNA electrophoresis gel of colony PCR products. Left to right: 3 colonies of F1436R mutation, 4 colonies of F1436A mutation, marker.

F1436A and F1436R were expressed in 2L cells each, and purified in complex with GDP.

F1436A F1436R

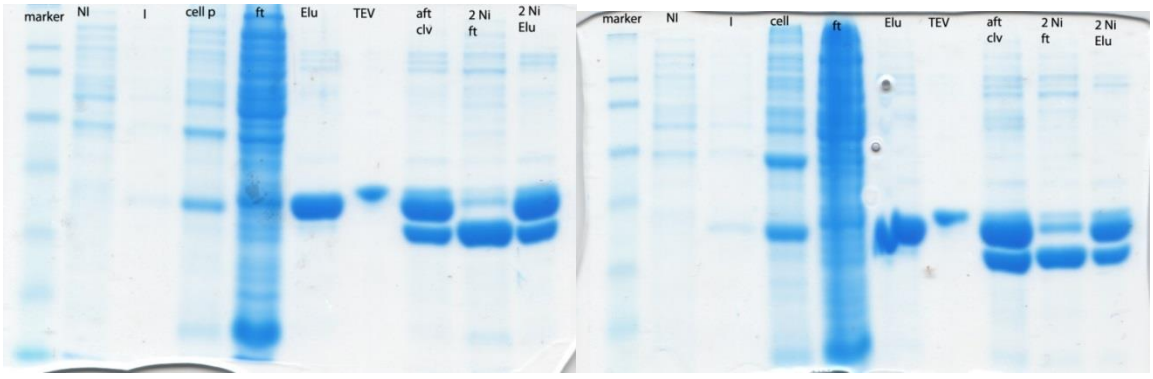


Figure 3-11: SDS-PAGE gels of samples collected along F1436A (left gel) and F1436R (right gel) purification. From left to right: marker, non-induced whole cell, induced whole cell, cell pallet, flow-through from nickel resin, elution from nickel column, TEV, mixture after TEV digestion, flow through from the second nickel resin (concentrated for crystallization setting up), elution from the second nickel resin.

1:10 ratio of fresh $AlCl_3$ and NaF were put into protein solution before crystallization screening. Some needle crystal hints showed up, but they were too small for data collection.

Grid screening optimization

1436A mutant crystallized in 0.2 M lithium sulfate, O.1 M Bis-Tris, pH 6.5, 25% PEG 3350, so grid screening buffers varying pH from 5.5 to 7.0 as well as PEG 3350 concentration from 20% to 35% was prepared and used as reservoir buffer.

After overnight incubation, needle shaped crystal grew from pH 6.5, 30% PEG 3350, pH 7, 30% PEG 3350 and pH7, 35% PEG 3350. But their size did not improve after runs of optimization.

ii. Surface entropy reduction

Wild type ROC was crystallized with GDP, but we failed to obtain the crystal with non-hydrolysable GTP or with GDP and AlF_4^-. Because surface glutamic acid and lysine might prevent crystallization, due to their long flexible side chains that prevent favorable lattice contacts (Derewenda and Vekilov 2006), we mutated three patches of glutamic acid or lysine (shown in Figure III-17) to alanine as an approach to reduce surface entropy, which could potentially promote new crystal lattice formation.

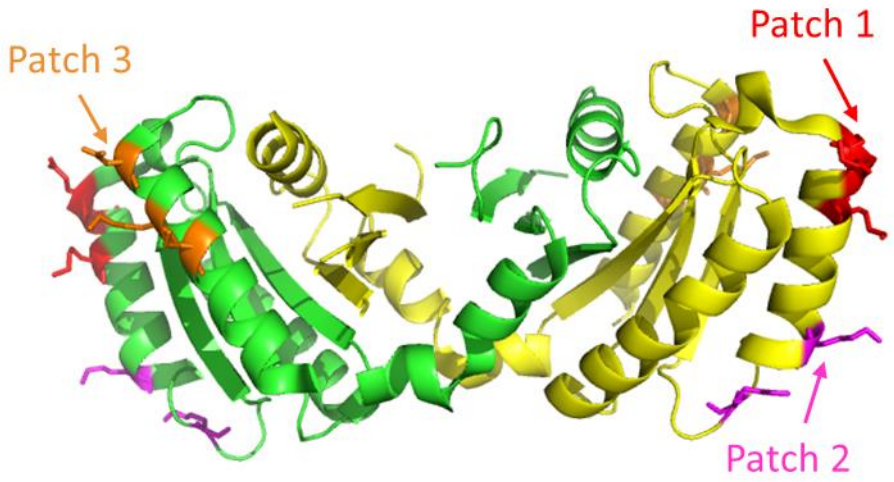


Figure 3-12: The illustration of 3 patches of surface mutation. Patch 1: 1459E/A, 1460K/A, 1463 K/A; Patch 2: 1471K/A, 1476K/A; Patch 3: 1495 D/A, 1499K/A, 1502 K/A

Based on the F1436A mutant, we constructed P1, P2, P3 and P1+2, P1+3, P2+3 mutants containing corresponding patches of mutations. Site-directed mutagenesis and purification were performed as described above.

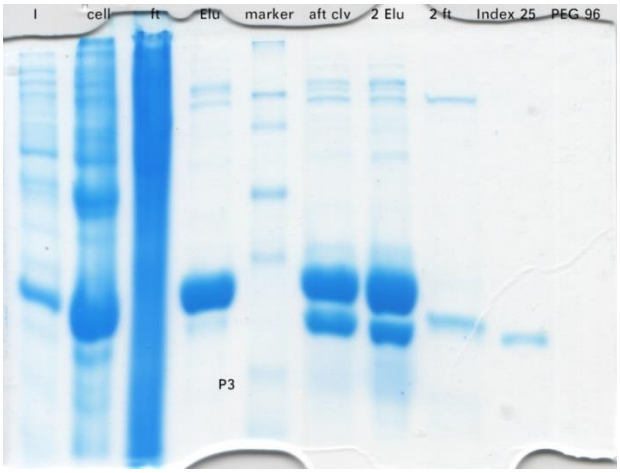


Figure 3-13: SDS-PAGE gel of samples collected along P3 purification. From left to right: induced whole cell, cell pallet, flow-through from nickel resin, elution from nickel resin, marker, mixture after TEV digestion, elution from the second nickel resin, flow through from the second nickel resin (concentrated for crystallization setting up).

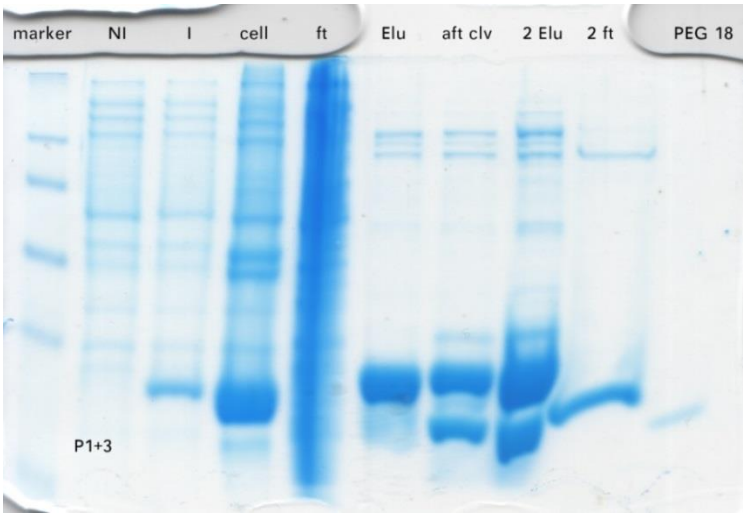


Figure 3-14: SDS-PAGE gel of samples collected along P1+3 purification. From left to right: marker, non-induced whole cell, induced whole cell, cell pallet, flow-through from nickel resin, elution from nickel resin, mixture after TEV digestion, elution from the second nickel resin, flow through from the second nickel resin (concentrated for crystallization setting up.)

g. We have obtained two different crystal forms of Roc with AlF_4 in the crystallization condition.

Previous GDP-bound Roc has thin rhombus plate morphology, but with the presence of AlF_4^-, P1+3 mutant formed hexagon plate crystals and P3 mutant formed di-pentagonal pyramid crystals.

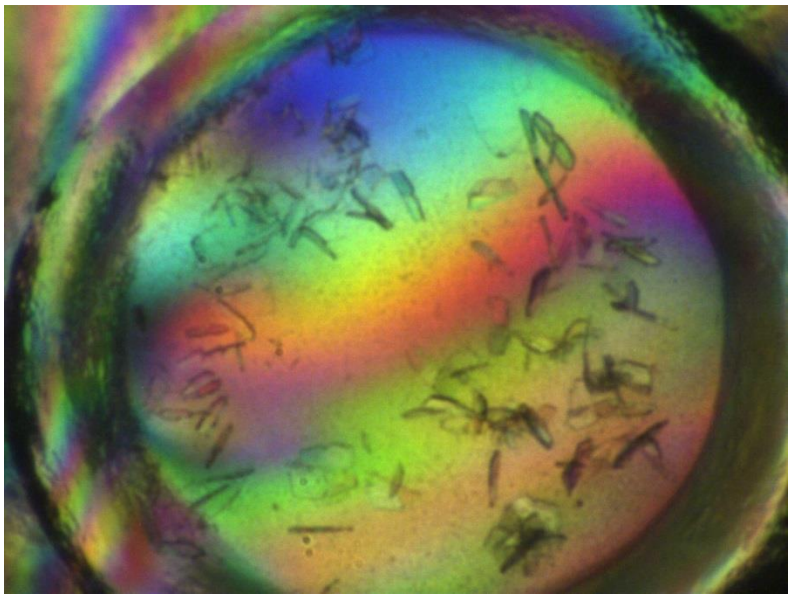


Figure 3-15: P1+3 hexagon crystals. Space group was C2.

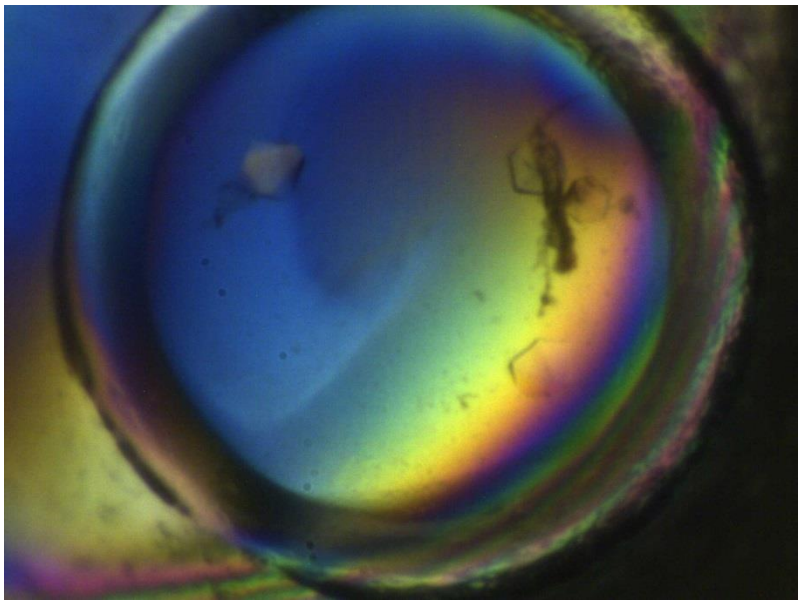


Figure 3-16: P3 di-pentagonal pyramid crystals. Space group is $P3_2$.

Data collection and initial processing is described as in II.5. The P1+3 hexagon crystal diffracted to 1.8 Å but the di-pentagonal crystal only diffracted to 3.6 Å.

Since a highly similar structure, is already present, we applied molecular replacement to initial data using Roc GDP complex structure as the model and obtain the initial phasing by using the Phaser program (Read 2001) and further modeling was carried using COOT (Emsley and Cowtan 2004). Finally structure was refined using REFMAC (Murshudov, Vagin et al. 1997). In the structure P1+3, no AlF_4^- was observed.

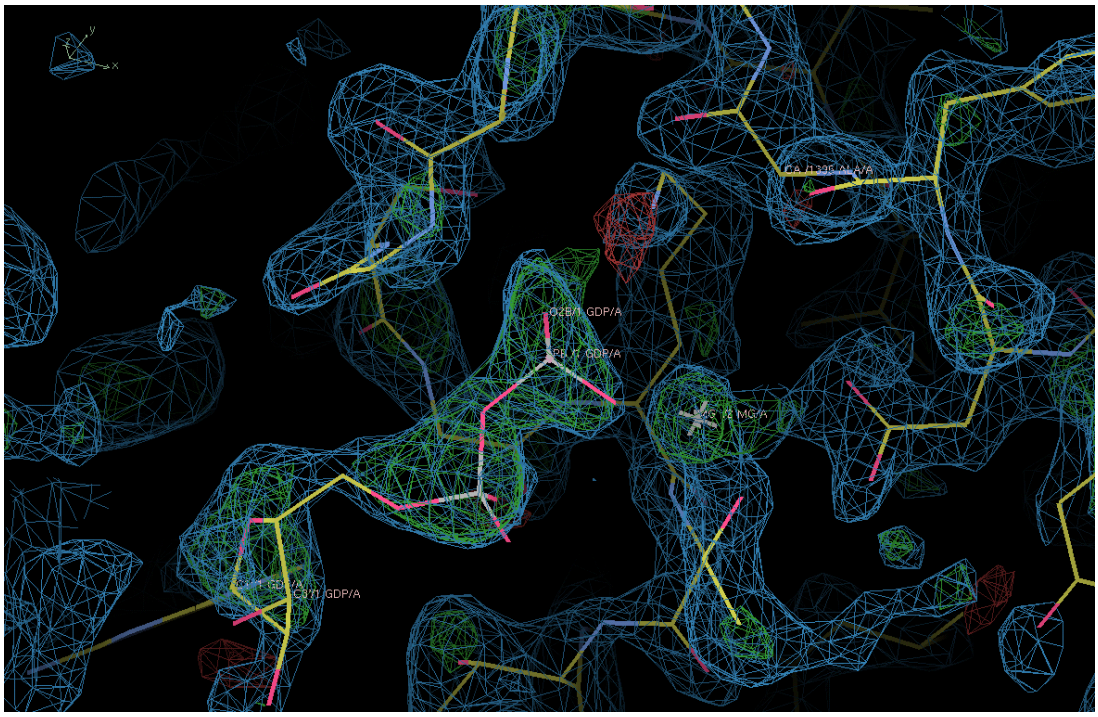


Figure 3-17: The electron density map of binding pocked of P1+3 in complex with GDP AlF_4^- crystal. We can observe the electron density for GDP and Mg^{2+} but no AlF_4^- . Drawn by coot (Emsley and Cowtan 2004).

Table 1 Roc P1+3-GDP-AlF_4 crystal

Statistics of data collection.

Values in parentheses are for the highest resolution shell.

Data collection	
Beamline	19-ID, APS
Wavelength (Å)	0.97918
Space group	C2
Unit-Cell parameters	a = 70.0 Å, b = 52.9 Å,c = 106.8 Å, α=90.0°, β=108.2°, γ=90.0°
Resolution (Å)	50 – 1.8
Total reflections	166,596
Unique reflections	35,627
R_{sym} (%)†	12.5 (81.9)
Redundancy	4.7(3.6)
Completeness (%)	99.7 (98.4)
<I/σ(I)>	15.3 (1.7)

† $R_{sym} = \Sigma\ |I_{obs} - I_{avg}|/\ \Sigma\ I_{avg}$.

h. We obtained crystal of Roc in the presence of GppNp.

Roc mutant 75 was the combination of F1436R and P3 mutations. It was purified in complex with GppNp and concentrated to 13.2 mg/ml. It crystalized in the condition of 0.2 M sodium acetate, 0.1 M sodium cacodylate pH=7.25, 25% PEG 8000. The data collection, processing and structure determination was carried out as above. In the binding site, GDP was observed instead of GppNp.

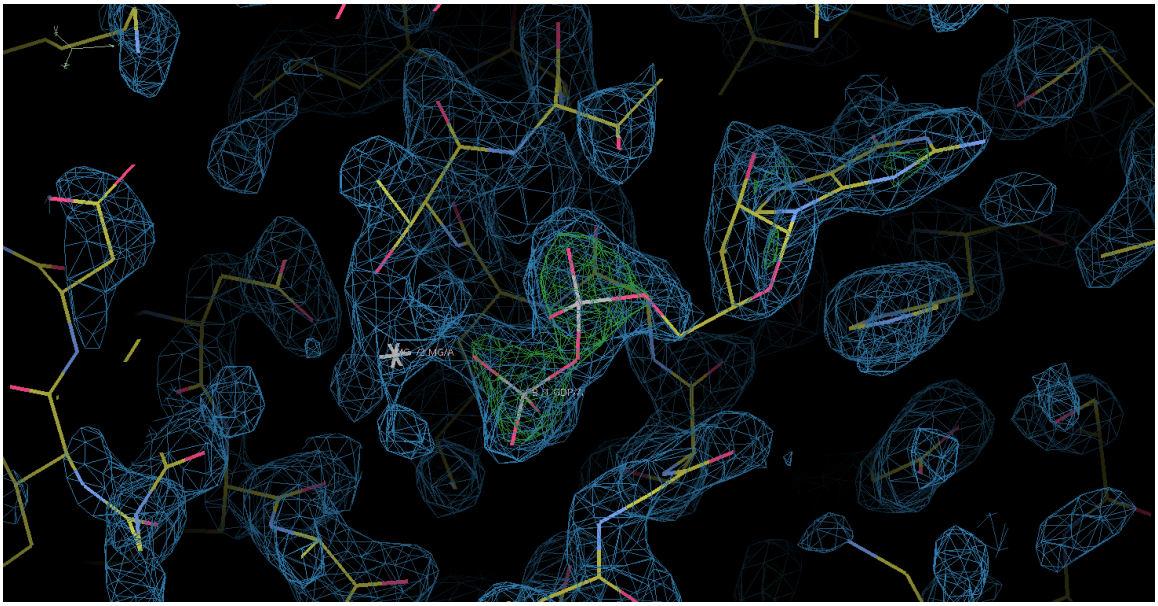


Figure 3-18: The electron density map of the binding pocket of mutant 75 in complex with GppNp crystal. We can observe the electron density for GDP and Mg^{2+} but no GppNp. Drawn by coot (Emsley and Cowtan 2004).

Table 2 Roc 75-GppNp crystal

Statistics of data collection.

Values in parentheses are for the highest resolution shell.

Data collection	
Beamline	19-ID, APS
Wavelength (Å)	0.97915
Space group	P1
Unit-Cell parameters	a = 44.1 Å, b = 44.0,c = 103.00 Å, α=97.3°, β=98.6°, γ=101.2°
Resolution (Å)	50 – 1.9
Total reflections	103,723
Unique reflections	53,050
R_{sym} (%)†	5.5 (20.9)
Redundancy	2.6(2.1)
Completeness (%)	91.0 (30.2)
<I/σ(I)>	21.5 (3.5)

† $R_{sym} = \Sigma |I_{obs} - I_{avg}| / \Sigma I_{avg}$.

2. Roc-COR domain crystallization

a. SUMO Tagged Roc-COR Cloning and Purification

i. Cloning

We selected four Roc-COR fragments of different length, amplified them with KOD enzyme (Invitrogen), then digested the fragments and pSUMO vector with the same restriction enzymes (*Nde* I and *BamH* I) at both ends.

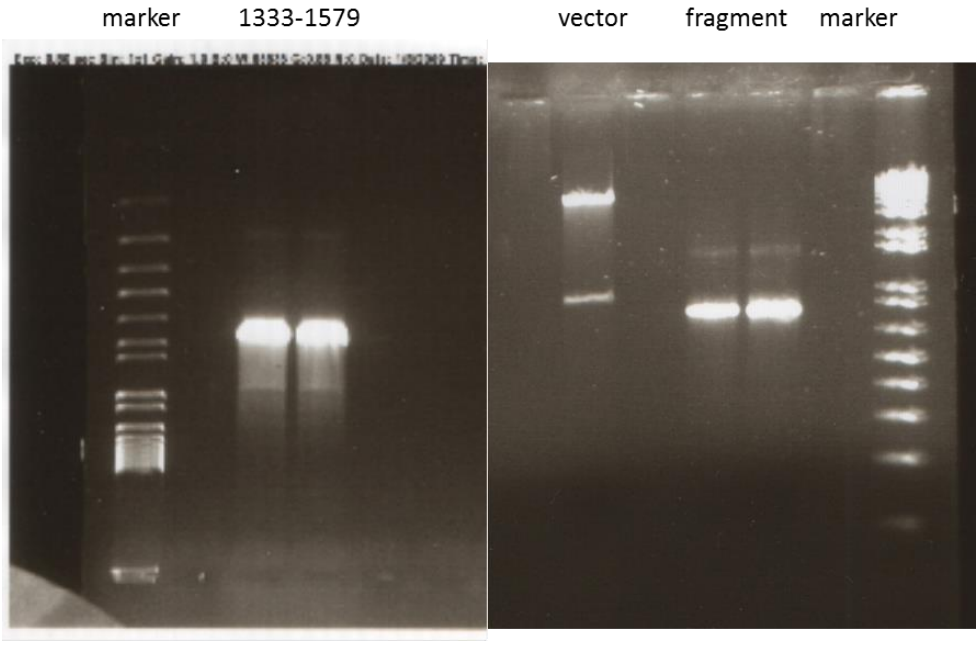


Figure 3-19: The DNA electrophoresis gels showing amplified Roc-COR fragment (1333-1579) (left gel) and the double digested vector and fragment (right gel).

Vector and amplified fragment were ligated by T4 ligase and thus transformed into *E. Coli* cell electronically.

ii. Insertion Confirmation

Colony PCR was performed first. After transformed competent cells growing on a Kanamycin resistant LB agar plate for overnight, a single colony was picked from the plate and grown in LB medium until dense. Plasmid was subsequently extracted from colony and used as template to amplify the DNA fragment between two primers (1333 or 1310 *Nde* I and 1579 or 1633 *BamH* I) at both ends. PCR product was analyzed with DNA gel electrophoresis to make sure the insertion should contain one and only one band with corresponding molecular weight (Figure 3-20).

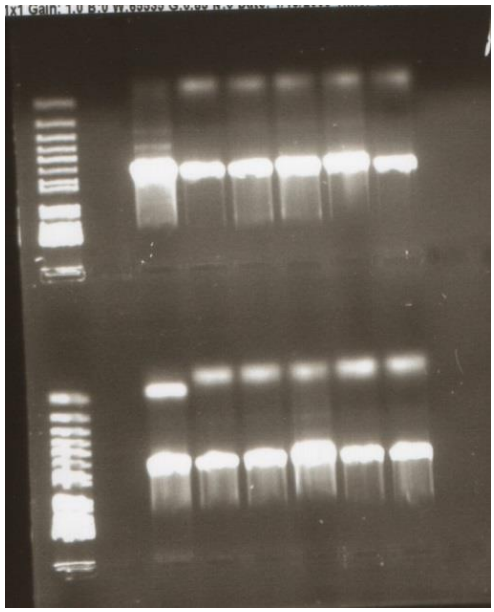


Figure 3-20: The DNA electrophoresis gel of colony PCR results of 4 constructs (3 clones from each construct). Upper left: 1333-1677. Upper right: 1310-1633. Lower left: 1333-1579. Lower right: 1333-1633.

The extracted plasmids was sent for sequencing with T7 forward and reverse primers and the result was aligned with correct sequence using BLASTP program (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) to confirm the insertions were correct.

iii. Small Scale Purification

For each construct of Roc-COR cloned, 50 ml cell culture was grown and lysed. Expressed proteins were purified through affinity tag with nickel resin. Samples were collected along purification and analyzed with SDS-PAGE gel.

Only 1333-1579 construct was able to be expressed and be purified in solution. (Figure 3-21)

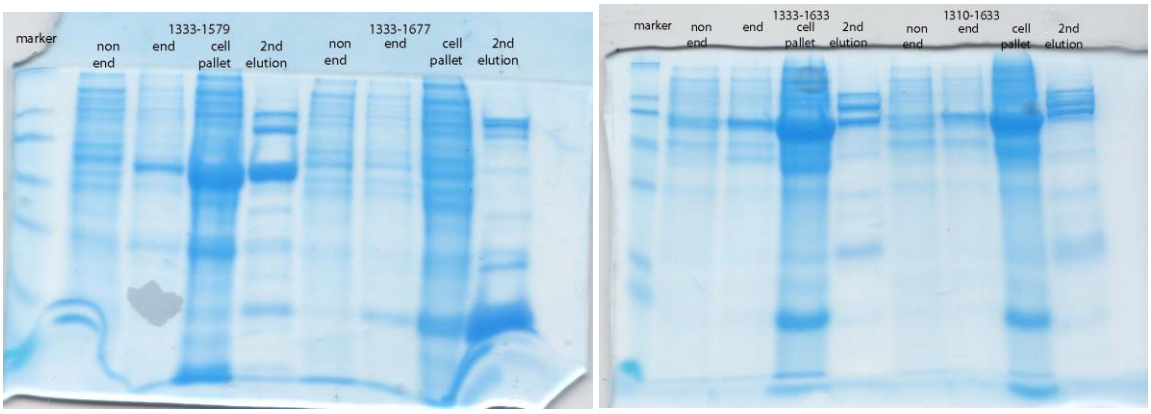


Figure 3-21: SDS-PAGE gels showing samples from small scale expression and purification of four constructs. From left to right: marker; non induced whole cell; induced whole cell; cell pallet; elution from nickel resin.

iv. Large Scale Purification

Protein from 2 liters of cell culture expression was purified through affinity tag, but the quality of protein was not satisfactory (Figure 3-22). It contained much molecular chaperone and SUMO.

