# Crystallization Process of Protein Rv0731c from *Mycobacterium Tuberculosis* for a Successful Atomic Resolution Crystal Structure at 1.2Å

Liang Cong Zhu

Office of Science, Community College Institute Program

Contra Costa College

Lawrence Berkeley National Laboratory

Berkeley, California

August 13, 2008

Prepared in partial fulfillment of the requirement of the Office of Energy's Community College Institute under the direction of Minmin Yu in the Physical Biosciences Division at Lawrence Berkeley National Laboratory, and supported by P.I. Li-Wei Hung

Participant: _____

Signature

Research Mentor: _____

Signature

# Table of Content

# Introduction

Proteins are bio-macromolecules consisting of basic 20 amino acids and have distinct three-dimensional folds. They are essential parts of organisms and participate in every process within cells. Proteins are crucial for human life, and each protein within the body has a specific function, such as antibodies, contractile proteins, enzymes, hormonal proteins, structural proteins, storage proteins and transport proteins [1]. Determining three-dimensional structure of a protein can help researchers discover the remarkable protein folding, binding site, conformation and etc, in order to understand well of protein interaction and aid for possible drug design.

The research on protein structure by X-ray protein crystallography carried by Li-Wei Hung's research group in the Physical Bioscience Division at Lawrence Berkeley National Laboratory (LBNL) is focusing on protein crystallography. The research in this lab is in the process of from crystallizing the proteins to determining the three dimensional crystal structures of proteins. Most protein targets are selected from *Mycobacterium Tuberculosis.* TB (Tuberculosis) is a possible fatal infectious disease. By studying TB target protein can help discover antituberculer drugs, and find treatment for TB. The high-throughput mode of crystallization, crystal harvesting, crystal screening and data collection are applied to the research pipeline (Figure 1). The X-ray diffraction data by protein crystals can be processed and analyzed to result in a three dimensional representation of electron density, producing a detailed model of protein structure.

Rv0731c is a conserved hypothetical protein with unknown function from *Mycobacterium Tuberculosis*. This paper is going to report the crystallization process and brief structure information of Rv0731c.

# Materials and Methods

## Protein material

Amount of 500 ul of 20.6 mg/ ml purified Rv0731c protein was received from LANL (Los Alamos National Laboratory).

## Crystallization process

### *a. Random screen plates*

Two sets of five screen random plates, including Hampton I&II (Hampton Research crystal screen and Hampton Research crystal screen 2), Wizard I&II (Emerald Biosystem Wizard I random sparse matrix crystallization screen and Emerald Biosystem Wizard II random sparse matrix crystallization screen) , Wizard III_PegIon (Emerald Biosystem Wizard III random sparse matrix crystallization screen and Hampton Research PEG/Ion screen), EM_64 (Emerald Biosystems Precipitant Synergy primary 64 crystallization screen) and Membfac_CSLite (Hampton Research Membfac screen and Hampton Research Crystal Screen Lite), were setup for a target Rv0731c protein. Intelli-Plate, a 96 well sitting drop vapor diffusion crystallographic plate with three droplets and a screen reservoir, was used for setting up random plates. A robotic device Phoenix (manufactured by Art Robbins Instruments) was used to setup random plates automatically. The droplet size was total 150nl which is mixed with 1:1 ratio of protein:reservoir and the reservoir volume is 50ul. The setup concentration of Rv0731c was 20.6 mg/ml. Random plates were stored in room temperature (22 °C). Total 12 initial crystal hits were found under 448 unique crystallization conditions.

### *b. Database entry and labeling*

These random plates then were entered into TBSGC (*Mycobacterium Tuberculosis* Structural Genomics Consortium) database. In the TBSGC database, all the information of Rv0731c, such as buffer type, original concentration, setup concentration, drop size and storage temperature, were recorded for an organized bookkeeping system. A plate ID is provided as the plate is entered into the database. The newly setup random plates are labeled with plate name and plate ID after being entered into the database.

### *c. Plate scanning and scoring*

Random plates were regularly scanned by the Cryscam, a robotic image viewing device, according to the viewing schedule. The images of each droplet containing protein sample were taken and stored in a local network system within LBNL and were also updated every 24 hours to posted automatically in the TBSGC database with updates of every 24 hours. These images were scored through complementary software called Cryscore according to the scoring scale (Table 1) to find any initial crystal hits. Any droplets with score 6 or above were viewed under the Nikon microscope to confirm initial crystal hits.

### *d. Optimization plates*

With the confirmed initial crystal hit, eight optimization plates were setup for the Rc0731c based on 2 out of 12 initial crystal hit conditions. For optimization plates, the Greiner-plates were used instead with total 1.0 μL droplet size. A Greiner-plate is divided into four different quadrants with various buffer pH values and gradient concentrations of precipitants and salts for improving crystal quality.  Most improved crystals were from

optimization conditions of 0.1M Tris pH 8.0-8.5, 30% PEG 3350, and 15-30% isopropanol.