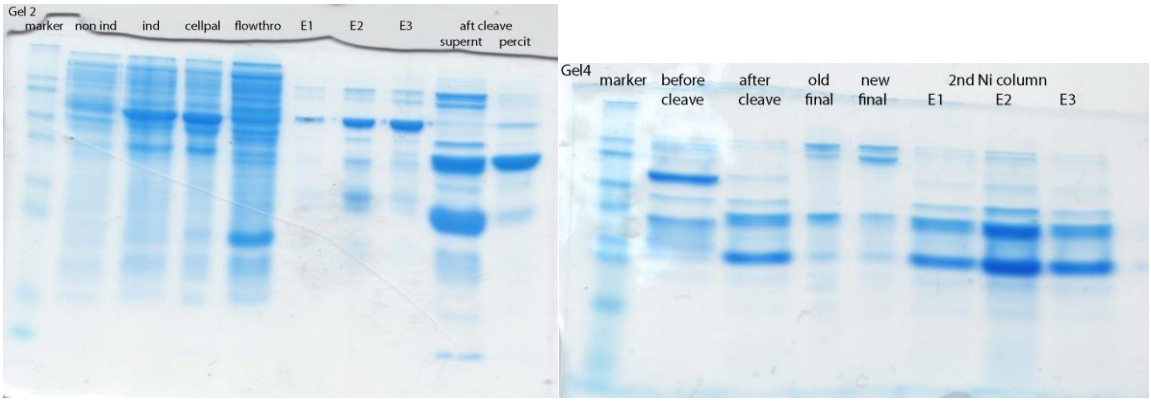


Figure 3-22: The SDS-PAGE gels of samples showing two batches of 1333-1579 construct purification. Left gel (batch 1) from left to right: marker, non-induced whole cell, induced whole cell, cell pallet, flow through nickel resin, the 1st, 2nd and 3rd elution fraction from nickel resin, supernatant after TEV digestion, precipitation after TEV digestion. Right gel (batch 2): marker, mixture before TEV digestion, mixture after TEV digestion, the concentrated 2nd nickel purified protein in batch 1, the concentrated 2nd nickel purified protein in batch 2, the three fraction from 2nd nickel elution.

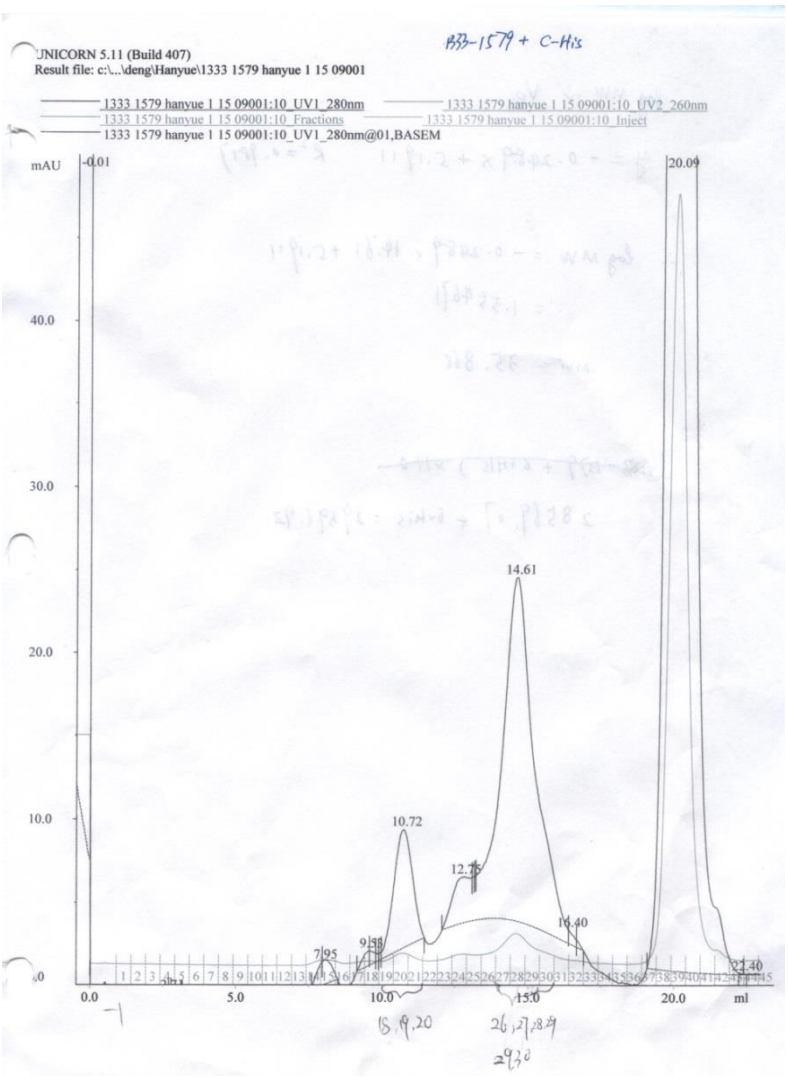
v. Crystallization

Multiple methods were applied in order to increase the purity of Roc-COR protein, such as detergent wash, GroEL/GroES chaperone assistance, but none of them was effective. The protein was used for crystallization screening anyway but no crystal was observed.

b. 1333-1579 C Terminal His-tagged Roc-COR Purification

The 1333-1579 fragment with permanent C-terminal His tag was cloned in the same way as described above. It could be expressed and purified, but the sample contained two adjacent bands which could not be separated even by size exclusion column. Western blot using anti-His antibody identified the upper band to be His-tag containing protein (Figure III-23) (Figure III-24).

Calculated from retention volume the molecular mass of particles in the peak was 47.7 kDa and its real molecular mass is 28.4 kDa. The protein might be heterodimer in solution containing one intact protomer and the other degraded slightly from the C terminus. The sample was used for crystallization screening anyway and no crystal was observed.



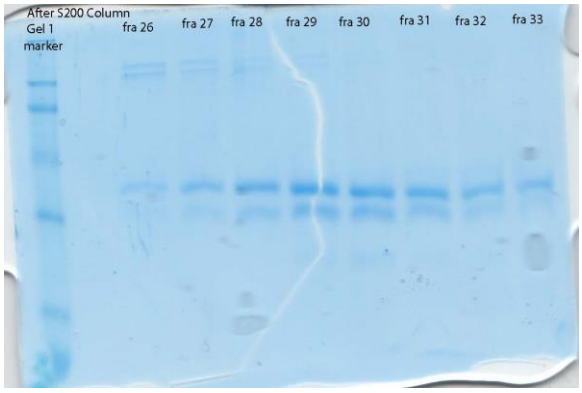


Figure 3-23: The size exclusion chromatograph of Roc-COR 1333-1579, the SDS-PAGE gel showing corresponding fractions.

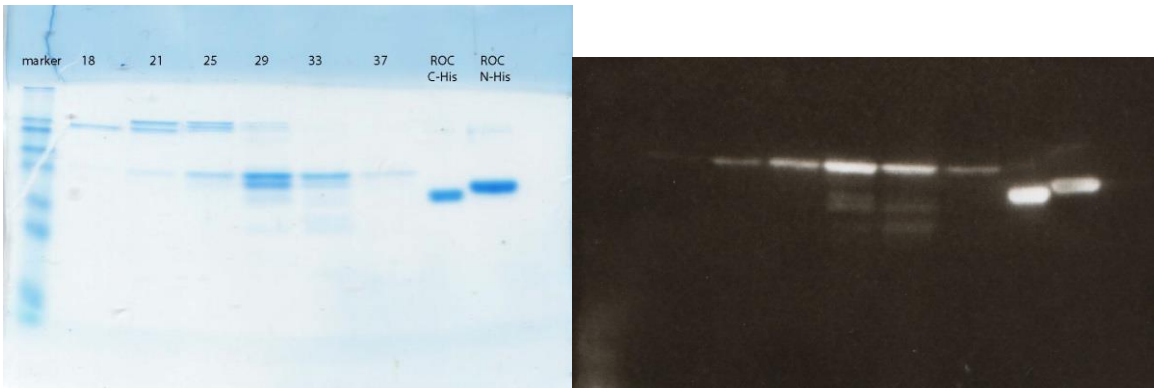


Figure 3-24: SDS-PAGE gel showing Roc-COR and parallel Western blot using anti-His antibody.

C. Discussion

1. The Information Obtained from Results

I obtained four Roc crystals in the presence of GTP analogs. One from glutaldehyde crosslinking and soaking had no GTP bound. In the structure of crystal of mutants in complex with GDP and AlF_4^-, only GDP was present. In the structure of crystal grew from Roc mutant purified in complex with GppNp, GDP was binding in the active site instead of GppNp.

2. Why I Didn't Obtain Roc-Cor or Roc GppNp Structure

The reason why we were not able to obtain the Roc-Cor tandem crystal might be that the linker between Roc domain and Cor domain is very flexible, thus the Roc-Cor tandem is mobile. In the prokaryotic Roc-Cor tandem structure, one Roc domain of the dimer was fixed by crystal lattice contact by neighboring molecule, while the other one was so mobile that it was invisible in the crystal structure (Gotthardt, Weyand et al. 2008). If human Roc-Cor tandem is elastic too, it will require large energy to overcome the energy barrier to crystallize.

The reason why only GDP was present in binding pocked instead of GppNp might be that GDP has higher affinity to Roc than GppNp (Liao, Wu et al. 2014), and the GDP was probably carried along from *E. coli.* Other literature about GTPase GppNp complex suggested thoroughly degrading GDP by alkaline phosphatase before putting GppNp into GTPase (Ye, Shima et al. 2005), but Roc is quite unstable without ligand bound so we cannot apply this method.

3. Surmises and Future Direction

To predict the structure of GppNp bound state of Roc, we need to take a look into Ras GTPase. It is consistent throughout Ras family that, the structural change of hydrolyzing GTP is described like loaded spring. The releasing of ɤ-phosphate causes switch I and switch II region to relax into GDP bound state (Vetter and Wittinghofer 2001).

The structures of Ras GppNp and GDP bound state superimpose well and their difference mainly exists in switch I and switch II region (Vetter and Wittinghofer 2001). When overlapping the cartoon structures of GDP and GppNp bound m-Ras (shown in Figure 3-25), the major change between them is present in the two switch regions. The switch II is disordered because it was mobile. The switch I is kicked away from active site upon binding GTP.

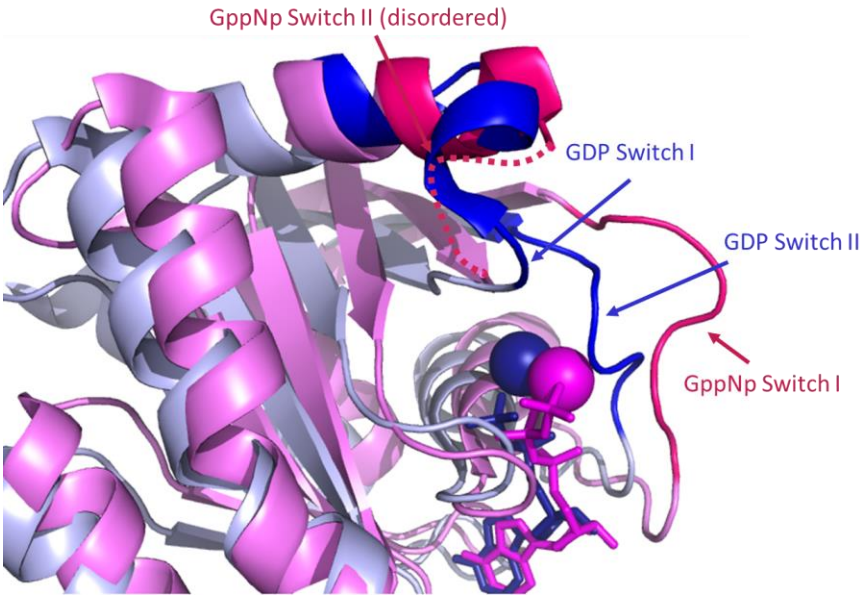


Figure 3-25: The comparison of GDP and GppNp bound m-Ras structure. The GDP bound m-Ras (gray) has its GDP and Mg^{2+} colored in dark blue and switch I and switch II region colored in blue. The GppNp bound m-Ras (pink) has its GppNp and Mg^{2+} colored in mauve and switch I and switch II region colored in hot pink. A part of switch II region is disordered (dotted line).

Since the GTP-bound state of Ras structures are relatively similar (Vetter and Wittinghofer 2001), we can reasonably surmise the GTP-bound Roc adopt the similar switch I and switch II conformational change.

The comparison between Roc and m-Ras also shed some light on the pathologically associated mutations in Roc. When Roc is aligned to m-Ras, its PD-related mutant R1441 corresponds to V113 in m-Ras. This Val interacts with α-helix connecting switch II region in GTP bound form. R1441 in Roc might adopt the similar function. Mutation in this position might cause the dislocation of switch II and further interfere with GTPase function.

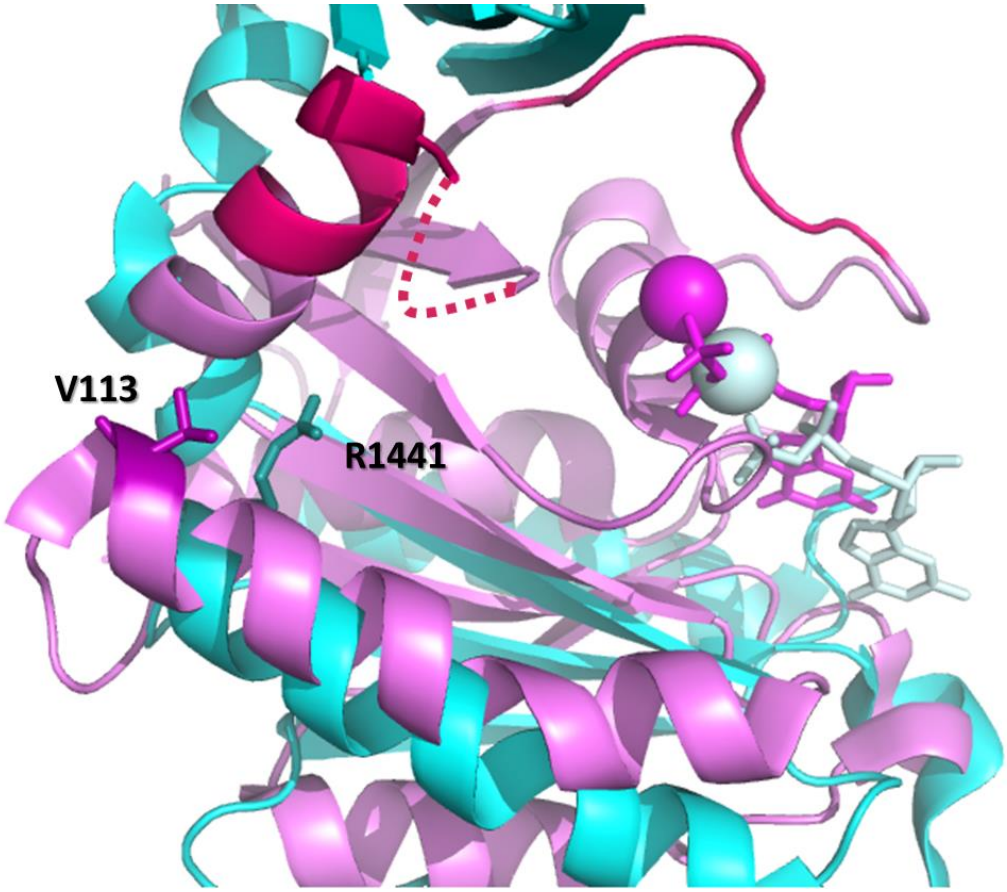


Figure 3-26: The superimposed Roc (teal) and m-Ras (pink). The R1441 and the V113 are highlighted in darker color and labeled.

The only available model for Roc-COR tandem is the prokaryotic *C. tepidum* Roco. Despite of the low similarity between bacterial and human LRRK2, we can still get some clue of pathological mutation Y1699C in Cor domain information from the prokaryotic structure. The conserved Y1699 is Y804 in *C. tepidum* Roco, which contacts with the α-helix connecting switch II (Gotthardt, Weyand et al. 2008). It is possible that Y1699 in human COR domain sits in the Roc/COR interface and interacts with Roc domain, the mutation of which decreases GTPase activity.

Roc domain is also a potential target for drug design. Inhibitors binding Roc domain might regulate kinase activity of LRRK2, and provide a new treatment for PD.

CHAPTER IV

A6

A. Introduction

A6 is a vaccinia virus protein expressed after virion gene replication and is packed tightly to the virion core. A6 is essential for poxvirus assembly and it was suggested to be involved in vaccinia virus morphogenesis by recruiting membrane as well as other important proteins from the infected cell to virion factories. Although the sequence of A6 is highly conserved through all vertebrate poxvirus (Meng, Embry et al. 2007), there are no known homologs outside the poxvirus family to suggest its function. Here we propose to reveal the structure of A6 by x-ray crystallography to obtain insights into its function. We hypothesize that certain functional domains of A6 share similar structures with a protein of known function. Therefore, searching through the structure database with the structure of A6 may provide us important clues about its function and mechanism.

1. Poxvirus

Poxvirus is a large, enveloped, double stranded DNA virus containing a complex genome varying from 130 to 230 kbp. Unlike other DNA viruses that release their genome upon contacting the nuclear membrane of infected cells and adopting the host DNA replication apparatus to replicate

themselves in the nucleus, poxvirus releases and replicates its DNA in the cytosome (Moss 2001). Thus, poxvirus requires genes encoding for proteins and enzymes used for transcription and translation (Moss 2013).

Research on poxvirus morphogenesis proteins draw our attention, because they might not only expand our knowledge on how poxvirus interacts with host cytosolic organelles and mediates its replication, but also shed light on the mechanism of cell regulation.

2. Replication Steps

As shown in Figure IV-1, the replication of poxvirus involves six steps.

It starts with the entry of virion particle into the infected cell (step 1), either upon binding unknown receptors or through endocytosis. Uncoated and exposed viral core particles go through early transcription and translation to produce the immunomodulatory proteins, enzymes, replication factors and translation factors (step 2). Then, the viral core particles are translocated to the outside of cell nucleus, and the nucleoprotein complex containing viral genome are released (step 3). The viral genome is replicated in certain domains of the infected cell called "**factories**" surrounded by ER derived cisternae where cell organelles are excluded. The genome is replicated as a concatemer and transcription factors are transcribed and translated, as well as enzymes and structural proteins (step 4). Then concatemer genome is resolved into linear double stranded DNA and packed with certain late viral proteins (A6 being one of them) into the immature virions (IVs) (step 5). How IVs mature into intracellular mature virions (IMV) remains unknown, with possible involvement of Golgi apparatus (step 6). After that IMVs are released in three ways. Most of virons are released through cell lysis and remain IMVs. Other virions bud through cell membrane and pick up viral envelope from plasma membrane of the inffected cell. Among them, cell-associated enveloped virions (CEVs) stay in cell until they are pushed by certain active tail

(probably microtubules) into contact with a second cell, and extracellular enveloped virions (EEVs) are simply released from the infected cell.

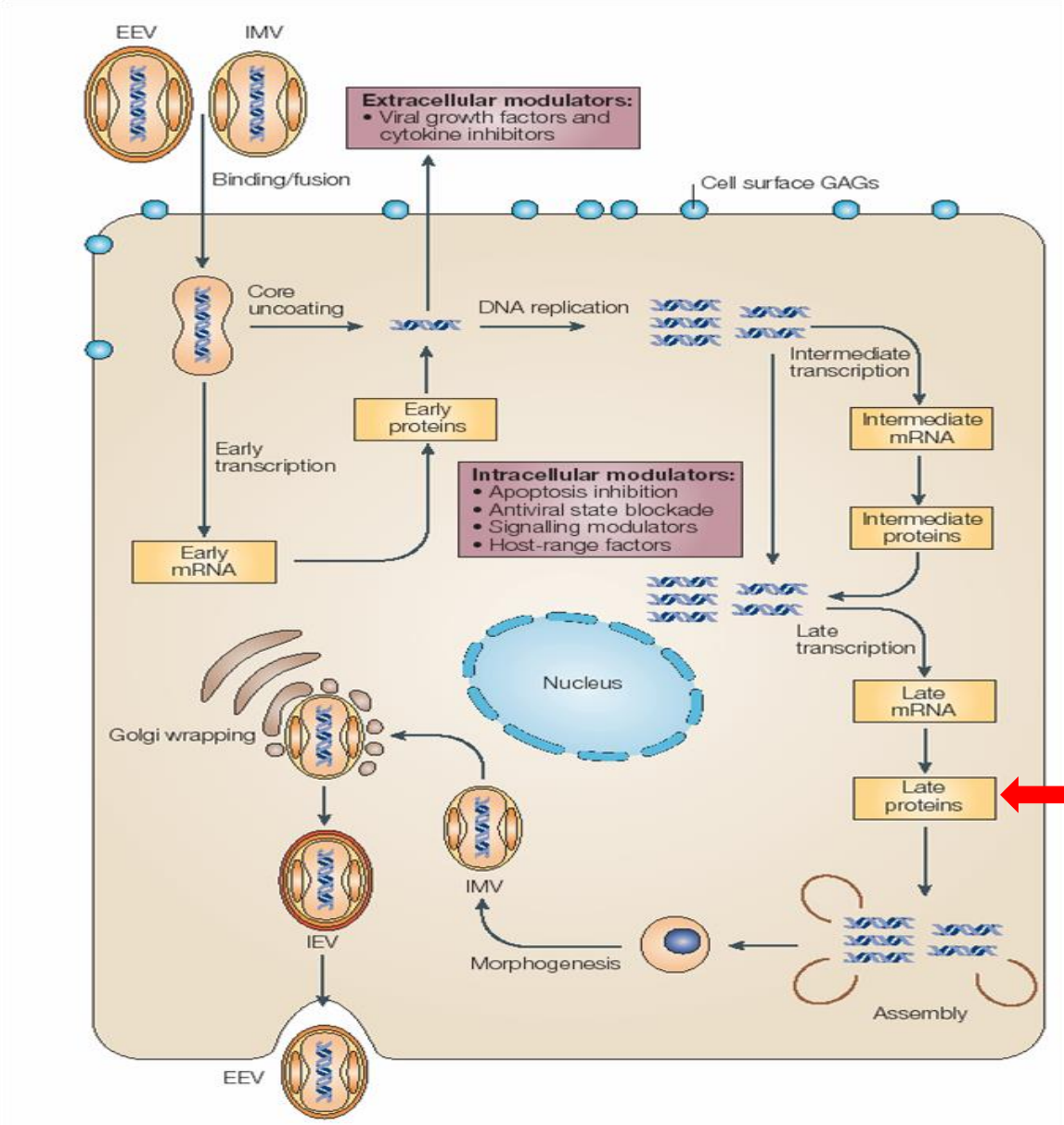


Figure 4-1: The illustration of poxvirus replication. A6 is one of the late proteins translated as indicated by the red arrow.(McFadden 2005)

3. Vaccinia Virus

Vaccinia virus is the most intensively studied member of poxvirus viruses. It has linear, double stranded DNA around 190 kbp. The large genome contains genes encoding proteins responsible for cytoplasmic viral DNA replication and transcription.

Unlike the lethal variola virus (small pox), the most notorious member of the poxvirus family, vaccinia virus causes only minor symptoms, while it can stimulate immunogenic response against the lethal smallpox. Therefore, a benign form of vaccinia virus is the major active component of the vaccine against small pox. With cytoplasmic replication site being the target for foreign gene introduction, the ability to infect many different cell lines, and its unique cytoplasmic transcriptional and translational apparatus, vaccinia virus can therefore serve as a vector carrying new genes to generate recombinant vaccines against other infectious diseases (Hruby 1990).

Under electron microscope, vaccinia viruses during replication go through morphological changes from double-layered crescents to closed-circle IVs, then to electron-dense nucleoid containing IVNs, and sequentially to membrane protein surrounded elliptic MVs. Lipid membrane wrapped viruses (WVs) are observed as final stage of viruses before released from host cell (Figure IV-2).

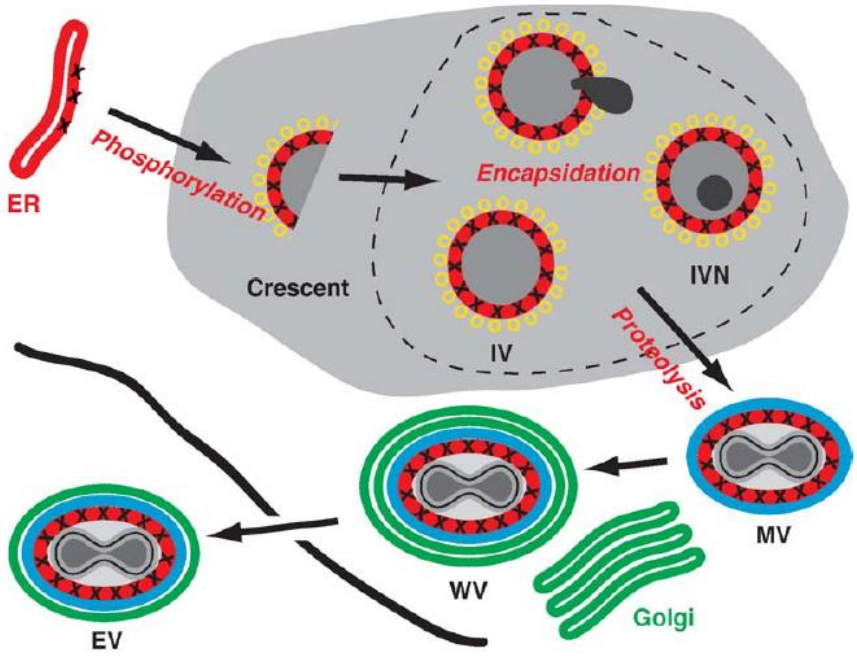


Figure 4-2: The illustration of poxvirus morphology development. (Condit, Moussatche et al. 2006). IV: immature virion. IVN: immature virion with nucleus. MV: mature virion. WV: wrapped virus. EV: extracellular virus.

4. The Importance of A6

Poxvirus genome contains 218 open reading frames, 91 of which are conserved through all vertebrate poxvirus species (Upton, Slack et al. 2003). Vaccinia virus A6 is encoded by one of those genes. A6 protein was identified by gel-free liquid chromatography and tandem mass spectroscopy of whole virions (Chung, Chen et al. 2006).

A6 contains 372 amino acids in total. Its molecular weight is around 43 kD and its calculated pI is 8.8. It has no homolog outside the poxvirus family, and no homologous structure is available to date.

A6 is expressed after viron replication and then packed tightly in the virion core. Although the detailed function of A6 is not known, vaccinia virus cannot mature with the absence of A6. Study

on epitope-tagged A6 and mutagenesis provided evidence of A6 being essential for the transition from IV to IVN and MV.

After mutating clusters of amino acids to Ala near C-terminus of A6, temperature-sensitive viruses with no or impaired replication ability at high temperature (40 ℃) were screened and selected. That loss of replication ability was due to the mutation of A6 was confirmed because the replication ability was restored by adding wild type A6. At the non-permissive temperature (40 ℃), two viron proteins (4a and 4b) were essentiallly prevented from maturation and remained as their precursors. Also, the lesion in A6 caused a block in virus morphogenesis and only crescents and a few IVs were observed in comparison to all stages (crescents, IVs, IVNs and MVs) were present at the permissive temperature 31 ℃ (Figure 4-3) (Meng, Embry et al. 2007).

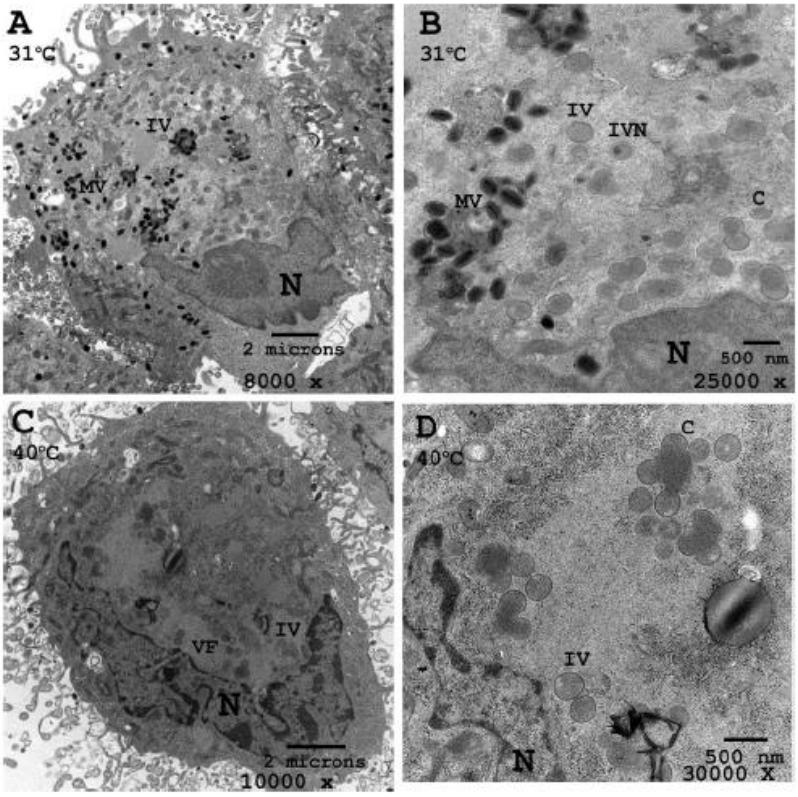


Figure 4-3: Absence of A6 block virus morphology maturation.

Virus containing with IPTG-inducible A6 was studied and A6 appears to be essential not only for early virus membrane recruitment, but also membrane protein localization. Transmission electron microscopy of cells inffect with recombinant viruses with A6 induced or repressed suggested that A6 is required for the membrane formation of virus crescent (Figure 4-4). Also, without A6, virion membrane proteins F9 and L1 cannot be detected by antibodies while majority of a virion core protein A10 remains as a precursor. A14 glycosylation ratio increases, same as observed when membrane biogenesis blocked (Figure 4-5). In absence of A6, results from immunofluorescence microscopy show that viron membrane proteins A13, A14, D8 and H3 change their localization from viron factories to the cytoplasm outside of the factories (Figure 4-6). Furthermore, viron membrane proteins not only co-localize but also co-sediment with ER protein when A6 expression is repressed, suggesting they associate with cell secretion apparatus. (Meng, Embry et al. 2012).

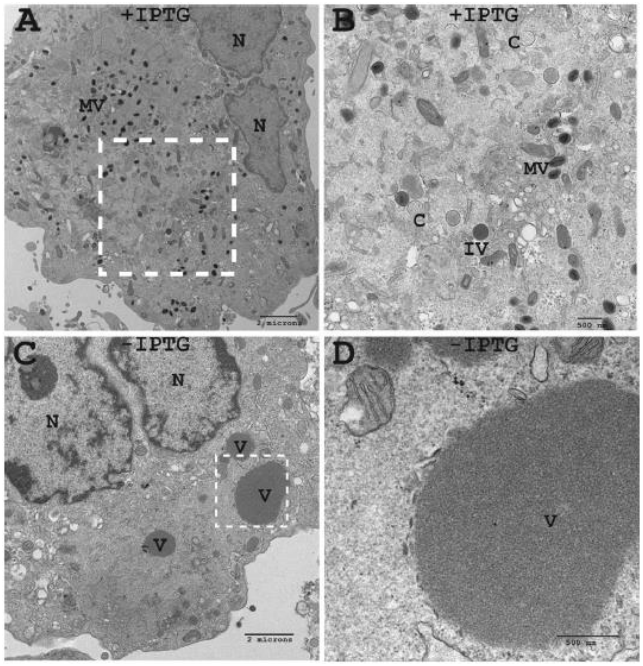


Figure 4-4: Absence of A6 block virus membrane formation. A and B are transmission electron microscopy images taken from cells infected with virus expressing A6, and C, D are from cells

infected with virus without A6. B, D are magnified from A, C correspondingly (Meng, Embry et al. 2012).

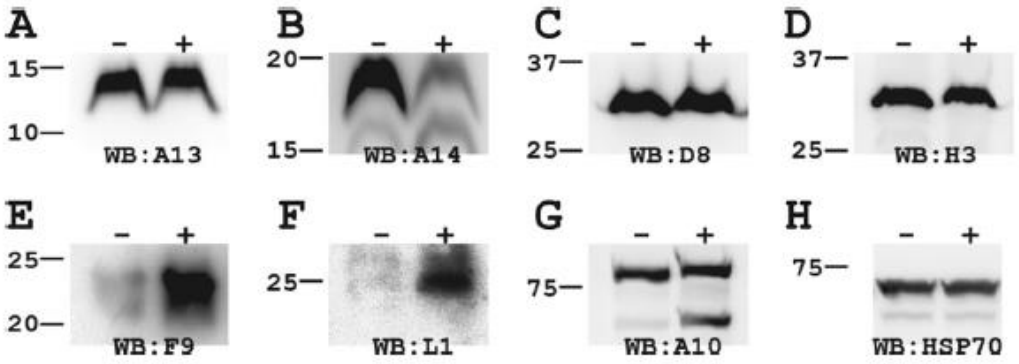


Figure 4-5: A6 effect the stability and post-translational modification of A14, F9, L1 and A10. The cells were infected by virus with (+) or without (-) IPTG. The western blot was performed on cell lysate using antibodies against each protein (Meng, Embry et al. 2012).

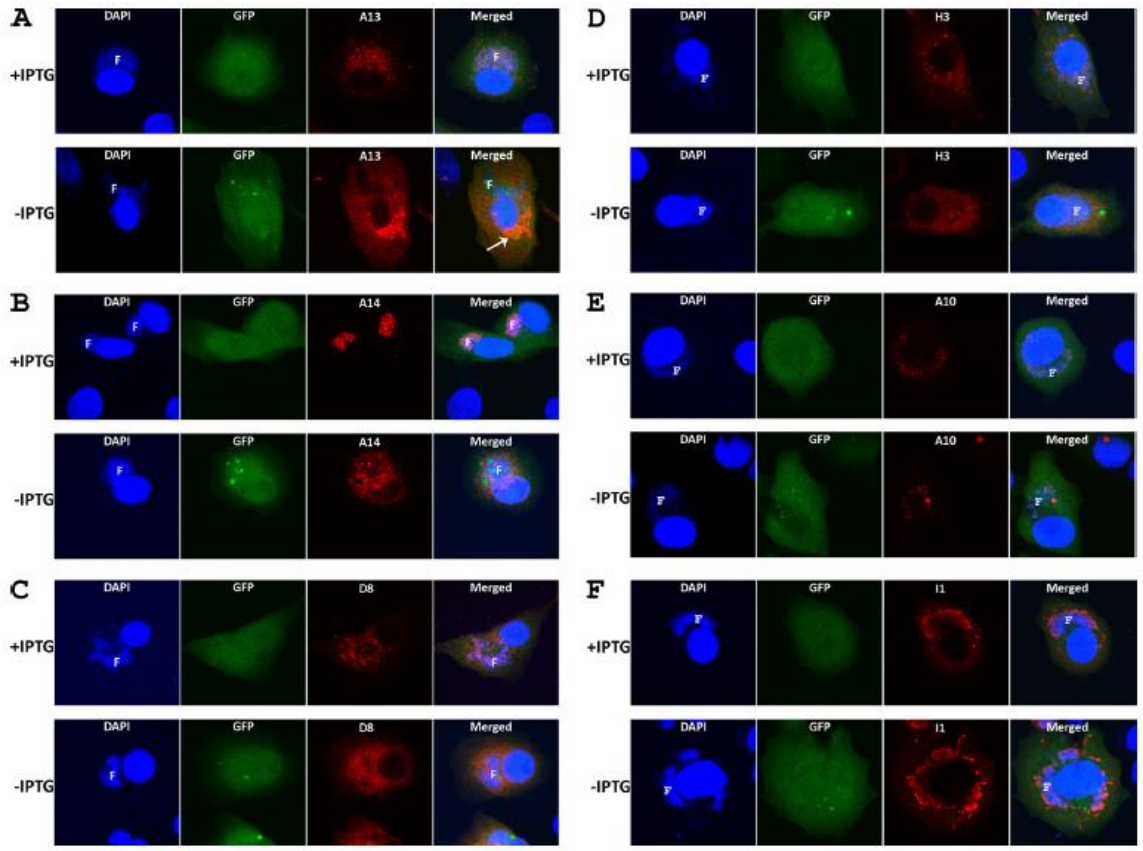


Figure 4-6: A6 effects the localization of A13, A14, D8 and H3. The cells were infected by virus with (+) or without (-) IPTG and then fixed, permiablized and stained with antibodies against each protein and DAPI. F stands for viron factories (Meng, Embry et al. 2012).

Our knowledge of A6 has largely come from morphogenesis observation via cell-biology-based microscopy. However, the detailed mechanism of how A6 interact with other virus proteins and cell skeleton and organelles remain unknown. Our dedication in solving the crystal structure of A6 aims to address the following questions. What proteins or elements in cell does A6 interact with? How virus utilize cell membranes and cytoskeleton for its own replication co-ordination?

In spite of no known A6 homologous structure based on amino acid sequence alignment, once we obtain the crystal structure of A6, we will be able to identify some functional domain using structural alignment, since the three dimensional structure is more conserved than amino acid

sequence. Such knowledge will expand our understanding of the role A6 plays during vaccinia virus replication.

B. Results

1. Cloning and Purification

A6 and its mutants were cloned into a modified pET28 vector, overexpressed them in *E. coli* and purified them in high quality.

The wild type A6 construct was obtained from our collaborator Dr Yan Xiang. We cloned wild type A6 and its mutants into a modified pET28 vector with an N-terminal His_6 tag SUMO (Small Ubiquitin-related Modifier) and *Ulp1* cleavable site. SUMO fusion can increase the solubility and stability of recombinant protein and *Ulp1* specifically recognizes SUMO and cleaves right at the N terminus of recombinant protein.

The plasmid was transferred electronically into BL21 DE3 gold strain with high expression ability. When cell growth reach OD_{600} 0.8~1.2, 1‰ IPTG was added into cell medium to induce protein expression at 18 ℃ for overnight. Cells were harvested the next morning using centrifugation at 15,000g.

The cell pellet was broken using a French press style high pressure homogenizer EmulsiFlex-C5. The His_6 tag binds to nickel resin though affinity chromatography under low imidazole concentration. Then the affinity is competed off by high concentration of imidazole. After that, the affinity tagged SUMO is cleaved by 1:50 to 1:100 ratio of *Ulp1* and the mixture of tag-cleaved protein, the tagged SUMO and uncleaved protein is applied to nickel resin again. Only tag-removed protein passed through. Finally, protein was concentrated using Millipore 30kD 15ml concentrator.

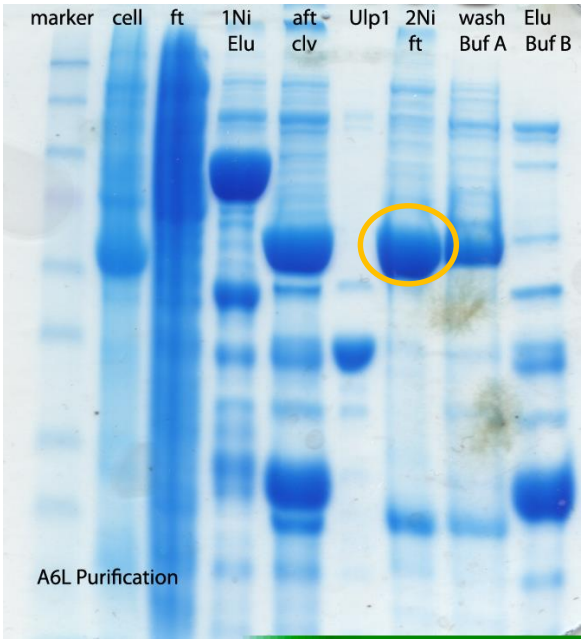


Figure 4-7: SDS-PAGE of samples collected along the purification of A6. The lanes from left to right are: protein marker, insoluble cell pellet, cell lysate flown through Ni resin, A6 with SUMO and His tag eluted from Ni resin, Ulp1 cleaved sample, A6 without His tag flown through Ni resin, A6 without His tag washed by 20mM imidazole from Ni resin, uncleaved material and Hig tagged SUMO eluted from Ni resin. Nickel purified A6 is shown in the orange circle.

2. Protein Characterization

a. Size Exclusion Chromatography

After obtaining tag-cleaved and purified A6 protein, we injected it into analytical size-exclusion column Superdex 200 (GE Superdex200) operated by GE AKTA purifier to determine its polymeric state in solution. Calculated from the retention volume of protein peak, A6 is monomeric in solution.

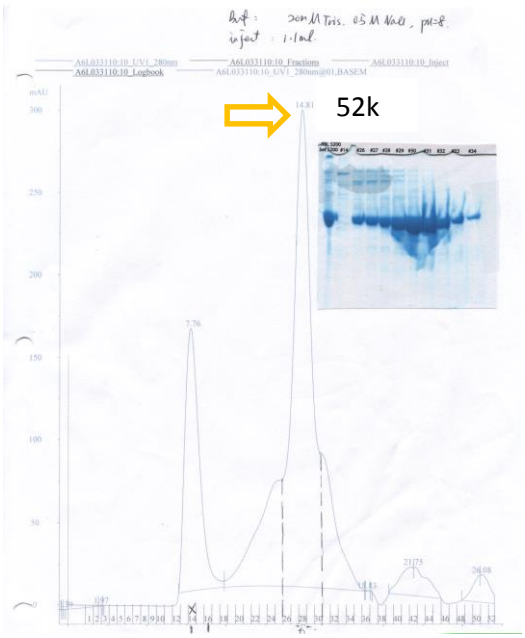


Figure 4-8: The size-exclusion chromatography figure of A6 and the SDS-PAGE of samples collected from the fractions around the protein peak. The protein peak is indicated by the arrow and from the retention volume. The calculated molecular mass is 52 kDa. Compared to A6's true molecular mass 43 kDa, it is a monomer in solution.

b. Dynamic Light Scattering

Dynamic light scattering serves as a powerful tool to determine the molecular weight distribution of particles in suspension or solution. After centrifuging 14,000 rpm for 15 minutes to remove any precipitant and insoluble particles, aqueous solution containing purified A6 was put in a cuvette in Malvern zetasizer. It was then hit by laser and the time scale of movement of scattered particles was calculated to reveal the partical size. The result of A6 DLS agreed with that from the size-exclusion chromatography.

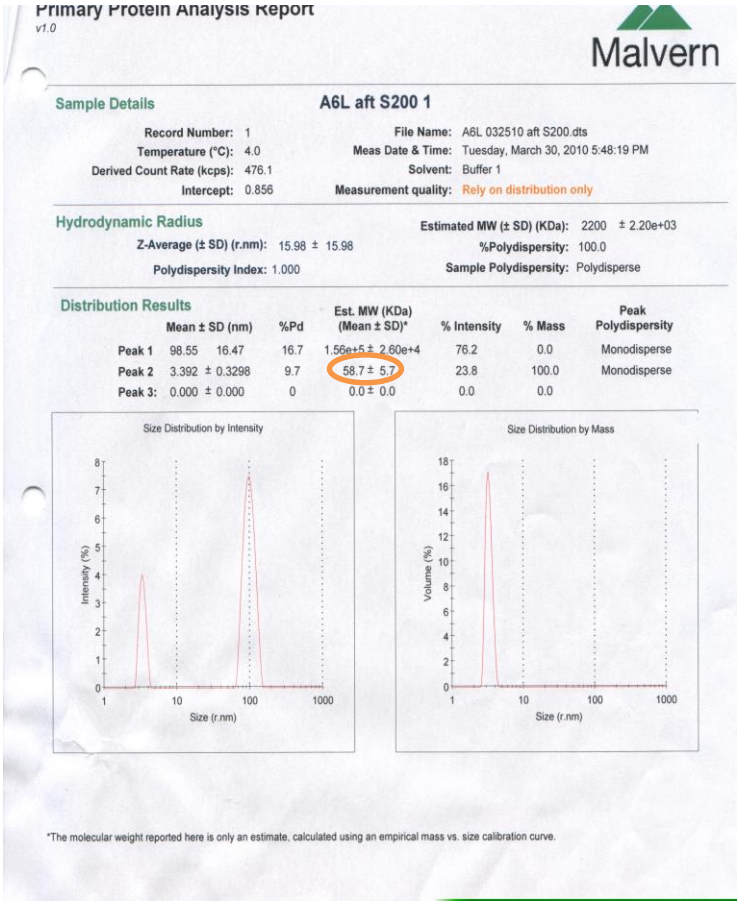


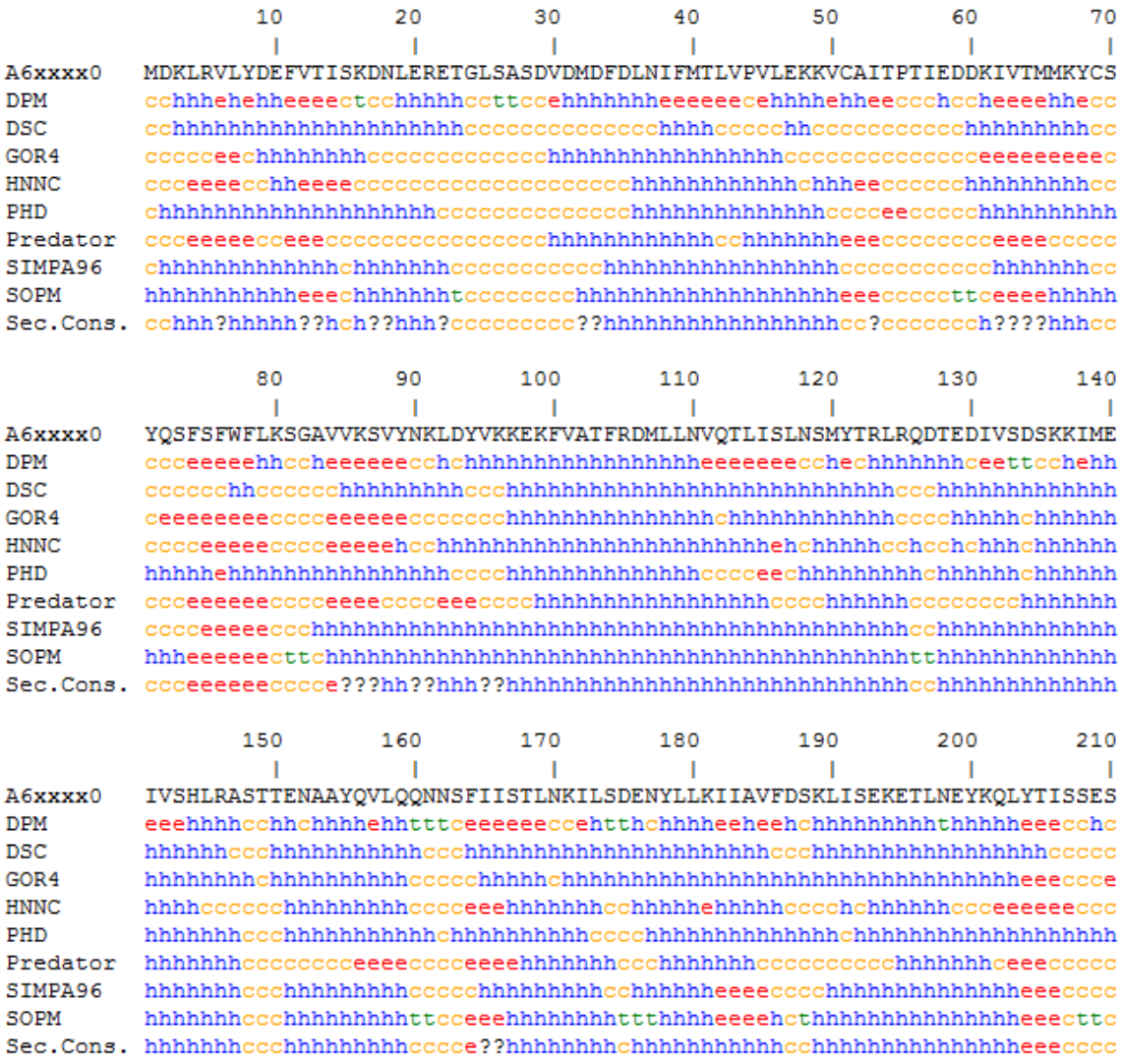
Figure 4-9: The Dynamic Light Scattering report of A6. The estimated molecular mass is 59 kDa, and it has monodisperse. Comparing to A6 true molecular mass 43 kDa, it is a monomer in solution.

c. Structure Prediction

i. Secondary Structure Prediction

We submitted A6 sequence to softwares DPM (Deleage and Roux 1987), DSC (King and Sternberg 1996), GOR4 (Garnier, Gibrat et al. 1996), HNNC (Guermeur, 1997), PHD (Rost, Sander et al. 1994), Predator (Fishman, 1997), SIMPA96 (Levin 1997), SOPM (Geourjon and Deleage 1994), SOPMA (Geourjon and Deleage 1995)) for secondary structure prediction and the consensus was integrated by the website

(http://www.bioinf.manchester.ac.uk/dbbrowser/bioactivity/NPS2.html). The result is shown as in Figure 4-10.