### Crystal harvesting and data collection

The crystals with score 8 and 9 were harvested with cryo-protectants using Hampton loop under the high magnified microscope and frozen in the liquid nitrogen. Crystal samples were screened at both beamline in Advance Light Source (ALS) and Stanford Synchrotron Radiation Laboratory (SSRL) for X-ray diffraction. The Br soaking crystal has the cryo condition of 80% mother liquor + 20% glycerol + 1N NaBr soaking for 3 mins, and the native crystal has a cryo condition of mother liquor + 20% glycerol. The Br MAD data were collected under multiple wavelengths of peak, inflection and high remote energy (Figure 2). The data were good to 1.8Å (Figure 3a). The native crystal was diffracted initially to 1.0Å, and the dataset was good at 1.2Å of an atomic resolution. (Figure 3b). The structure solution from Br MAD (Multiwavelengths anomalous diffraction) data provided an interpretable electron density map of a three dimensional protein model.

## Results and Discussions

Rv0731c is a conserved hypothetical protein with unknown function. This protein is from *Mycobacterium Tuberculosis* which has 318 amino acids with molecular weight of about 35 kDalton. The protein sequence of Rv0731c is shown in the Table 2.

Crystallization process of Rv0731c was from 5x2 random plates with 448 unique crystallization conditions. There were total 12 random crystal hits were found from these random screen plates. The data of random crystal hits are listed in the Table 3.

Based on 2 conditions of initial crystal hits, 8 optimization plates were setup for Rv0731c. Crystals with scores 8 or better from both random and optimization plates were harvested and screened through X-ray diffraction in the ALS beamline. This structure was determined quickly within 20 crystals being screened at synchrotron. The best diffraction was improved from 2.4 Å of initial crystal hits to 1.0 Å after crystal optimization. The X-ray diffracted resolution was improved nicely to an atomic resolution level.

The Br soaking crystal MAD data were collected at multi-wavelengths of peak, inflection and high remote energy. MAD data were processed well to 1.8Å. The native crystal dataset was good to an atomic resolution of 1.2 Å. The initial crystal structure was determined from Br soaking MAD data and native data which was good to extend the phasing to an atomic resolution of 1.2Å. The structure refinement is underway.

The three dimensional protein structure of Rv0731c was determined with Br soaking MAD method. The initial crystal structure shows several β-strands are surrounded by multiple α-helixes (Figure 4). Rv0731c is a protein from *Mycobacterium Tuberculosis* with unknown function. The sequence homologous blast search did not give any convincing hits lower than $e^{-3}$ value. Once the protein model building is more completed along with the ongoing structure refinement, more structure analysis will be done and hopefully it may help for protein function study.

# Reference

[1] S. E. Manahan, Environmental science and technology, CRC Press, 1997.

# Acknowledgements

I would like to give my sincere appreciations to the following individuals and organization for giving me this precious opportunity to do research at Lawrence Berkeley National Laboratory:

My mentor, Dr. Minmin Yu
Principal Investigator, Dr. Li-Wei Hung

Brian Jan, Andy Zhang and other lab members in Hung's research group

Structural genomics consortiums:
LANL protein production facility
ISFI & TBSGC members

Staff at BCSB in ALS and staff at SMB in SSRL

Supporting organizations:
CSEE in LBL
Community College Initiative Internship Program
Center for Science Excellence at Contra Costa College
NIH/NIGMS
National Science Foundation
Contract Number for Berkeley Lab (DE-AC02-05CH11231)

Table 1. Scoring Scale

| 0: clear drop | 1: unable to accurately record or salt crystal | 2: light precipiate | 3: heavy precipitate | 4: phase separation | 5: organized precipitate | 6: crystal cluster | 7: microcrystal | 8: small single crystal (=<100um) | 9: large single crystal (>100um) |
|---|---|---|---|---|---|---|---|---|---|

Table 2. Amino Acid Sequence of Rv0731c

| |
|---|
| MTQTGSARFEGDSWDLASSVGLTATMVAAARAVAGRAPGALVND<br>QFAEPLVRAVGVDFFVRMASGELDPDELAEDEANGLRRFADAMAI<br>RTHYFDNFFLDATRAGIRQAVILASGLDSRAYRLRWPAGTIVFEVD<br>QPQVIDFKTTTLAGLGAAPTTDRRTVAVDLRDDWPTALQKAGFDN<br>AQRTAWIAEGLLGYLSAEAQDRLLDQITAQSVPGSQFATEVLRDIN<br>RLNEEELRGRMRRLAERFRRHGLDLDMSGLVYFGDRTDARTYLAD<br>HGWRTASASTTDLLAEHGLPPIDGDDAPFGEVIYVSAELKQKHQDTR |