10 20 30 40 50 60 70
A6xxxx0 MDKLRVLYDEFVTISKDNLERETGLSASDVDMDFDLNIFMTLVPVLEKKVCAITPTIEDDKIVTMMKYCS
DPM
DSC
GOR4
HNNC
PHD
Predator
SIMPA96
SOPM
Sec.Cons.
80 90 100 110 120 130 140
A6xxxx0 YQSFSFWFLKSGAVVKSVYNKLDYVKKEKFVATFRDMLLNVQTLISLNSMYTRLRQDTEDIVSDSKKIME
DPM
DSC
GOR4
HNNC
PHD
Predator
SIMPA96
SOPM
Sec.Cons.
150 160 170 180 190 200 210
A6xxxx0 IVSHLRASTTENAAYQVLQQNNSFIISTLNKILSDENYLLKIIAVFDSKLISEKETLNEYKQLYTISSES
DPM
DSC
GOR4
HNNC
PHD
Predator
SIMPA96
SOPM
Sec.Cons.

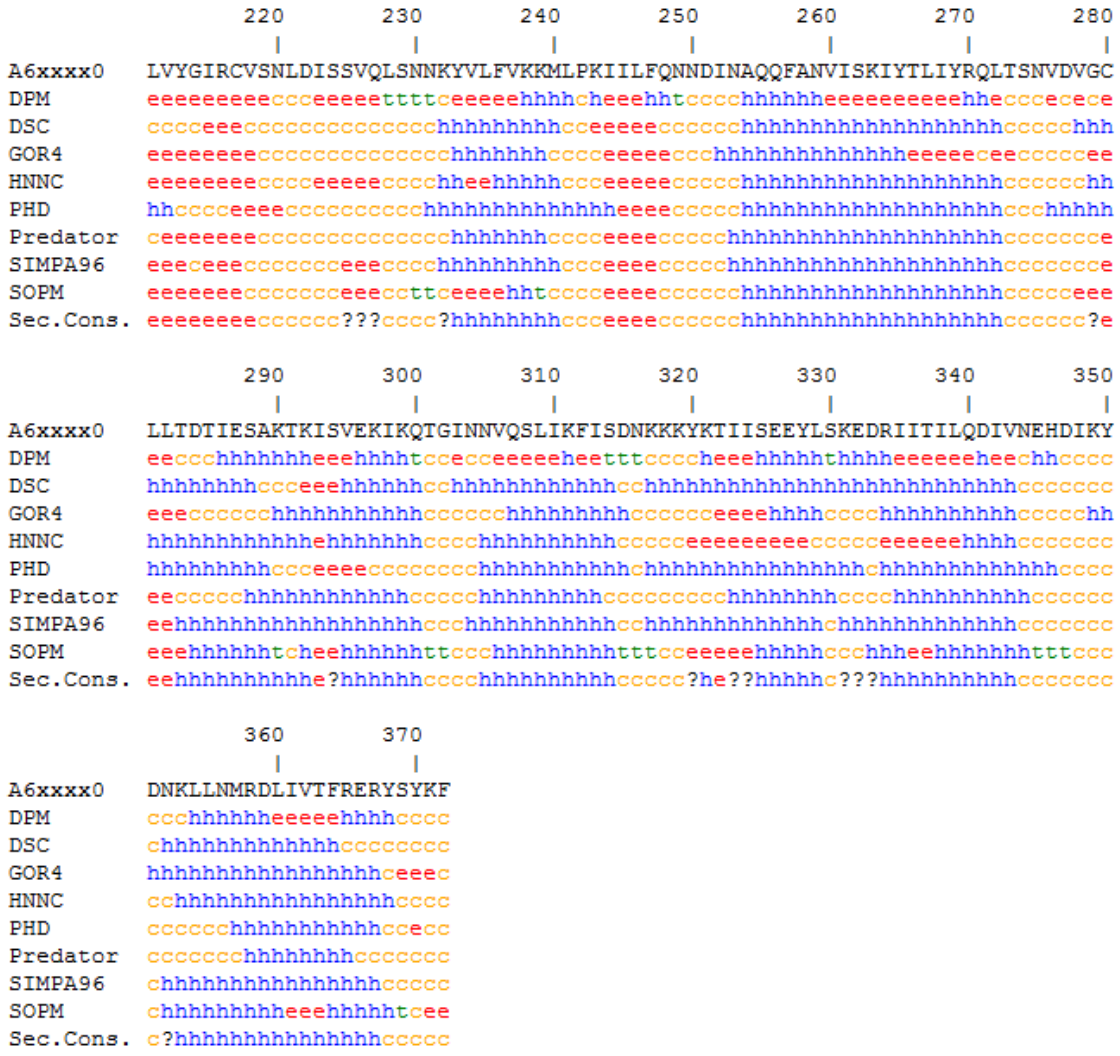


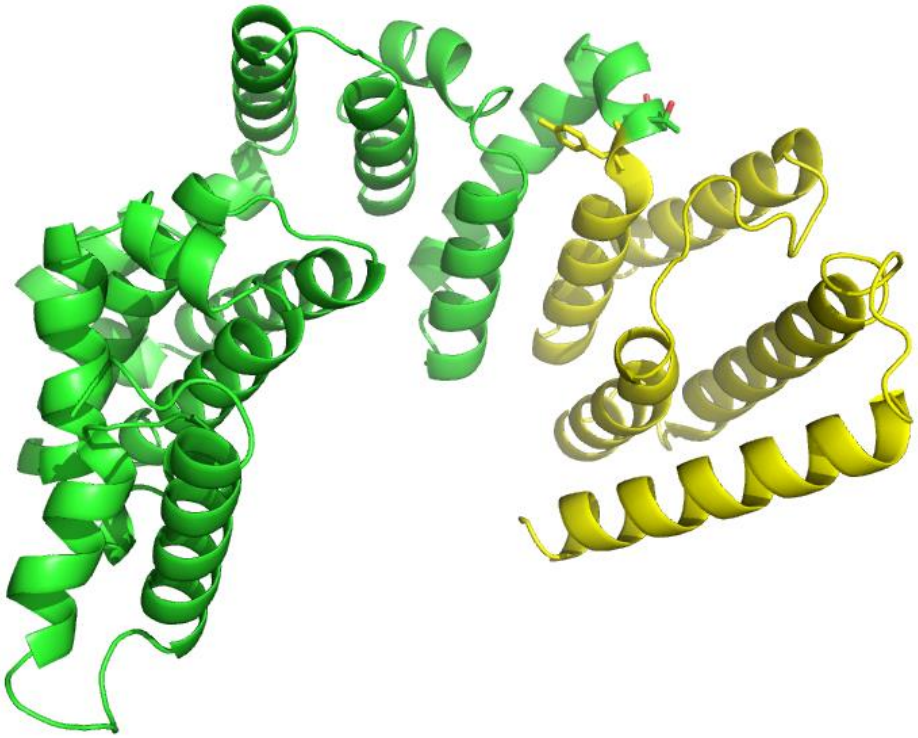
Figure 4-10: The consensus result of A6 secondary structure prediction.

According to the consensus, A6 is a well ordered protein which may contain 57.8% alpha helix, 7.53% extended strand, 25.27% random coil and 9.41% ambiguous states.

ii. Tertiary Structure Prediction

The A6 sequence was submitted in I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) for structure prediction (Zhang 2008). Five models were generated with C-score between -4.35 to -3.09. Only one model has the 121st and 122nd amino acids at the turn of α-helix,

while other models have them inside one α-helix. Combining the domain identification result in later section, we prefer this model. (Figure 4-11)



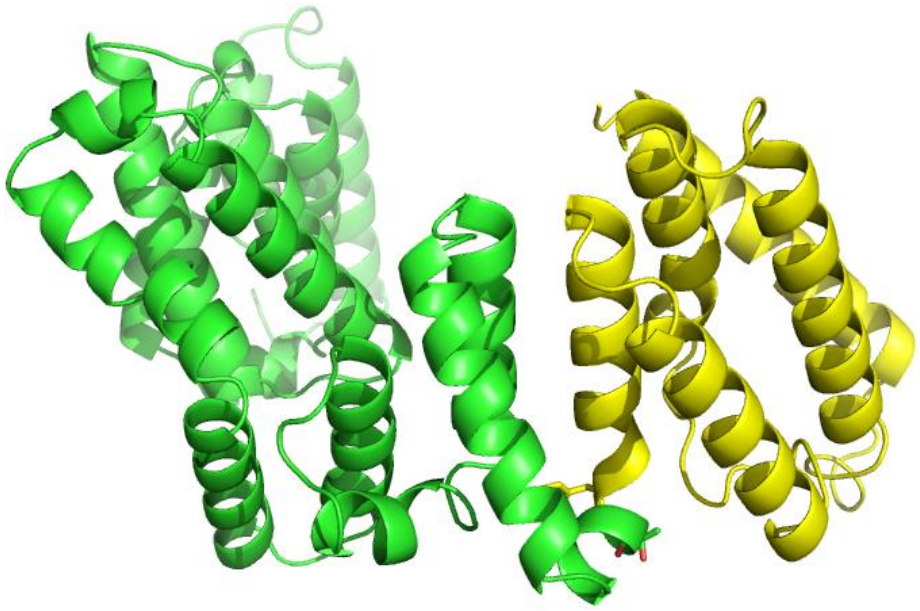


Figure 4-11: The ribbon structure of one I-TASSER predicted A6 model. N-terminal domain is colored in yellow and C-terminal is colored in green. Seeing from the top (upper image) and side (lower image). With 121st and 122nd amino acids shown in sticks.

3. Crystallization Screening

Purified wild type A6 in buffer [20 mM Tris, 0.5 M NaCl, 1 mM DTT, pH 8] was concentrated to 11.6 mg/ml and used for crystallization screening. Screening plates were set up in sitting drops using Art Robins Gryphon robot. 960 conditions were covered including Index, Crystal Screen, PEG/ION, SaltRx (from Hampton Research), Wizard 1 2 3 4 (Emerald BioStructures), JCSG I II III IV (from Qiagen). But no crystal hits were observed.

4. Surface Entropy Reduction (SER)

Wild type A6 is recalcitrant to crystallization, so we applied mutagenesis based on surface entropy reduction to assist crystallization.

One limiting step of protein molecules from solution to crystal is the formation of favorable crystal lattice contacts. Site specific alanine substitutions targeting patches of amino acids with long flexible side chains (usually Glu and Lys) would reduce the surface entropy and could overcome the energy barrier for lattice build up. While other conditions remain unchanged, the protein engineering by SER approach makes crystal forming thermodynamically favored (Cooper, Boczek et al. 2007).

a. Site Direct Mutation

Surface Glu and Lys might prevent nonspecific aggregation and precipitation towards crystallization, for their long flexible side chains interact with solvent molecules (Derewenda and Vekilov 2006). Using program http://services.mbi.ucla.edu/SER/ we predicted several clusters of Glu or Lys (shown in Figure 4-11) to be sites of mutation and made several constructs containing each site or combined sites with Ala substitutions.

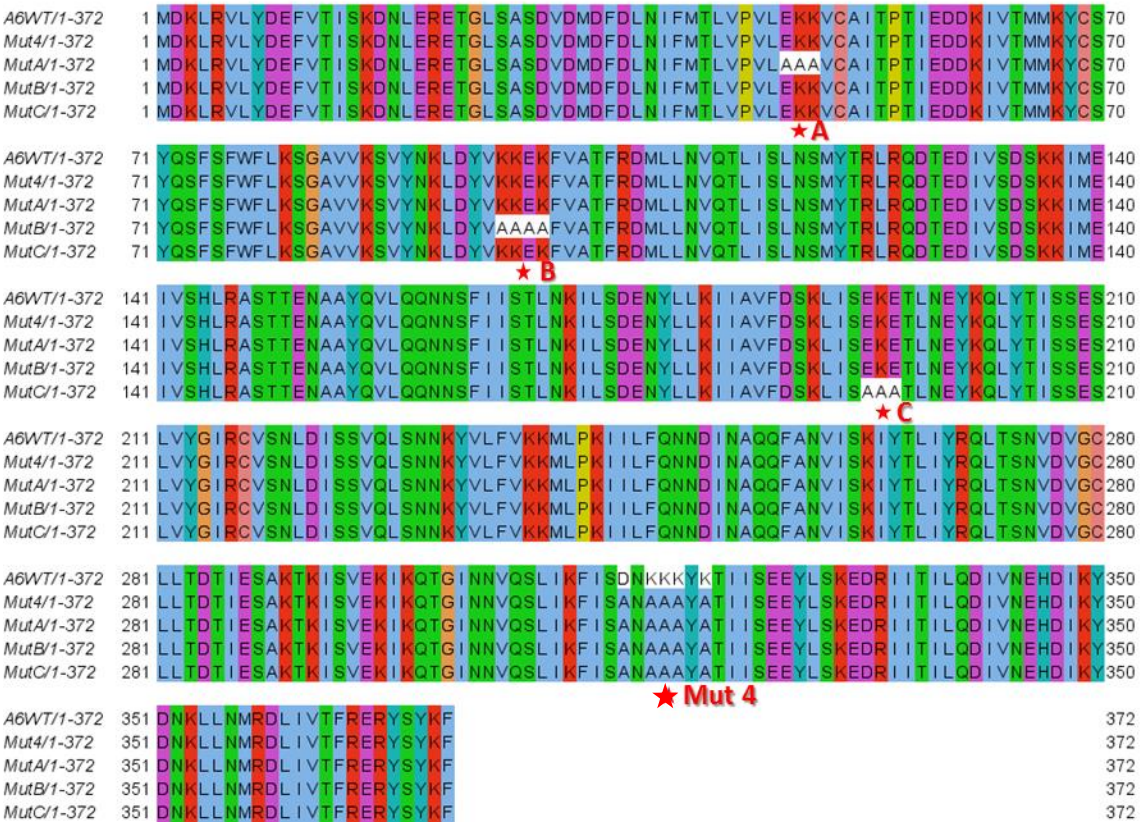


Figure 4-11: The alignment of sequences of A6 wild type and mutants. The sites of mutation are labeled with stars underneath.

5. Purification and Crystallization of Mutant AC

Mutant AC (alanine mutations at both A and C sites) was generated by overlapping PCR method and cloned in the same way as the wild type A6 as described above. Protein purification and characterization were carried out in the same way as described in II.B.

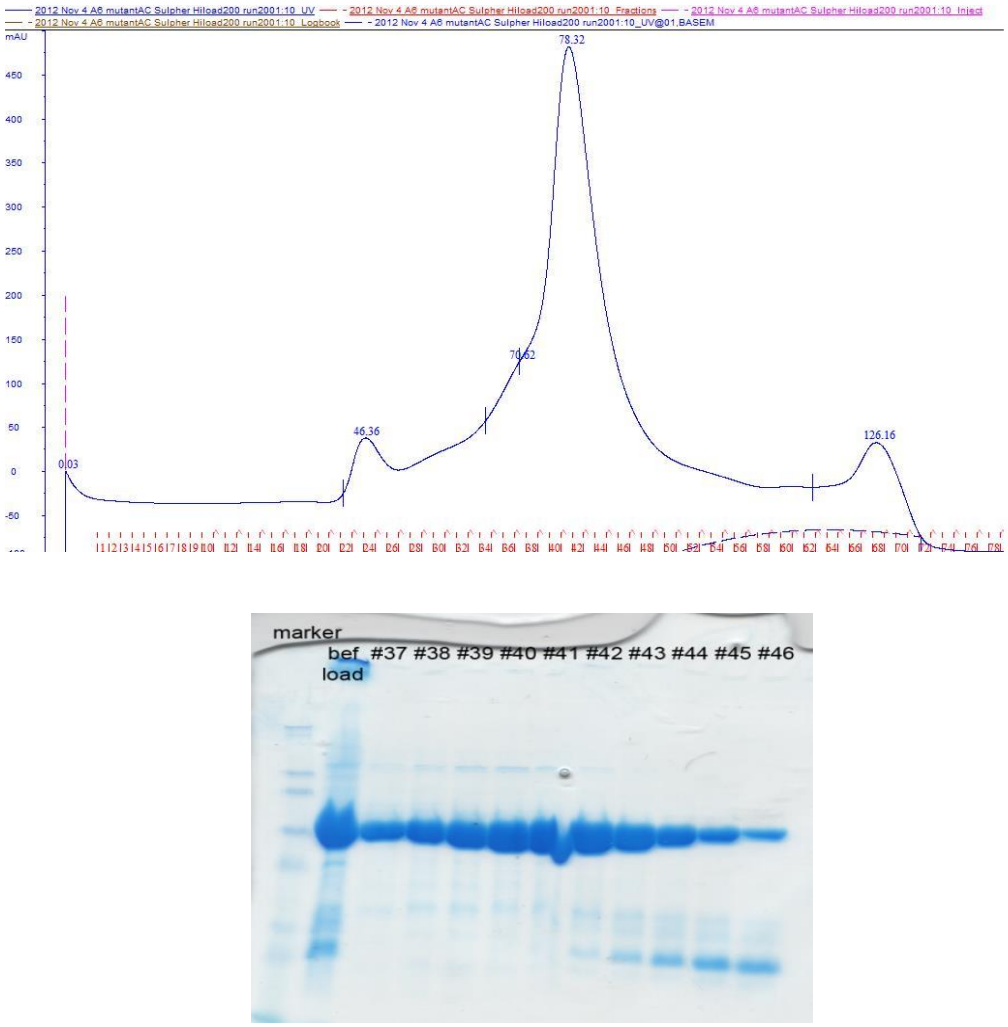


Figure 4-12: The chromatograph of A6 mutant AC in GE HiLoad S-200 size exclusion column and the SDS-PAGE gel showing the protein samples from the corresponding fractions.

Mutant AC formed as rectangular block crystals under the condition of 0.1 M Tris, 8% PEG8000, pH7.0.

Figure 4-13: The A6 mutant AC crystals.

The crystal was harvested and data was collected and processed as described in II.F. Crystal diffracted to 4 Å and one unit cell contained more than 20 molecules. The data was undecipherable.

a. Native Gel Analysis

The large amount of molecules in one unit cell suggested that mutant AC might not be homogenous in the crystal. A6 may adopt dynamic conformations in solution. To further characterize A6 protein, we carried out native PAGE analysis. Native PAGE gel system is the polyacrylamide gel electrophoresis run in non-denaturing conditions so that the protein or protein complex analyzed remains its natural state. The speed of protein or protein complex migrating is determined not only by its size but also by the charge and shape.

We used native gel to analyze the homogeneity of AC. Although AC appeared to be homogenous monomer in size-exclusion chromatography, on native gel the AC sample was separated into one major band and one minor band. It indicated AC may carry different charges or bind some small molecules like lipids.

b. Ion Exchange Chromatography of A6

We further try to separate any possible contaminant using ion exchange chromatography. Ion exchange separates polar molecule based on the surface charge.

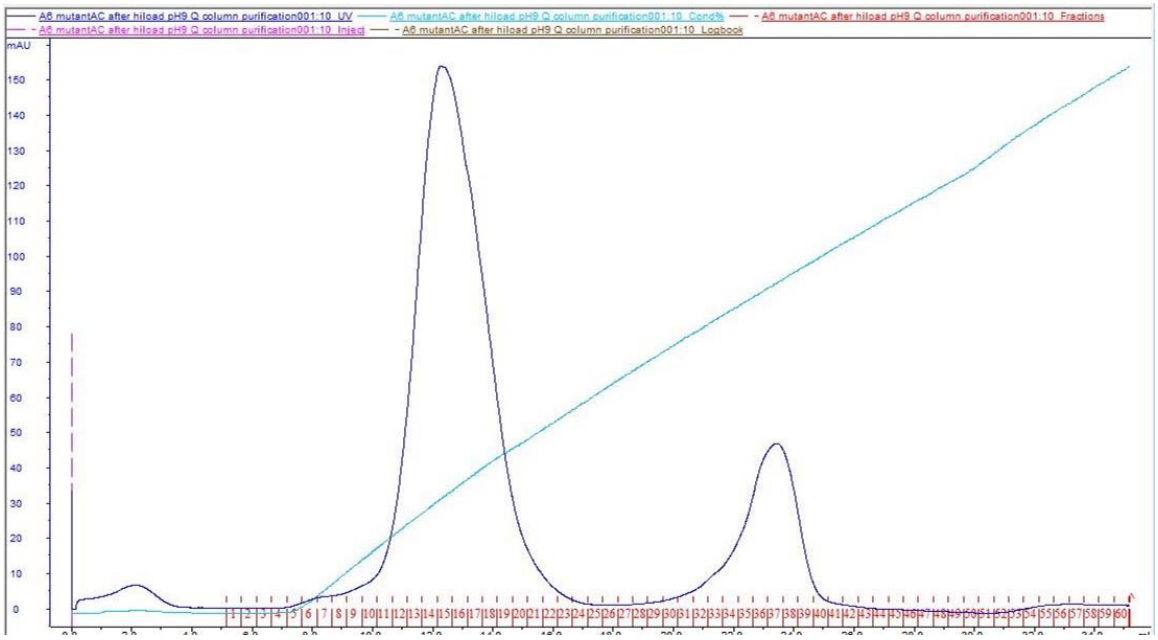


Figure 4-14: The chromatograph of HQ ion exchange analysis of AC. The light blue slash indicates the cond %.

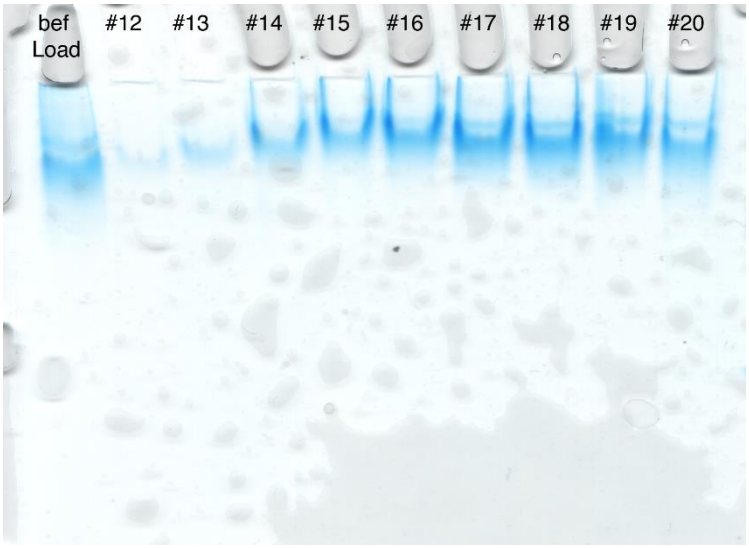


Figure 4-15: The native gel analysis of samples collected from HQ ion exchange chromatography.

6. Domain Identification

During the study of recombinant virus with IPTG inducible A6, a steady C terminal fragment of A6 around 25kD was observed (Meng, Embry et al. 2012). It is quite possible that a small portion of full length A6 is degraded naturally by protease inside the cell. A6 might contain a stable C-terminal domain which is more likely to crystalize.

a. Limited Proteolysis

We used proteases trypsin and chymotrypsin to cut off the loose and flexible peptide from protein and leave rigid region, which might assist crystallization. Trypsin cleaves at the carboxyl side of Lys and Arg while chymotrypsin prefers to cut after Tyr, Trp and Phe.

Various ratio of trypsin or chymotrypsin were applied for various amount of time. It was noticed that A6 was gradually degraded to a steady fragment around 30 kDa from both digestions.

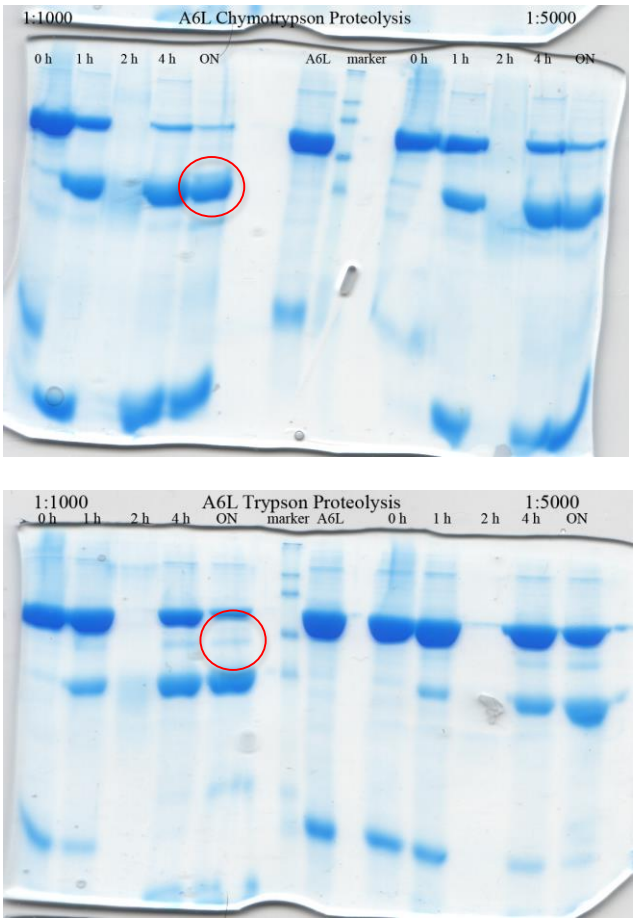
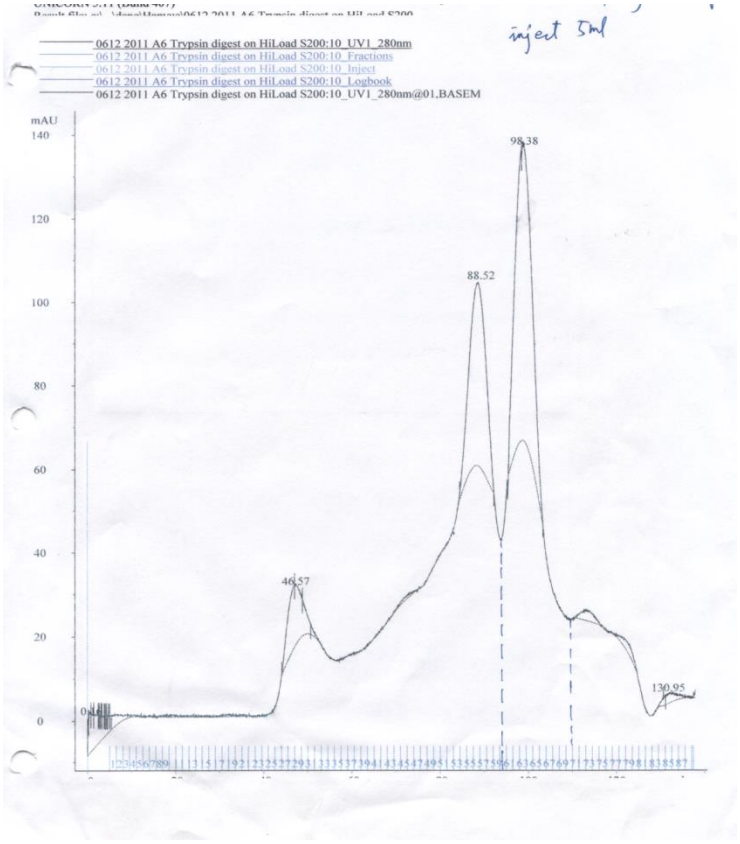
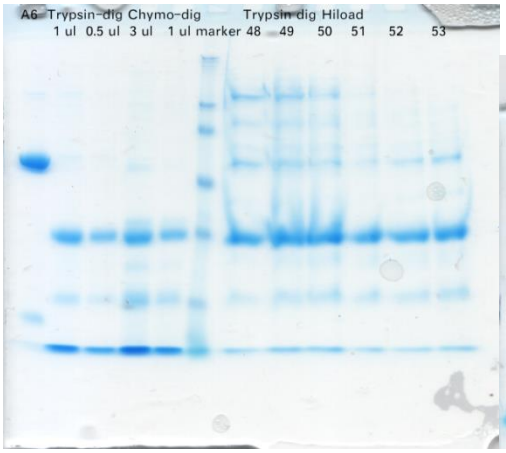


Figure 4-16: SDS-PAGE of A6 digestion by trypsin or chymotrypsin. The length of digestion is indicated upon each lane. The circled bands are the stable bands sent for mass fingerprint.

b. HiLoad Size Exclusion Chromatography

Then 1:1000 molar ration of trypsin or chymotrypsin was used to digest overnight and loaded the sample onto a HiLoad sizing column. From the chromatograph (Figure IV-18) we could tell that there were two distinct peaks, each representing a complex of peptides. The two peptides complexes can be concentrated to 15mg/ml and 8 mg/ml. I used the material to set up crystallization trails but no crystal was observed.





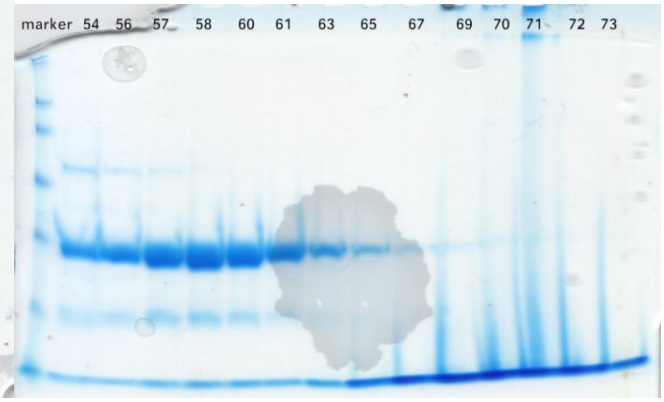
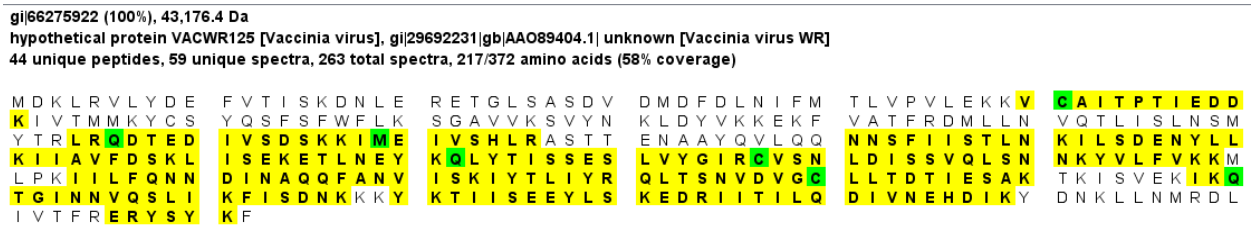


Figure 4-17: The chromatograph of HiLoad size-exclusion of trypsin digested A6 and the SDS-PAGE showing the sample collected from each fraction.

c. Mass Fingerprint Identification

Chymotrypsin and trypsin digestion both generated steady fragments around 30 kDa. After digestion, these two fragments were sent to the core facility for mass fingerprint analysis. When selecting peptide threshold to be 80% and peptide count to be 2, Vaccinia Virus A6 was identified to be 100% for both peptide. The result is shown in Figure 4-19. The highlighted area are the 80% identified peptides. Both fragments appeared to be primarily the C terminal part of A6, with some N terminal peptide attached.

A



B

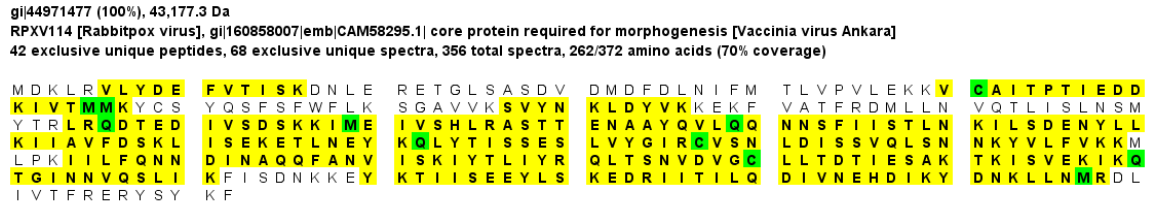


Figure 4-18: The Mass fingerprint result of the band from chymotrypsin digestion (A) and the one from trypsin digestion (B). They are both identified as part of A6, primarily C terminus of A6.

d. Western Blot Confirmation

Intact A6, trypsin digested A6 and chymotrypsin digested A6 were run in SDS-PAGE and transferred to membrane. Then antibodies generated from mice by collaborator targeting C-terminus or N-terminus of A6 were used as first antibody and anti-mice antibody was used as secondary antibody. The result confirmed that the fragments around 37 kD are C-terminus parts of A6.

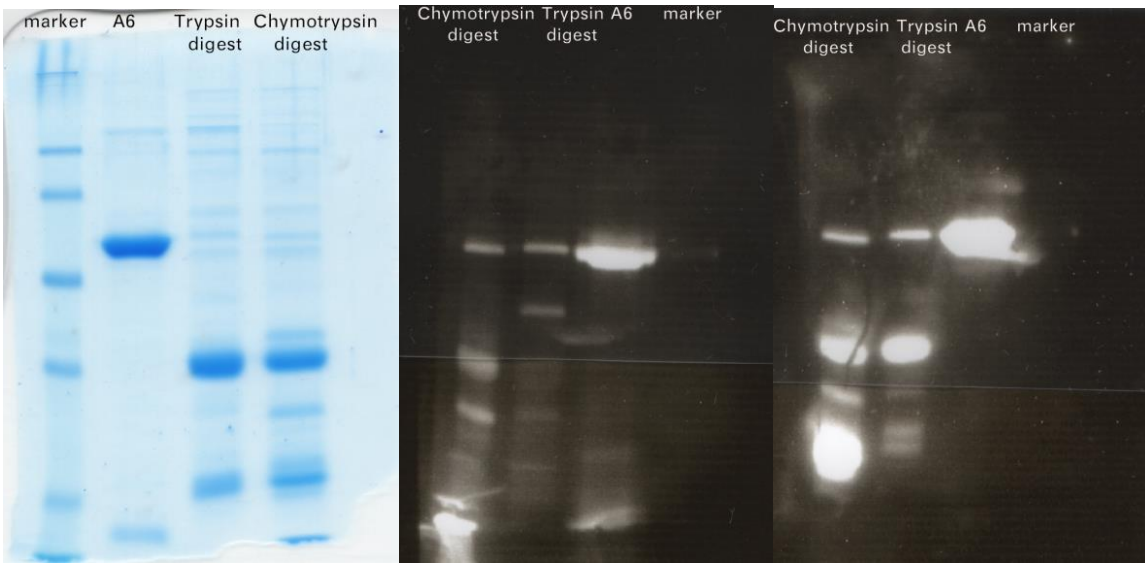


Figure 4-19: The SDS-PAGE gel, the N-terminus antibody as first antibody western blot and the C-terminus antibody as first antibody western blot. It shows that the steady bands around 37kD are the N-terminal part of A6.

e. N-terminal Sequencing

We sent the steady fragments around 37 kDa generated from limited proteolysis to Iowa State University for N-terminal sequencing. The result indicated the starting residue of chymotrypsin digested part is 122nd A.A. and that of trypsin digested part is 124th A.A. We made construct of this domain.

7. N Terminal Domain of A6

Truncated version of wild type A6 and mutants were constructed and purified as described above.

The truncated mutant C formed cubic crystals under the condition of 0.1 M MES, 2 M ammonium phosphate, pH 6.0.

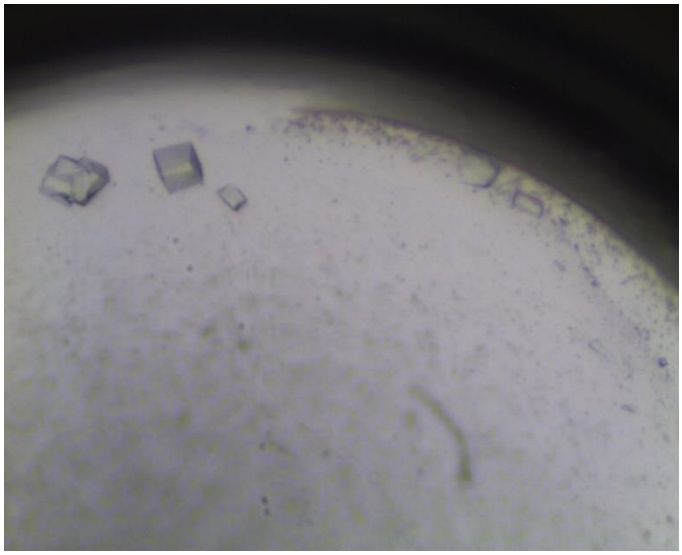


Figure 4-20: The crystal of truncated mutant C.

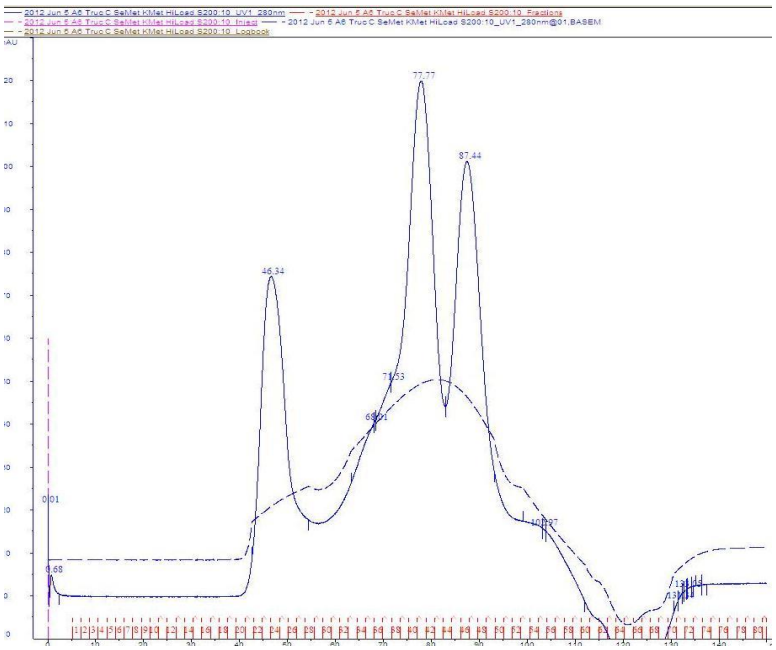
Data collection and processing was carried out as described in II.F. The crystal diffracted to 8Å and data was not enough for structure determination.

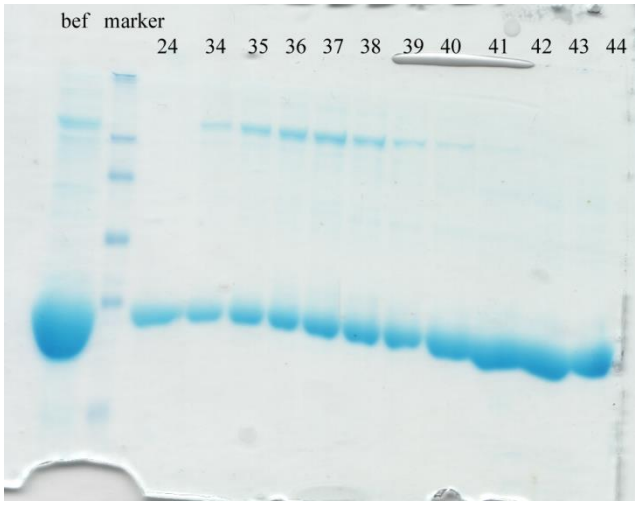
a. Lysine methylation

Reductive alkylation on lysine residues is a chemical modification on molecule surface that reduces surface entropy and promotes crystallization. We use formaldehyde as substrate and dimethylamine-borane complex as reducing agent following a previous protocol (Rayment 1997). The first step of reaction is the free amine of lysine on the protein surface forming a Schiff base with formaldehyde and then reduced to a secondary amine. After that, monomethyl lysine reacts rapidly with formaldehyde to dimethyl amine.

We mixed 40 ul freshly made 1 M formaldehyde and 20 ul freshly made dimethylamine-borane complex solution with each ml of diluted protein sample (1mg/ml) and gently rock the container in 37 ℃ for 2 hours. After that, the same amount of formaldehyde and dimethylamine-borane complex was added into the reaction. 2 hours later, a final 10 ul of dimethylamine-borane complex was supplied and the reaction was incubated overnight.

Lysine methylation reduces protein solubility and increases mass, thus lots of precipitation occurs after the reaction. After centrifuging, size-exclusion chromatography can be applied to obtain homogenous sample. (Walter, Meier et al. 2006) The sample was concentrated and set up for crystallization. But no crystal hit was observed.





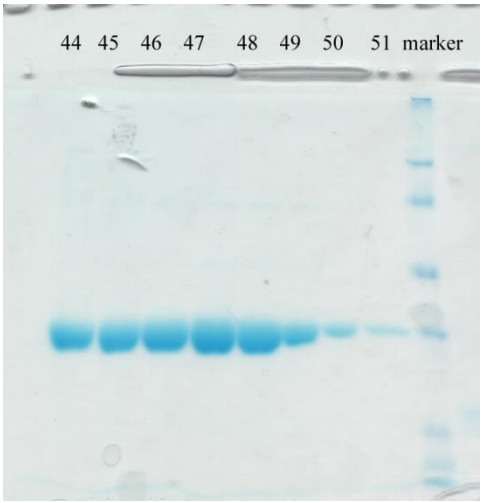


Figure 4-21: The HiLoad size exclusion chromatograph of lysine-methylated truncated mutant C seleno-methionine protein and the SDS-PAGE gel showing samples collected from corresponding fractions.

8. Antibody Assisted Crystallization

a. Rationale

In the past few years, the use of antibodies in co-crystallizing proteins that are otherwise recalcitrant to crystallization has witnessed great successes. For example, the determination of G-protein coupled receptor β(2) adrenoceptor structure greatly benefited from a specific single domain antibody (sdAb) (Rasmussen, Choi et al. 2011). The use of antibodies in co-crystallization has at least two benefits: 1) stabilizing flexible regions of the target protein to reduce entropy; 2) creating crystal lattices through antibody-mediated interactions. 3) The available structure of antibodies will be useful for structure determination of the complexes by using molecular replacement method.

b. Fragment Antigen-Binding (Fab) Complex Crystallization

i. Rationale

When A6 fails to crystallize itself, Fab generated from monoclonal antibody was used to assist crystallization. The fragment antigen-binding (Fab fragment) contains one variable region and one consistant region of each heavy chain and light chain, thus Fab retains the antigen binding ability of antibody while its size is significantly smaller. Fab has become a powerful assistant in crystallization. It can form a layer facilitating crystal lattice formation, and if a crystal of protein with antibody is obtained, no additional phasing is needed because the structure of antibody is already known and molecular replacement can be applied.

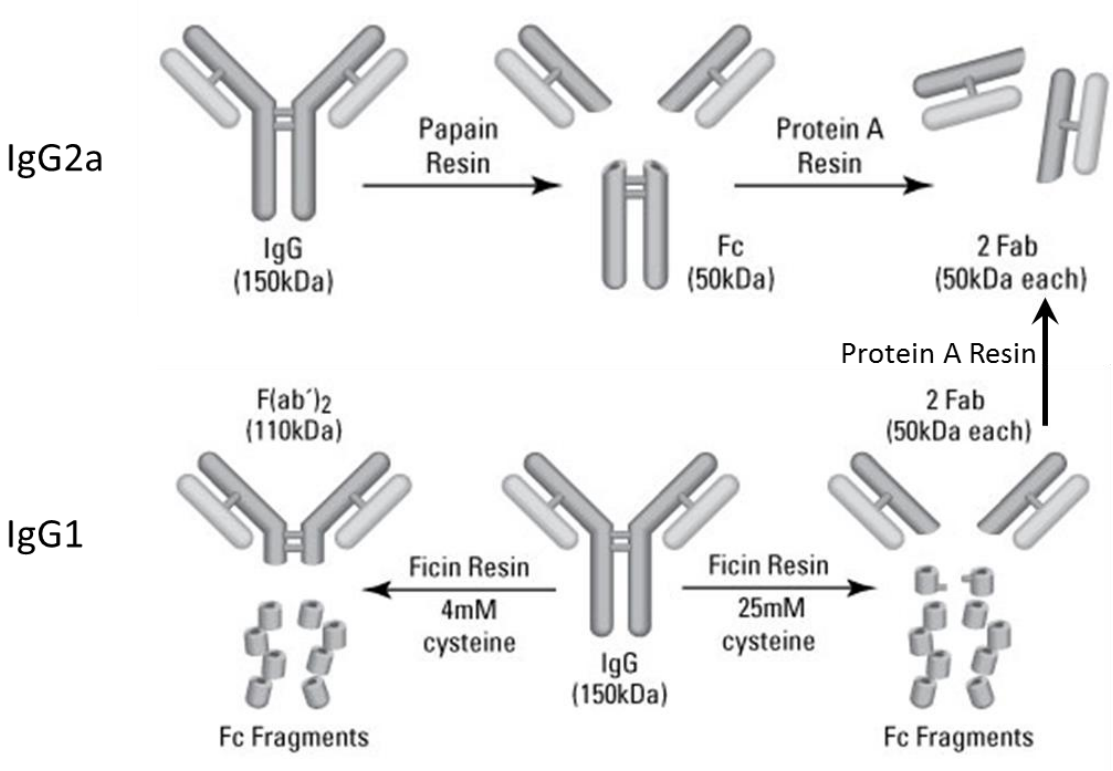


Figure 4-22: Fab generation rationale. From Pierce Chemical.

ii. Fab Generation Procedure

We provided A6 protein to our collaborator in San Antonio and they used it to generate monoclonal antibody against the N-terminal or C-terminal of A6.

Papain, a nonspecific thiol-endopeptidase derived from papaya, or ficin, that derived from figs latex, was used to digest the antibody into $(Fab')_2$ and Fc fragments or Fc pieces, followed by cysteine to mildly separate $(Fab')_2$ to two Fab fragments.

Protease (papain or ficin) was first activated in the buffer containing 10 mM Cysteine in 37 ℃ for 15 minutes. Then cysteine was removed by buffer exchange. This activated protease was added into monoclonal antibody around 1:100 molar ratio and digested for 2 hours. Iodoacetamide (IAA) was used to stop the digestion and immobilized Protein A was used to capture Fc fragments and undigested antibody and Fab will stay in flow-through solution.

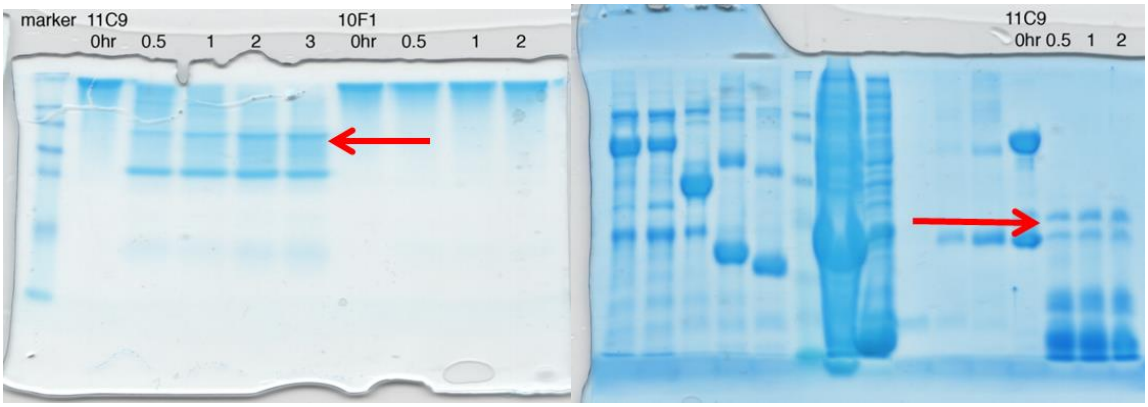
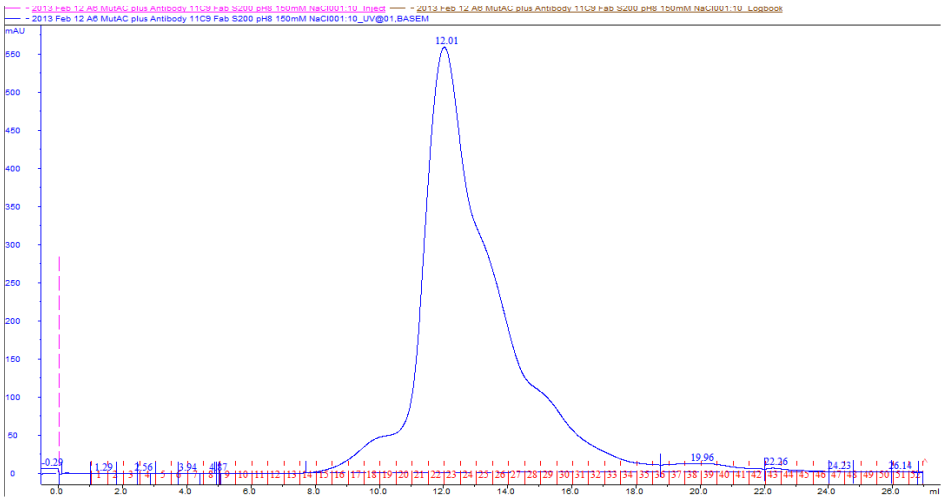


Figure 4-23: The non-reducing (left) and reducing (right) SDS-PAGE gels of test digestion of monoclonal antibodies 11C9. The 50 kD band in the left gel and the two 25kD bands in the gel highlighted by the red arrows are the non-reduced and reduced sample showing generated Fab.

After Fab is purified, we can mix it with A6 and run the mixture in size exclusion column and obtain the complex. The complex was concentrated and used for crystallization set up.

iii. Complex of Mutant AC Plus Fab



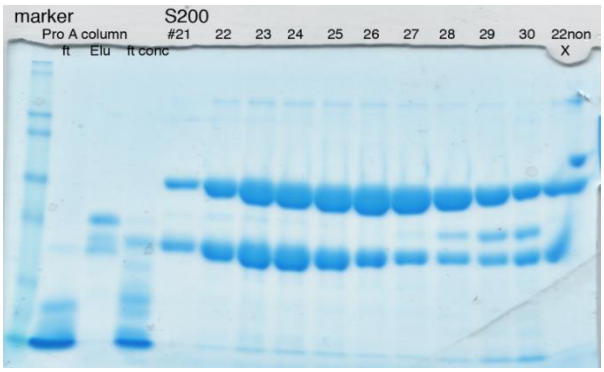
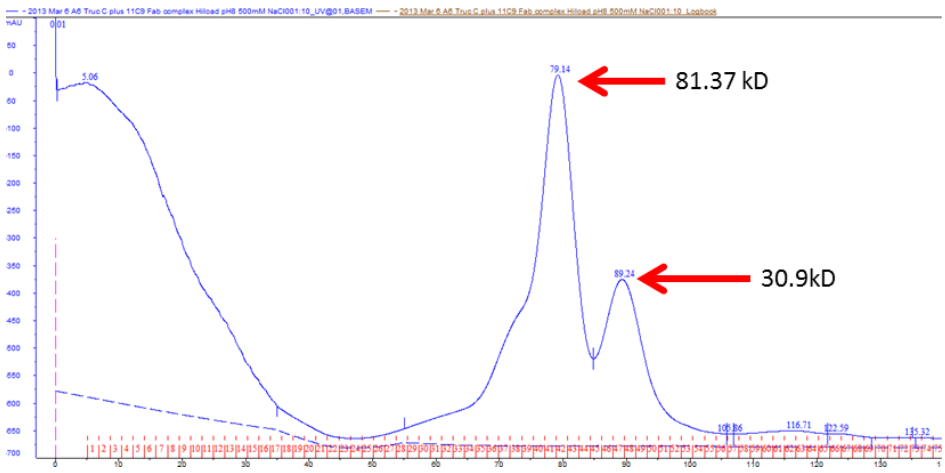


Figure 4-24: The size exclusion chromatograph of A6 mutant AC plus 11C9 Fab and the SDS-PAGE gel showing samples from corresponding fractions. The calculated molecular mass of this complex peak is 170 kD, suggesting this complex might be a hetero-tetramer containing two AC molecules and two Fab fragments.

This peak was concentrated and used for crystallization screening, but only Fab crystals were obtained. This might be caused by low affinity between AC mutant and the Fab because Fab was generated from wild type A6.

iv. The Complex of Truncated Mutant C and Fab



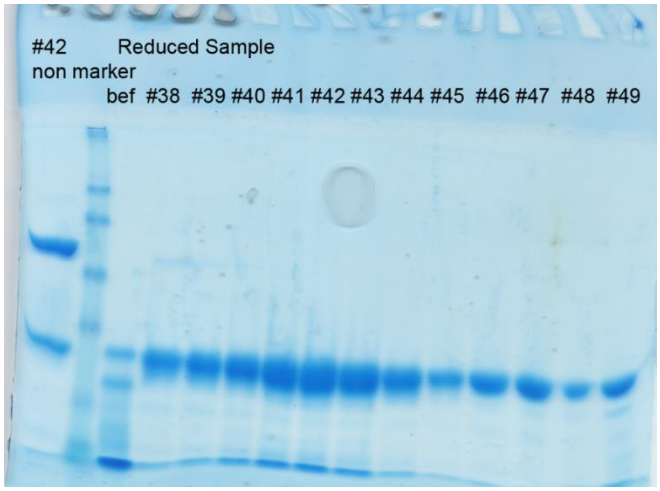


Figure 4-25: The size exclusion chromatograph of truncated A6 mutant C plus 11C9 Fab and the SDS-PAGE gel showing samples from corresponding fractions. The calculated molecular mass of this complex peak is 81.37 kD, suggesting this complex might be a hetero-dimer containing one truncated C molecule and one Fab fragment.

c. Single Domain Antibody (SdAb) Complex Crystallization

i. Rationale

One type of heavy chain antibodies in absence of light chains are found in sera of camelidae (Hamers-Casterman, Atarhouch et al. 1993). Single domain antibody (nanobody), the smallest segment of immunoglobin retaining antigen binding ability, contains only one variable region derived from heavy chain. It is remarkably smaller than traditional antibody, thus more stable, while it can be produced faster and more economically (Vincke and Muyldermans 2012). In crystallography, SdAb has proven to be a powerful auxiliary tool (Lam, Pardon et al. 2009). It can facilitate crystallization while it binds to protein that is recalcitrant to crystallization on its own and may provide crystal lattice contact. Once complex crystal formed and X-ray diffraction data obtained, SdAb can also provide the phasing information during structure determination because the structure of SdAb is already known. It is more convenient and at lower cost than conventional

antibody or Fab generated from antibody because recombinant SdAb is expressed in *E. coli* or yeast.

In our case, purified protein (A6) was injected to a llama six times in mixture of other proteins. Heavy chain antibody mRNA was extracted from lymphocytes, ligated to pHEN4 phagemid vector and subsequently transformed to *E. coli* TG1 cells (Conrath, Lauwereys et al. 2001). In this way, a VHH library was constructed. The phages containing this generated VHH library infected *E. coli* in the presence of M13K07 helper phage and expressed VHHs. A6 protein was used to screen its binders through multiple runs of phage display. Enriched colonies were applied to a phage ELISA to analyze the binding affinity. 40 to 50 positive individual candidates were selected and sent for sequencing to obtain distinct VHHs. After that, all the selected nanobody genes were cloned into pHEN6 vector for expression in *E. coli* (Lam, Pardon et al. 2009).

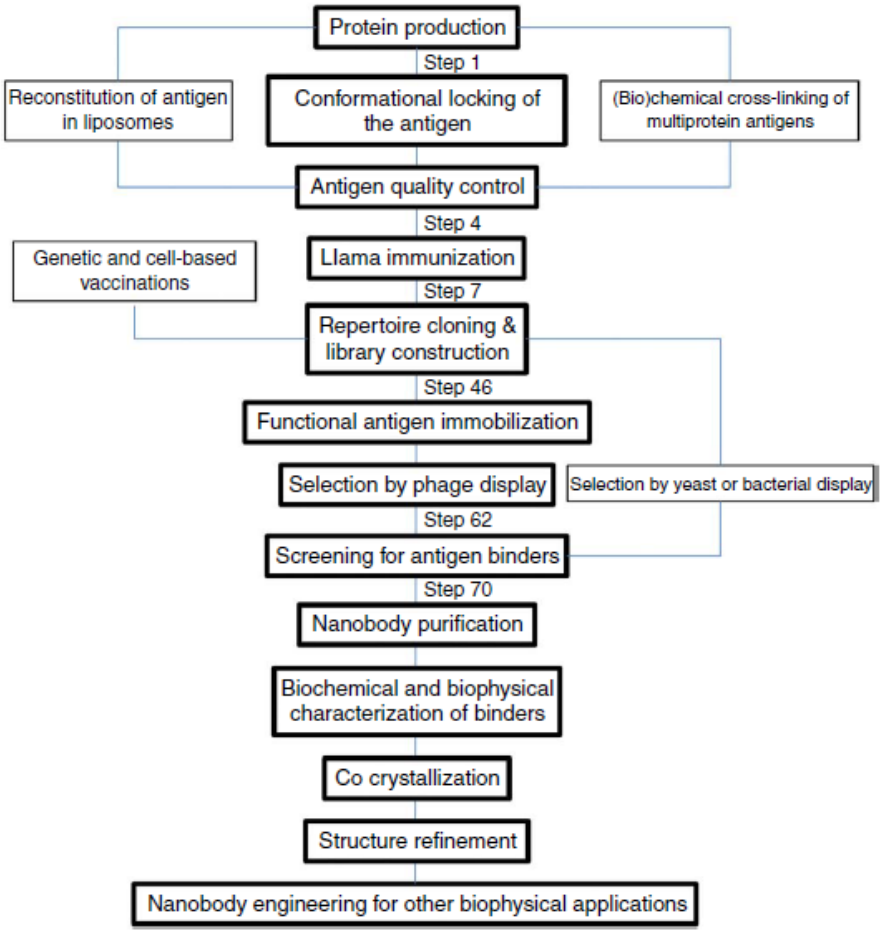


Figure 4-26: Workflow of single domain antibody (SdAb/nanobody) generation (Pardon, Laeremans et al. 2014).

ii. Periplasmic Expression

The major advantage of periplasmic expression is that only correctly folded protein can be secreted to periplasm. Besides, SdAb might be toxic to cell, so periplasmic expression avoid cytoplasmic overexpression. The expression protocol starts from strains frozen in -80 ℃, which are grown in Terrific Broth (TB) plus 2% glucose and antibiotic in 25 ℃ or 30 ℃ overnight. Next morning, cells are centrifuged down and inoculated into TB. Cells are grown to high density

(OD~1.5) and induced with 1‰ IPTG. Protein express is allowed for 4 to 5 hours in 23℃ before harvest cell.

iii. Osmotic Shocking Purification

Correctly folded SdAb is secreted to the periplasm of cell. So instead of breaking cell completely like our conventional routine, we use osmotic shocking to avoid the unfolded cytoplasmic protein.

Cells are resuspended in 15 ml hypertonic Tris-sucrose solution (100 mM Tris, 0.75 M sucrose, pH 8.0) with gentle pipetting and 2 ml of 3 mg/ml lysozyme is added to digest peptidoglycan. The tube is rocked at 4℃ for 15 to 20 minutes. 2.8 ml of 10 mM EDTA is then added drop to drop into solution while gently shaking the tube to prevent high local concentration and rock for 30 minutes. This step is to permeablize the outer membrane of *E. coli*. Then 2 ml of 0.5 M $MgCl_2$ is added to chelate any extra EDTA. The buffers are mixed well and centrifudged at 10,000 rpm for 20 minutes. The supernatant is collected and mixed with Nickel resin. SdAb with permanent His tag at C terminal is captured by nickel resin and later eluted with 250 mM imidazole buffer.

Constructs were expressed and purified as described above (by labmate Lingyi Tang). Then purified A6 was mixed with each SdAb and injected into size exclusion column.

iv. Complex Analysis

Antibody 458 seems to have low binding affinity to A6 wild type because it caused A6 peak to shift, but the ratio of A6 and 458 in the complex peak was not stoichiometric indicating that the

complex binding was not tight enough (Figure 4-27).

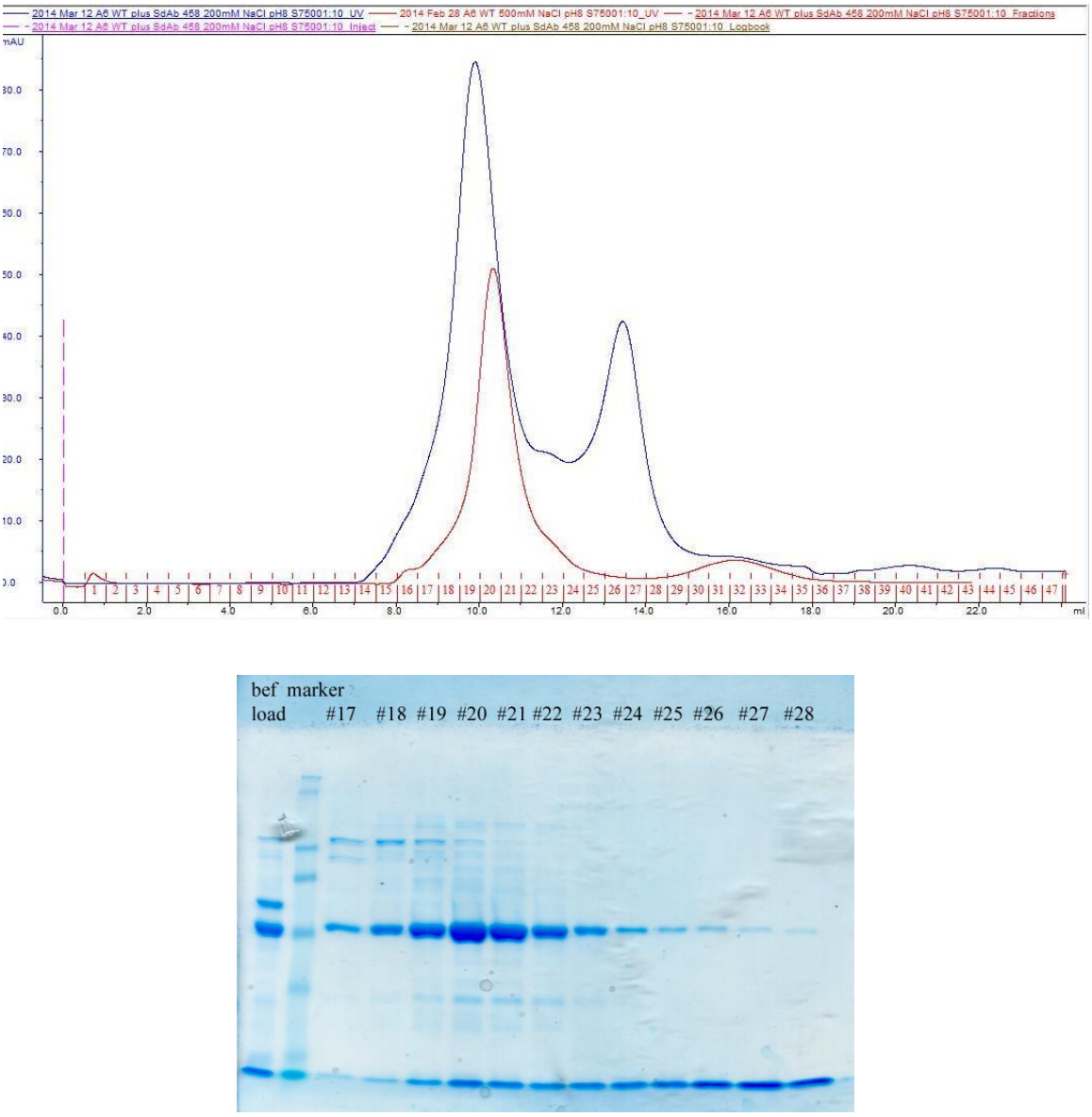


Figure 4-27: The size exclusion chromatograph of A6 wild type plus SdAb 458 complex on S75 (GE Health) column and SDS-PAGE showing samples collected before injection and from corresponding fractions.

The binding between antibody 458 and A6 is even weaker when NaCl concentration increases to 500 mM (Figure 4-28).

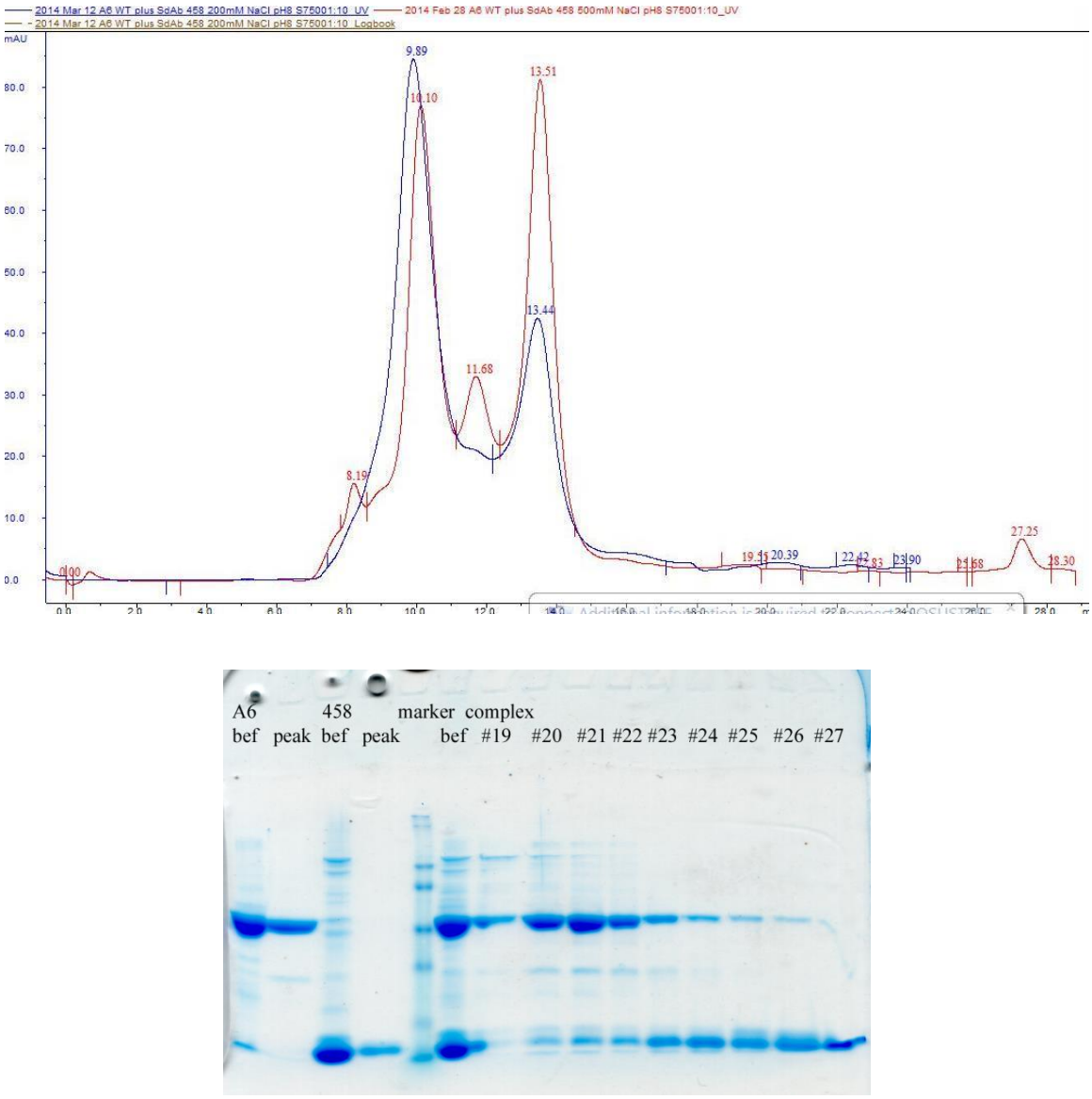
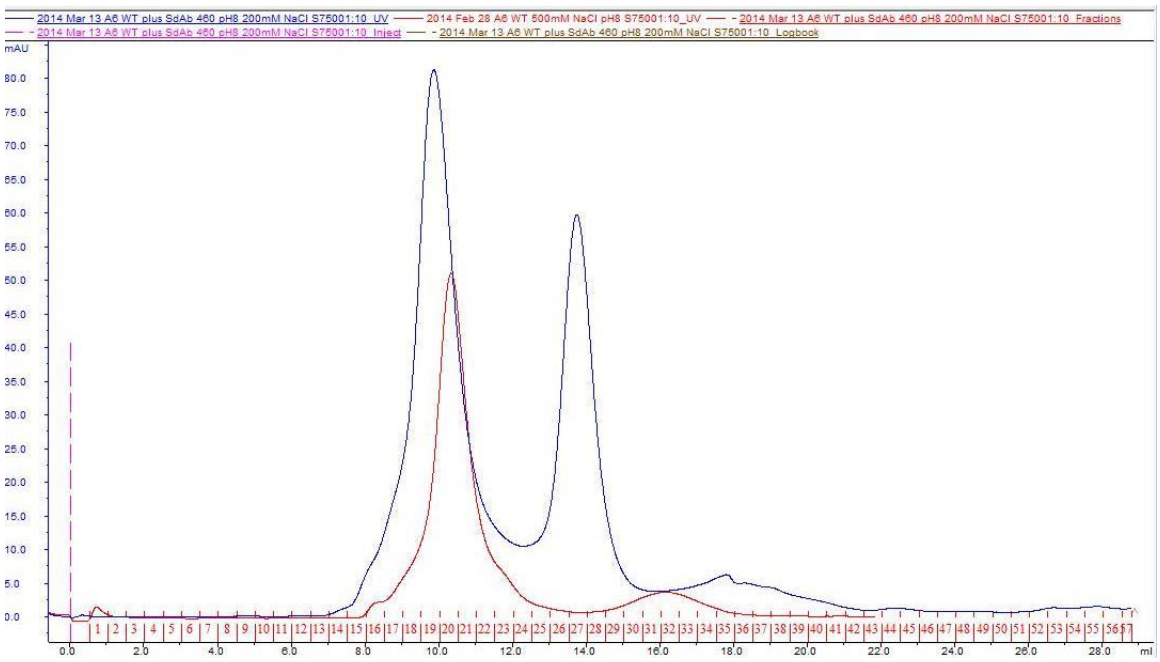


Figure 4-28: The comparison of size exclusion chromatographs of A6 wild type plus SdAb 458 complex on S75 (GE Health) column in 200 mM NaCl (blue) and 500 mM NaCl (red). The SDS-PAGE showing samples collected before injection and from corresponding fractions in 500mM NaCl.

Antibody 460 seems to have strong binding affinity to A6 wild type for it caused A6 peak shifted more than 458, and the ratio of A6 and 460 in the complex peak was consistent (Figure 4-29).



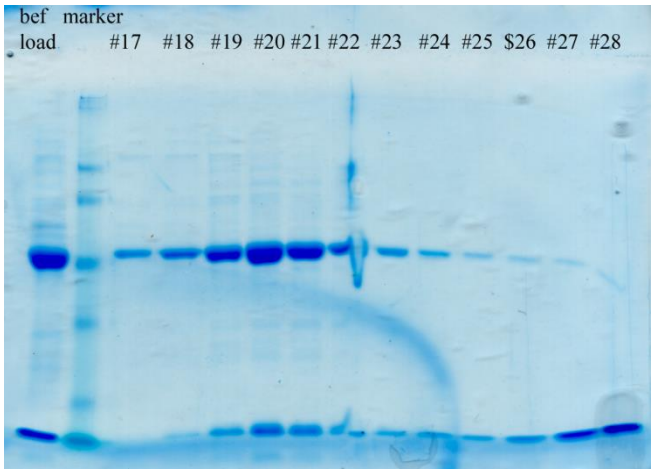
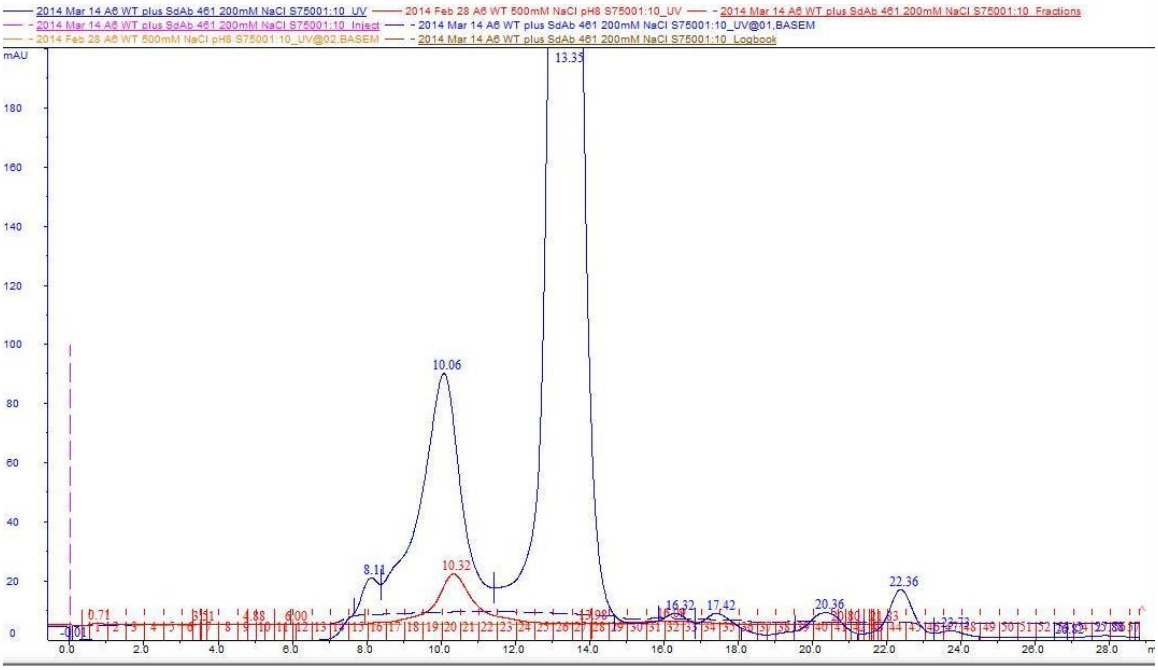


Figure 4-29: The size exclusion chromatograph of A6 wild type plus SdAb 460 complex on S75 (GE Health) column and SDS-PAGE showing samples collected before injection and from corresponding fractions.

Antibody 461 and A6 wild type do not bind, for the peak fractions do not contain any complex (Figure 4-30).



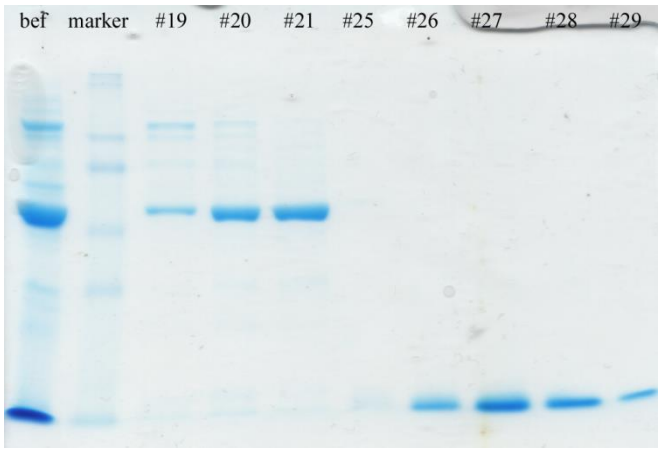


Figure 4-30: The size exclusion chromatograph of A6 wild type plus SdAb 461 complex on S75 (GE Health) column and SDS-PAGE showing samples collected before injection and from corresponding fractions.

Antibody 458 and A6 mutant AC seems to bind according to the SDS-PAGE gel, but the peak shift is not obvious on S75 chromatography, probably because Mutant AC itself is a dimer and its molecular mass already exceeds 75 kD (Figure 4-31).

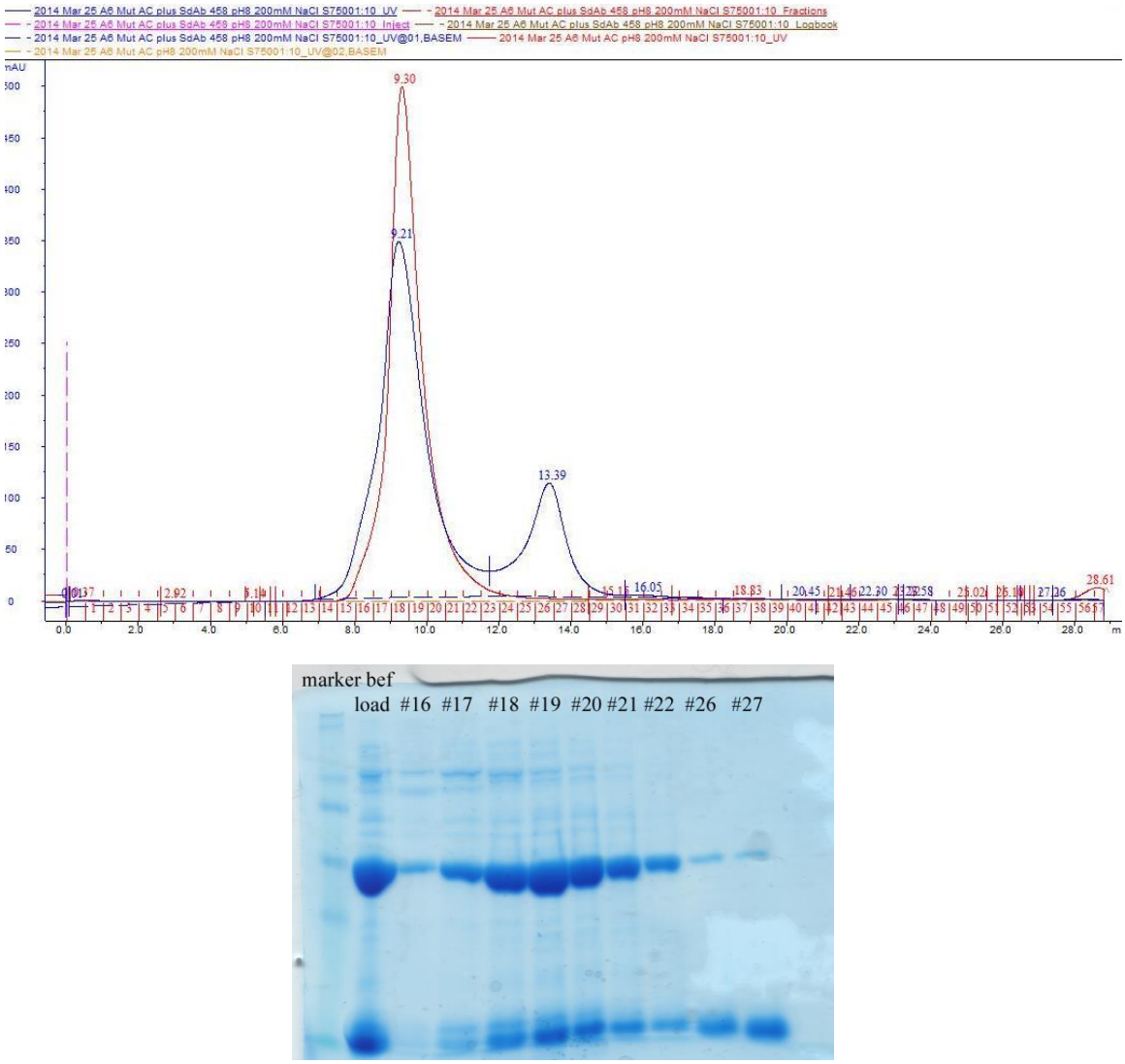


Figure 4-31: The size exclusion chromatograph of A6 mutant AC plus SdAb 458 complex on S75 (GE Health) column and SDS-PAGE showing samples collected before injection and from corresponding fractions.

A6 truncated mutant C and SdAb 458 have strong binding due to the complex peak shifts from A6 truncated mutant C peak, and the ratio of truncated mutant and 458 in the complex peak was consistent (Figure 4-32).

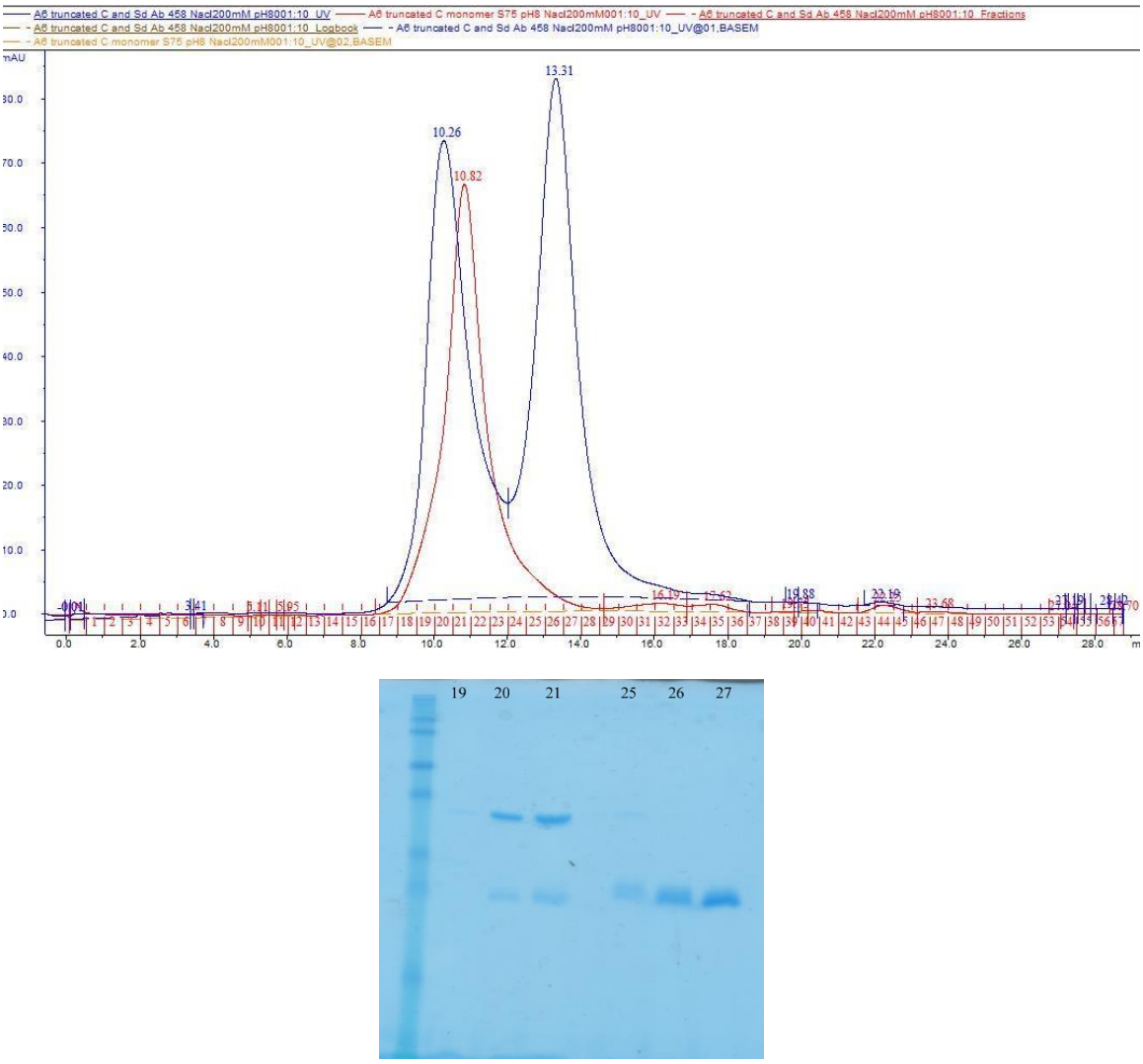


Figure 4-32: The size exclusion chromatograph of A6 truncated C plus SdAb 458 complex on S75 (GE Health) column and SDS-PAGE showing samples collected before injection and from corresponding fractions.

9. Homologues

M95 in Myxoma virus (Johnston, Barrett et al. 2005) and Y97 from Yaba monkey tumor virus (Brunetti, Amano et al. 2003) are two homologues of A6. M95 has 57% identities in with amino acid sequences and Y95 has 56%. We constructed these two proteins in pSUMO vector and expressed in *E. coli*.

pCG 66 (SUMO-M95) was purified through two step nickel and then run in SEC. The chromatograph showed two distinct peaks with calculated MW of 123 kD and 38.5 kD (Figure IV-33). Considering its real MW is 43kD, they might be a tetramer peak and a monomer peak. The nonreducing and reducing SDS gel showed that both peaks contained disulfide-bond-linked oligomers (Figure 4-34).

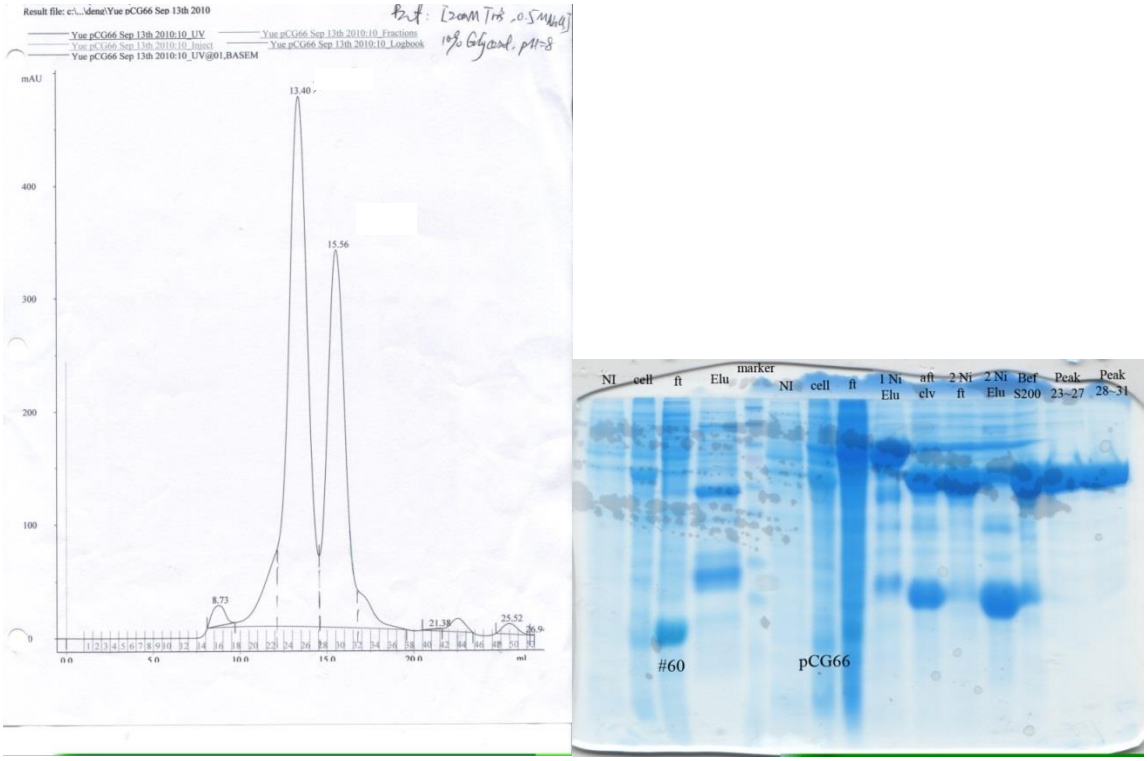


Figure 4-33: The analytical SEC of pCG66 chromatograph and the SDS-PAGE showing samples collected along purification.

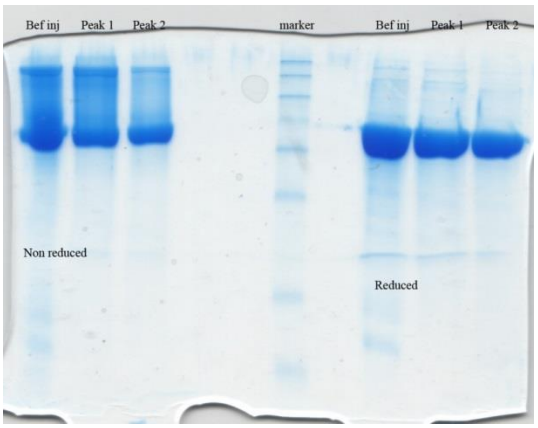


Figure 4-34: The nonreduced (left) and reduced (right) samples of pCG66 HiLoad before injection and two peaks.

Protein from the nickel purification was then supplied with 1 mM DTT before injected to HiLoad SEC. The chromatogram showed a single peak with calculated MW of 49 kD. Purified protein was concentrated to 6.8 mg/ml and used for crystallization, but no crystal was observed (Figure 4-35).

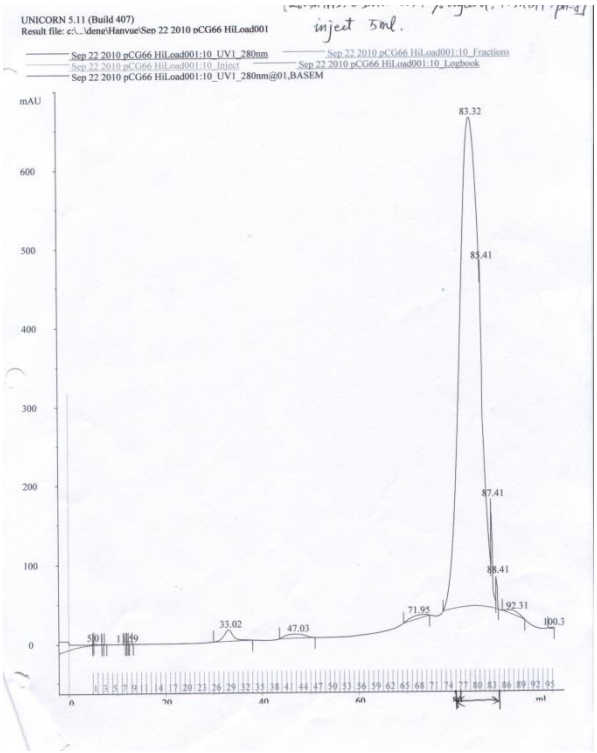
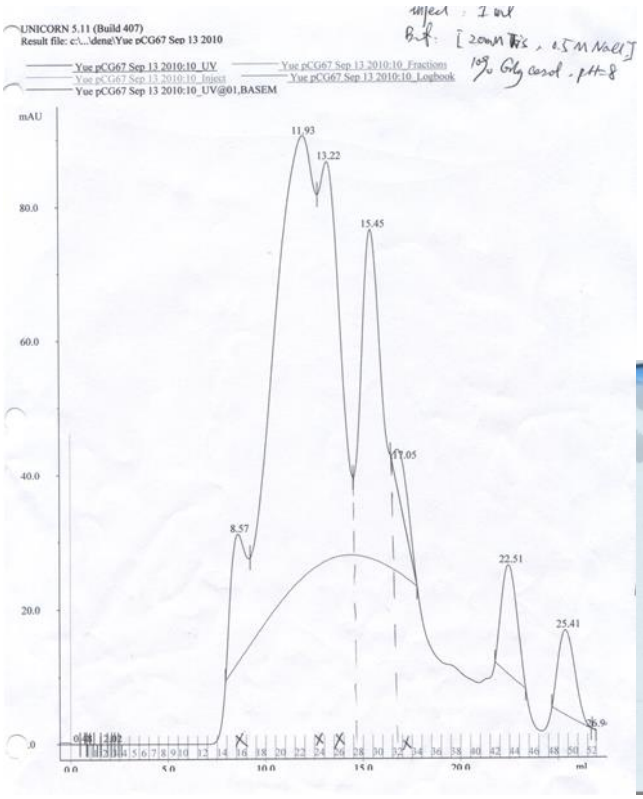
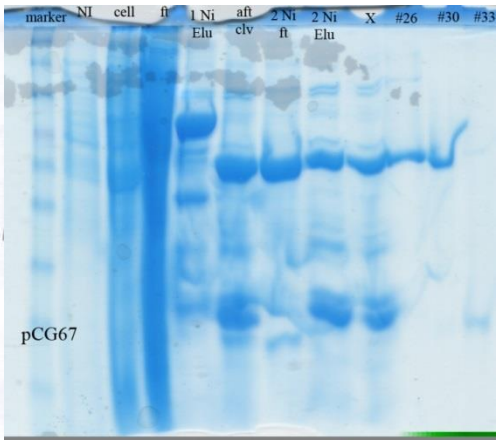
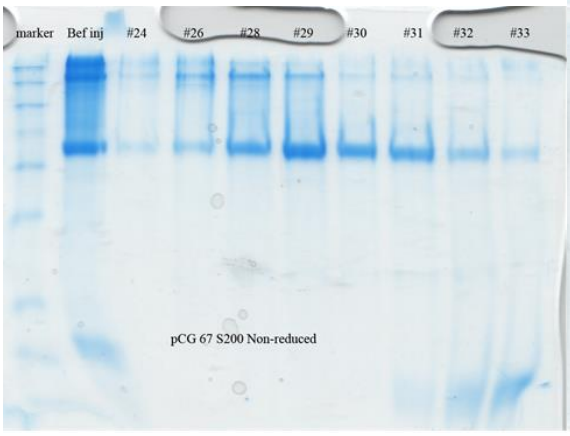


Figure 4-35: The HiLoad chromatograph of pCG66 containing DTT.

pCG 67 (SUMO-Y97) was purified through two step nickel and then run in SEC. The chromatograph showed multiple peaks with calculated MW of 271.8 kD, 135.5 kD and 40.7 kD. The nonreducing and reducing SDS gel showed that all peaks contained disulfide-bond-linked oligomers (Figure 4-36).







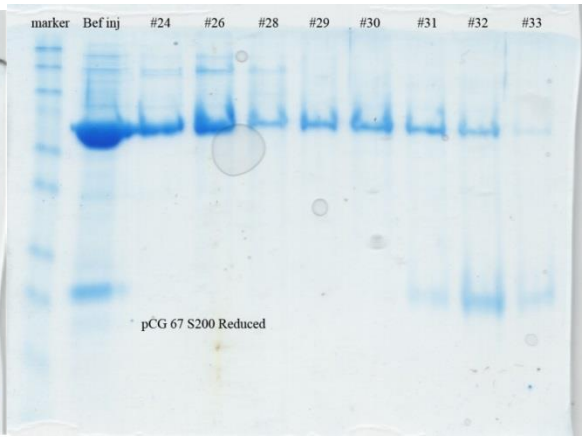


Figure 4-36: The analytical SEC of pCG67 chromatograph and the SDS-PAGE showing samples collected along purification.

DTT didn't improve the homogeneity of protein, for the 5 mM DTT containing sample presented the similar chromatograph on SEC (Figure 4-37).

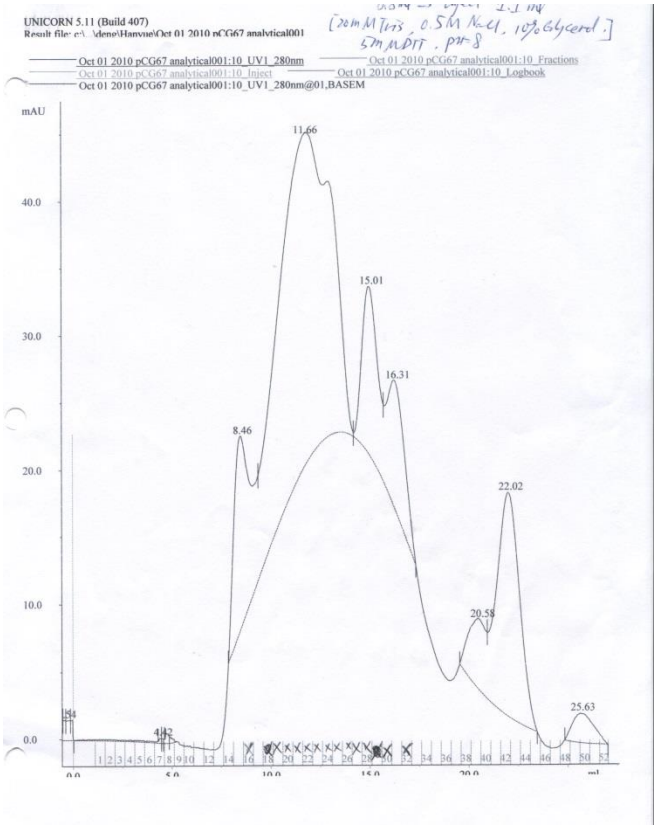


Figure 4-37: The HiLoad SEC of pCG67 chromatograph.

C. Discussion

1. Possible Reasons for Not Obtaining A6 Structure

The reason why A6 never formed nice diffracting crystals may not be known until its structure is finally determined. One thing we can propose is that it contains a lot of lysines, and thus needs to

overcome large surface entropy to crystalize. Our surface entropy reduction based protein engineering did promote crystallization, but the crystal packing was not ideal due to heterogeneity. Ion exchange chromatography, detergent washing and high salt washing were applied but none was able to improve the protein quality.

2. Possible Function Prediction

We identified two domains, N-terminal and C-terminal. Our collaborator found out both domains are essential for deletion of either disabled A6 function (unpublished data).

A11 is another protein required for poxvirus primary membrane forming. It associates with membrane and colocalizes with membrane proteins during virus replication. Its binding to A6 was proved by immunoprecipitation and its localization in cell alters in the presence or absence of A6. Moreover, without A6, A11 does not associate with membrane or colocalize with membrane proteins (Wu, Meng et al. 2012).

H7 is a protein critical for forming crescents and immature virions. It is expressed in cytoplasm but not packed into mature viruses (Satheshkumar, Weisberg et al. 2009). From collaborator's immunoprecipitation data (unpublished), A6 binds to H7.

A6 might bind to A11 and H7 directly, and together they assemble a protein complex responsible for membrane recruitment into viron factory.

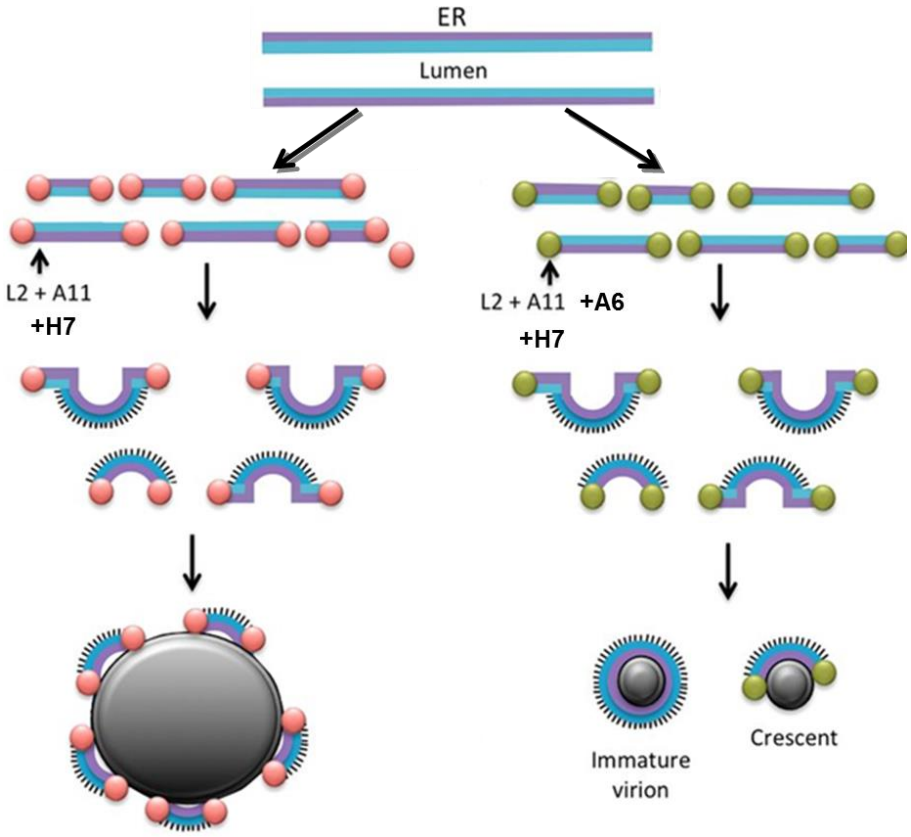


Figure 4-38: The depiction of viron membrane assembly model. Adopted from (Maruri-Avidal, Weisberg et al. 2013).

Under fluorescence microscope, our collaborator observed A6 travels in cell along microtubules into viron factories. It is reasonable to surmise that A6 binds to one of the microtubule motors such as kinesin or dynein.

3. Alternative Approach and Future Direction

The first choice is to exhaust all SdAb and A6 constructs combinations. SdAbs with strong binding affinity as well as weak binding affinity may assist A6 crystallization.

Next, MBP fusion protein may be worth looking into.

Maltose binding protein (MBP) infusion is another entropy reduction method. It can increases A6 expression level and solubility, as well as can encourage the crystal lattice formation when it is readily crystallizable itself. Moreover, once the crystal formed and data collected, MBP can provide the phase information since its structure is already known. Thus, we can solve the A6 structure by molecular replacement (Moon, Mueller et al. 2010).

We constructed MBP linked A6 or its mutants with a AAA linker in between. Nickel purification and size exclusion chromatography steps are the same as described above, except for the 40 mM maltose in the sizing column running buffer to prevent the nonspecific binding between MBP and column beads. Five mM maltose was kept in the final buffer to stabilize MBP (carried out by postdoctoral researcher Dr Bing Zhang in our lab). The protein was unable to express, possibly because the linker wasn't optimal.

The N-terminal domain of A6 is also worth looking into. Our postdoctoral researcher Bing Zhang recently crystalized it and now he is getting ready to collect data on it.

Once the A6 structure is determined, screening for function of the domains can be carried out as described in III.1. A6 binding positions also can be mapped to find out other virion protein binding sites.

Although Roc and A6 projects remain unclosed, I decided to move on to a new chapter of my life.

The years of PhD study is, however, rewarding.

ABBREVIATIONS

Book Part	Abbreviation
Clorobium tepidu.	C. tepidu
Escherichia coli.	E. coli
Escherichia coli.	E. coli
Size Exclusion Chromatagraphy	SEC
Melting temperature.	T_m
Surface Entropy Reduction.	SER
Fragment antigen-binding.	Fab
Single domain antibody.	SdAb
Wild type.	WT
Leucine-rich repeat kinase 2.	LRRK2
Guanine diphosphate.	GDP
Guanine triphosphate.	GTP
Molecular replacement	MR
Millimole	mM
Micromole.	μM
Multiple Anomalous Scattering.	MAD
Single Anomalous Scattering.	SAD
Multiple isomorphous replacement	MIR
Single isomorphous replacement.	SIR

REFERENCES

Aasly, J. O., M. Toft, I. Fernandez-Mata, J. Kachergus, M. Hulihan, L. R. White and M. Farrer (2005). "Clinical features of LRRK2-associated Parkinson's disease in central Norway." Ann Neurol **57**(5): 762-765.

Ayer, R. (1989). "Determination of unit cell." J Electron Microsc Tech **13**(1): 16-26.

Benard, V. and G. M. Bokoch (2002). "Assay of Cdc42, Rac, and Rho GTPase activation by affinity methods." Methods Enzymol **345**: 349-359.

Bergfors, T. (2007). "Succeeding with seeding: some practical advice. " Evolving Methods for Macromolecular Crystallography**.** Editors: Read, R. & Sussman, J. Springer Press, 1-10.

Betzenhauser, M. J., L. E. Wagner, 2nd, M. Iwai, T. Michikawa, K. Mikoshiba and D. I. Yule (2008). "ATP modulation of Ca2+ release by type-2 and type-3 inositol (1, 4, 5)-triphosphate receptors. Differing ATP sensitivities and molecular determinants of action." J Biol Chem **283**(31): 21579-21587.

Bigay, J., P. Deterre, C. Pfister and M. Chabre (1985). "Fluoroaluminates activate transducin-GDP by mimicking the gamma-phosphate of GTP in its binding site." FEBS Lett **191**(2): 181-185.

Biosa, A., A. Trancikova, L. Civiero, L. Glauser, L. Bubacco, E. Greggio and D. J. Moore (2013). "GTPase activity regulates kinase activity and cellular phenotypes of Parkinson's disease-associated LRRK2." Hum Mol Genet **22**(6): 1140-1156.

Blundell, T. L. and L. N. Johnson (1976). Protein crystallography. New York, Academic Press.
Carrington, J. C. and W. G. Dougherty (1988). "A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing." Proc Natl Acad Sci U S A **85**(10): 3391-3395.

Bosgraaf, L. and P. J. Van Haastert (2003). "Roc, a Ras/GTPase domain in complex proteins." Biochim Biophys Acta **1643**(1-3): 5-10.

Bower, J. H., D. M. Maraganore, S. K. McDonnell and W. A. Rocca (1999). "Incidence and distribution of parkinsonism in Olmsted County, Minnesota, 1976-1990." Neurology **52**(6): 1214-1220.

Brunetti, C. R., H. Amano, Y. Ueda, J. Qin, T. Miyamura, T. Suzuki, X. Li, J. W. Barrett and G. McFadden (2003). "Complete genomic sequence and comparative analysis of the tumorigenic poxvirus Yaba monkey tumor virus." J Virol **77**(24): 13335-13347.

Chen, V. B., W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallogr D Biol Crystallogr **66**(Pt 1): 12-21.

Chung, C. S., C. H. Chen, M. Y. Ho, C. Y. Huang, C. L. Liao and W. Chang (2006). "Vaccinia virus proteome: identification of proteins in vaccinia virus intracellular mature virion particles." J Virol **80**(5): 2127-2140.

Condit, R. C., N. Moussatche and P. Traktman (2006). In A Nutshell: Structure and Assembly of the Vaccinia Virion. Advances in Virus Research. M. Karl and J. S. Aaron, Academic Press. **Volume 66:** 31-124.

Conrath, K. E., M. Lauwereys, M. Galleni, A. Matagne, J. M. Frere, J. Kinne, L. Wyns and S. Muyldermans (2001). "Beta-lactamase inhibitors derived from single-domain antibody fragments elicited in the camelidae." Antimicrob Agents Chemother **45**(10): 2807-2812.

Cookson, M. R. (2010). "The role of leucine-rich repeat kinase 2 (LRRK2) in Parkinson's disease." Nat Rev Neurosci **11**(12): 791-797.

Cooper, D. R., T. Boczek, K. Grelewska, M. Pinkowska, M. Sikorska, M. Zawadzki and Z. Derewenda (2007). "Protein crystallization by surface entropy reduction: optimization of the SER strategy." Acta Crystallogr D Biol Crystallogr **63**(Pt 5): 636-645.

Deleage, G. and B. Roux (1987). "An algorithm for protein secondary structure prediction based on class prediction." Protein Eng **1**(4): 289-294.

Deng, J., P. A. Lewis, E. Greggio, E. Sluch, A. Beilina and M. R. Cookson (2008). "Structure of the ROC domain from the Parkinson's disease-associated leucine-rich repeat kinase 2 reveals a dimeric GTPase." Proc Natl Acad Sci U S A **105**(5): 1499-1504.

Derewenda, Z. S. and P. G. Vekilov (2006). "Entropy and surface engineering in protein crystallization." Acta Crystallogr D Biol Crystallogr **62**(Pt 1): 116-124.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-2132.

Gao, P. F., G. Q. Cao, H. T. Zhao, G. X. Zhang, Y. S. Jiang and Q. D. Wang (2009). "Molecular cloning and characterization of pigeon (Columba liva) ubiquitin and ubiquitin-conjugating enzyme genes from pituitary gland library." Int J Biol Sci **5**(1): 34-43.

Garnier, J., J. F. Gibrat and B. Robson (1996). "GOR method for predicting protein secondary structure from amino acid sequence." Methods Enzymol **266**: 540-553.

Geourjon, C. and G. Deleage (1994). "SOPM: a self-optimized method for protein secondary structure prediction." Protein Eng **7**(2): 157-164.

Geourjon, C. and G. Deleage (1995). "SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments." Comput Appl Biosci **11**(6): 681-684.

Giacovazzo, C. and D. Siliqi (2004). "Phasing via SAD/MAD data: the method of the joint probability distribution functions." Acta Crystallogr D Biol Crystallogr **60**(Pt 1): 73-82.

Goedert, M. (2001). "Alpha-synuclein and neurodegenerative diseases." Nat Rev Neurosci **2**(7): 492-501.

Gotthardt, K., M. Weyand, A. Kortholt, P. J. Van Haastert and A. Wittinghofer (2008). "Structure of the Roc-COR domain tandem of C. tepidum, a prokaryotic homologue of the human LRRK2 Parkinson kinase." EMBO J **27**(16): 2239-2249.

Greggio, E., S. Jain, A. Kingsbury, R. Bandopadhyay, P. Lewis, A. Kaganovich, M. P. van der Brug, A. Beilina, J. Blackinton, K. J. Thomas, R. Ahmad, D. W. Miller, S. Kesavapany, A. Singleton, A. Lees, R. J. Harvey, K. Harvey and M. R. Cookson (2006). "Kinase activity is required for the toxic effects of mutant LRRK2/dardarin." Neurobiol Dis **23**(2): 329-341.

Greggio, E., J.-M. Taymans, E. Y. Zhen, J. Ryder, R. Vancraenenbroeck, A. Beilina, P. Sun, J. Deng, H. Jaffe, V. Baekelandt, K. Merchant and M. R. Cookson (2009). "The Parkinson's disease kinase LRRK2 autophosphorylates its GTPase domain at multiple sites." Biochemical and Biophysical Research Communications **389**(3): 449-454.

Greggio, E., I. Zambrano, A. Kaganovich, A. Beilina, J. M. Taymans, V. Daniels, P. Lewis, S. Jain, J. Ding, A. Syed, K. J. Thomas, V. Baekelandt and M. R. Cookson (2008). "The Parkinson disease-associated leucine-rich repeat kinase 2 (LRRK2) is a dimer that undergoes intramolecular autophosphorylation." J Biol Chem **283**(24): 16906-16914.

Hamers-Casterman, C., T. Atarhouch, S. Muyldermans, G. Robinson, C. Hamers, E. B. Songa, N. Bendahman and R. Hamers (1993). "Naturally occurring antibodies devoid of light chains." Nature **363**(6428): 446-448.

Hanington, P. C., D. R. Barreda and M. Belosevic (2006). "A novel hematopoietic granulin induces proliferation of goldfish (Carassius auratus L.) macrophages." J Biol Chem **281**(15): 9963-9970.

Hassler, R. (1938) Zur Pathologie der Paralysis agitans und des postenzephalitischen Parkinsonismus. A second mutation in the α-synuclein gene (A30P) in a family with early-onset Parkinson's disease. J. Psychol. Neurol. 48, 387–486

Hengen, P. (1995). "Purification of His-Tag fusion proteins from Escherichia coli." Trends Biochem Sci **20**(7): 285-286.

Horowitz, M. P. and J. T. Greenamyre (2010). "Gene-environment interactions in Parkinson's disease: the importance of animal modeling." Clin Pharmacol Ther **88**(4): 467-474.

Hruby, D. E. (1990). "Vaccinia virus vectors: new strategies for producing recombinant vaccines." Clin Microbiol Rev **3**(2): 153-170.

Ito, G., T. Okai, G. Fujino, K. Takeda, H. Ichijo, T. Katada and T. Iwatsubo (2007). "GTP binding is essential to the protein kinase activity of LRRK2, a causative gene product for familial Parkinson's disease." Biochemistry **46**(5): 1380-1388.

Johnston, J. B., J. W. Barrett, S. H. Nazarian, M. Goodwin, D. Ricciuto, G. Wang and G. McFadden (2005). "A poxvirus-encoded pyrin domain protein interacts with ASC-1 to inhibit host inflammatory and apoptotic responses to infection." Immunity **23**(6): 587-598.

Johnson, L. L., D. A. Bornemeier, J. A. Janowicz, J. Chen, A. G. Pavlovsky and D. F. Ortwine (1999). "Effect of species differences on stromelysin-1 (MMP-3) inhibitor potency. An explanation of inhibitor selectivity using homology modeling and chimeric proteins." J Biol Chem **274**(35): 24881-24887.

King, R. D. and M. J. Sternberg (1996). "Identification and application of the concepts important for accurate and reliable protein secondary structure prediction." Protein Sci **5**(11): 2298-2310.

Lam, A. Y., E. Pardon, K. V. Korotkov, W. G. J. Hol and J. Steyaert (2009). "Nanobody-aided structure determination of the EpsI:EpsJ pseudopilin heterodimer from Vibrio vulnificus." Journal of Structural Biology **166**(1): 8-15.

Langston, J. W., P. Ballard, J. W. Tetrud and I. Irwin (1983). "Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis." Science **219**(4587): 979-980.

Langer, G., S. X. Cohen, V. S. Lamzin and A. Perrakis (2008). "Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7." Nat Protoc **3**(7): 1171-1179.

Lees, A. J., J. Hardy and T. Revesz (2009). "Parkinson's disease." Lancet **373**(9680): 2055-2066.

Levin, J. M. (1997). "Exploring the limits of nearest neighbour secondary structure prediction." Protein Eng **10**(7): 771-776.

Lewis, P. A., E. Greggio, A. Beilina, S. Jain, A. Baker and M. R. Cookson (2007). "The R1441C mutation of LRRK2 disrupts GTP hydrolysis." Biochem Biophys Res Commun **357**(3): 668-671.

Li, Y., L. Dunn, E. Greggio, B. Krumm, G. S. Jackson, M. R. Cookson, P. A. Lewis and J. Deng (2009). "The R1441C mutation alters the folding properties of the ROC domain of LRRK2." Biochim Biophys Acta **12**(7): 23.

Liao, J., C. X. Wu, C. Burlak, S. Zhang, H. Sahm, M. Wang, Z. Y. Zhang, K. W. Vogel, M. Federici, S. M. Riddle, R. J. Nichols, D. Liu, M. R. Cookson, T. A. Stone and Q. Q. Hoang (2014). "Parkinson disease-associated mutation R1441H in LRRK2 prolongs the "active state" of its GTPase domain." Proc Natl Acad Sci U S A **111**(11): 4055-4060.

Marin, I. (2006). "The Parkinson disease gene LRRK2: evolutionary and structural insights." Mol Biol Evol **23**(12): 2423-2433.

Maruri-Avidal, L., A. S. Weisberg and B. Moss (2013). "Direct Formation of Vaccinia Virus Membranes from the Endoplasmic Reticulum in the Absence of the Newly Characterized L2-Interacting Protein A30.5." Journal of Virology **87**(22): 12313-12326.

McFadden, G. (2005). "Poxvirus tropism." Nat Rev Micro **3**(3): 201-213.

Meng, X., A. Embry, L. Rose, B. Yan, C. Xu and Y. Xiang (2012). "Vaccinia virus A6 is essential for virion membrane biogenesis and localization of virion membrane proteins to sites of virion assembly." J Virol **86**(10): 5603-5613.

Meng, X. Z., A. Embry, D. Sochia and Y. Xiang (2007). "Vaccinia virus A6L encodes a virion core protein required for formation of mature virion." Journal of Virology **81**(3): 1433-1443.

Moon, A. F., G. A. Mueller, X. Zhong and L. C. Pedersen (2010). "A synergistic approach to protein crystallization: combination of a fixed-arm carrier with surface entropy reduction." Protein Sci **19**(5): 901-913.

Moss, B. (2013). "Poxvirus DNA replication." Cold Spring Harb Perspect Biol **5**(9).

Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr **53**(Pt 3): 240-255.

Nallamsetty, S. and D. S. Waugh (2007). "A generic protocol for the expression and purification of recombinant proteins in Escherichia coli using a combinatorial His6-maltose binding protein fusion tag." Nat. Protocols **2**(2): 383-391.

Paduch, M., F. Jelen and J. Otlewski (2001). "Structure of small G proteins and their regulators." Acta Biochim Pol **48**(4): 829-850.

Pardon, E., T. Laeremans, S. Triest, S. G. Rasmussen, A. Wohlkonig, A. Ruf, S. Muyldermans, W. G. Hol, B. K. Kobilka and J. Steyaert (2014). "A general protocol for the generation of Nanobodies for structural biology." Nat Protoc **9**(3): 674-693.

Paul-Dauphin, S., F. Karaca, T. J. Morgan, M. Millan-Agorio, A. A. Herod and R. Kandiyoti (2007). "Probing Size Exclusion Mechanisms of Complex Hydrocarbon Mixtures: The Effect of Altering Eluent Compositions." Energy & Fuels **21**(6): 3484-3489.

Polymeropoulos, M. H., C. Lavedan, E. Leroy, S. E. Ide, A. Dehejia, A. Dutra, B. Pike, H. Root, J. Rubenstein, R. Boyer, E. S. Stenroos, S. Chandrasekharappa, A. Athanassiadou, T. Papapetropoulos, W. G. Johnson, A. M. Lazzarini, R. C. Duvoisin, G. Di Iorio, L. I. Golbe and R. L. Nussbaum (1997). "Mutation in the alpha-synuclein gene identified in families with Parkinson's disease." Science **276**(5321): 2045-2047.

Rasmussen, S. G., H. J. Choi, J. J. Fung, E. Pardon, P. Casarosa, P. S. Chae, B. T. Devree, D. M. Rosenbaum, F. S. Thian, T. S. Kobilka, A. Schnapp, I. Konetzki, R. K. Sunahara, S. H. Gellman, A. Pautsch, J. Steyaert, W. I. Weis and B. K. Kobilka (2011). "Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor." Nature **469**(7329): 175-180.

Rayment, I. (1997). "Reductive alkylation of lysine residues to alter crystallization properties of proteins." Methods Enzymol **276**: 171-179.

Read, R. J. (2001). "Pushing the boundaries of molecular replacement with maximum likelihood." Acta Crystallogr D Biol Crystallogr **57**(Pt 10): 1373-1382.

Rossmann, M. G. (1990). "The molecular replacement method." Acta Crystallogr A **46 (Pt 2)**: 73-82.

Rost, B., C. Sander and R. Schneider (1994). "PHD--an automatic mail server for protein secondary structure prediction." Comput Appl Biosci **10**(1): 53-60.

Rupp, B. (2010). Biomolecular crystallography : principles, practice, and application to structural biology. New York, Garland Science.

Sakash, J. B. and E. R. Kantrowitz (2000). "The contribution of individual interchain interactions to the stabilization of the T and R states of Escherichia coli aspartate transcarbamoylase." J Biol Chem **275**(37): 28701-28707.

Satake, W., Y. Nakabayashi, I. Mizuta, Y. Hirota, C. Ito, M. Kubo, T. Kawaguchi, T. Tsunoda, M. Watanabe, A. Takeda, H. Tomiyama, K. Nakashima, K. Hasegawa, F. Obata, T. Yoshikawa, H. Kawakami, S. Sakoda, M. Yamamoto, N. Hattori, M. Murata, Y. Nakamura and T. Toda (2009). "Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease." Nat Genet **41**(12): 1303-1307.

Krüger, R., W. Kuhn, T. Müller, D. Woitalla, M. Graeber, S. Kösel, H. Przuntek, J. Epplen, L. Schols and O. Riess (1998). " Ala30Pro mutation in the gene encoding α-synuclein in Parkinson's disease." Nature Genet **18**: 106–108.

Satheshkumar, P. S., A. Weisberg and B. Moss (2009). "Vaccinia virus H7 protein contributes to the formation of crescent membrane precursors of immature virions." J Virol **83**(17): 8439-8450.

Schneider, T. R. and G. M. Sheldrick (2002). "Substructure solution with SHELXD." Acta Crystallogr D Biol Crystallogr **58**(Pt 10 Pt 2): 1772-1779.

Sen, S., P. J. Webber and A. B. West (2009). "Dependence of leucine-rich repeat kinase 2 (LRRK2) kinase activity on dimerization." J Biol Chem **284**(52): 36346-36356.

Smith, W. W., Z. Pei, H. Jiang, V. L. Dawson, T. M. Dawson and C. A. Ross (2006). "Kinase activity of mutant LRRK2 mediates neuronal toxicity." Nat Neurosci **9**(10): 1231-1233.

Taylor, G. (2003). "The phase problem." Acta Crystallogr D Biol Crystallogr **59**(Pt 11): 1881-1890.

Taylor, G. L. (2010). "Introduction to phasing." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 325-338.

Terwilliger, T. C. and J. Berendzen (1999). "Automated MAD and MIR structure solution." Acta Crystallogr D Biol Crystallogr **55**(Pt 4): 849-861.

Terwilliger, T. C. (2000). "Maximum-likelihood density modification." Acta Crystallogr D Biol Crystallogr **56**(Pt 8): 965-972.

Tesmer, J. J. G., D. M. Berman, A. G. Gilman and S. R. Sprang "Structure of RGS4 Bound to AlF4−-Activated Giα1: Stabilization of the Transition State for GTP Hydrolysis." Cell **89**(2): 251-261.

Thony, B., F. Neuheiser, L. Kierat, M. Blaskovics, P. H. Arn, P. Ferreira, I. Rebrin, J. Ayling and N. Blau (1998). "Hyperphenylalaninemia with high levels of 7-biopterin is associated with mutations in the PCBD gene encoding the bifunctional protein pterin-4a-carbinolamine dehydratase and transcriptional coactivator (DCoH)." Am J Hum Genet **62**(6): 1302-1311.

Tickle, I. J. and H. P. Driessen (1996). "Molecular replacement using known structural information." Methods Mol Biol **56**: 173-203.

Upton, C., S. Slack, A. L. Hunter, A. Ehlers and R. L. Roper (2003). "Poxvirus orthologous clusters: toward defining the minimum essential poxvirus genome." J Virol **77**(13): 7590-7600.

Vetter, I. R. and A. Wittinghofer (2001). "The guanine nucleotide-binding switch in three dimensions." Science **294**(5545): 1299-1304.

Vincke, C. and S. Muyldermans (2012). Introduction to Heavy Chain Antibodies and Derived Nanobodies. Single Domain Antibodies. D. Saerens and S. Muyldermans, Humana Press. **911:** 15-26.

Wagner, L. E., 2nd, M. J. Betzenhauser and D. I. Yule (2006). "ATP binding to a unique site in the type-1 S2- inositol 1,4,5-trisphosphate receptor defines susceptibility to phosphorylation by protein kinase A." J Biol Chem **281**(25): 17410-17419.

Walter, T. S., C. Meier, R. Assenberg, K. F. Au, J. Ren, A. Verma, J. E. Nettleship, R. J. Owens, D. I. Stuart and J. M. Grimes (2006). "Lysine methylation as a routine rescue strategy for protein crystallization." Structure **14**(11): 1617-1622.

VITA

YUE HAN

Candidate for the Degree of

Doctor of Philosophy

Thesis: THE STRUCTURAL STUDY OF PARKINSON DISEASE RELATED PROTEIN LRRK2 AND VACCINIA VIRUS PROTEIN A6

Major Field: Biochemistry and Molecular Biology

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Biochemistry and Molecular Biology at Oklahoma State University, Stillwater, Oklahoma in May, 2014.

Completed the requirements for the Master of Science in Biochemistry and Molecular Biology at Oklahoma State University, Stillwater, Oklahoma in May, 2012.

Completed the requirements for the Bachelor of Science in Life Science at University of Science and Technology of China, Hefei, Anhui, China in 2007.

Experience:

Doctoral Research Oklahoma State University	2008~2014
Post-graduate Research University of Science and Technology of China	2007~2008
Undergraduate Research Project University of Science and Technology of China	2006