Table 3. Random Crystal Hits

| plate number | well | crystal type | screen type | screen tube | crystallization condition |
|---|---|---|---|---|---|
| 600077 | B10L | single crystal | HP_I_II | HP_I, #22 | 0.1M Tris pH 8.5, 30% PEG 4000, 0.2M NaAc |
| 600106 | B2L | needle clusters | WZ_I_II_r2 | WZ_I, #14 | 0.1M cacodylate pH 6.5, 1M sodium citrate |
| 600106 | B8L | needle clusters | WZ_I_II_r2 | WZ_I, #20 | 0.1M imidazole pH 8.0, 0.4M Na $H2PO_4$ , 1.6 M K $H2PO_4$, 0.2M NaCl |
| 600106 | E4L | needle clusters | WZ_I_II_r2 | WZ_II, #4 | 0.1M cacodylate pH 6.5, 2.0M $(NH_4)_3SO_4$, 0.2M NaCl |
| 600106 | G7L | needle clusters | WZ_I_II_r2 | WZ_II, #31 | 0.1M Tris pH 7.0, 1M sodium citrate, 0.2M NaCl |
| 600108 | A03L | needle clusters | EM_64screen_r1 | EM_64screen, #3 | 0.1M HEPES pH7.5, 2M $(NH_4)_2SO_4$, 1% MPD |
| 600108 | A4L | needle clusters | EM_64screen_r1 | EM_64screen, #4 | 0.1M Tris pH 8.5, 5% PEG 400, 2M $(NH_4)_2SO_4$, 0.05M MgSO4, |
| 600108 | A9L | needle cluster | EM_64screen_r1 | EM_64screen, #9 | 0.1M Na/K phosphate pH 6.5, 2% PEG 400 |
| 600108 | B6L | needles | EM_64screen_r1 | EM_64screen, #18 | 0.1M Tris pH 8.5, 2% PEG 400, 2M $Li_2SO_4$ |
| 600108 | B10L | microcrystals | EM_64screen_r1 | EM_64screen, #22 | 0.1M imidazole pH 6.5, 30% isopropanol, 1.3M NaCl, 0.1M $CaCl_2$ |
| 600108 | C12B | plates | EM_64screen_r1 | EM_64screen, #36 | 0.1M Tris pH 8.5, 30% PEG 3350, 30% isopropanol |
| 600108 | D5L | microcrystals | EM_64screen_r1 | EM_64screen, #41 | 0.1 M Tris pH 8.5, 25% PEG 400, 20% PEG 3350, 0.1M $MgCl_2$ |
| 600109 | C12L | needle clusters | Membfac_CSLite | Membfac, #36 | 0.1M HEPES pH 7.5, 1M sodium citrate , 0.25M $NaH_2PO_4$ |

Figure 1. Crystallographic pipeline in this lab

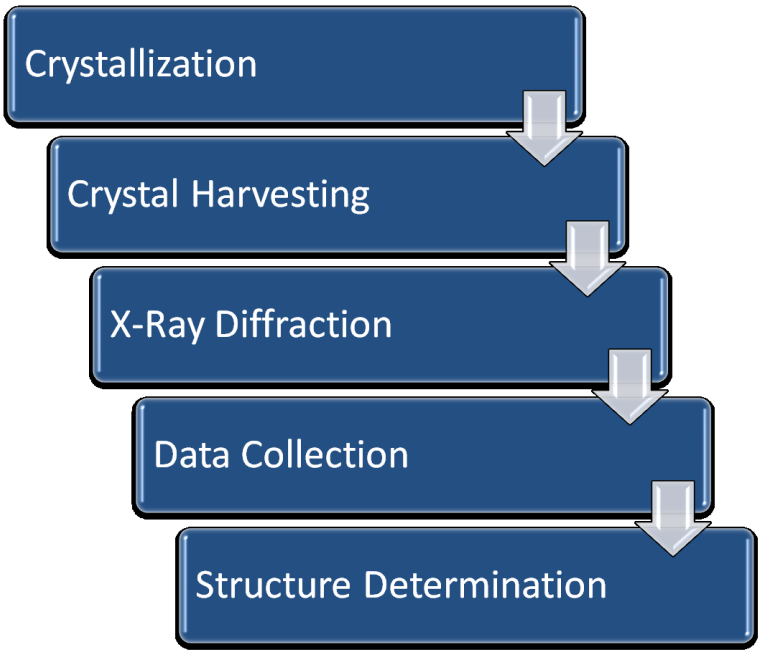


Figure 2. Br Fluorescence Scan Diagram

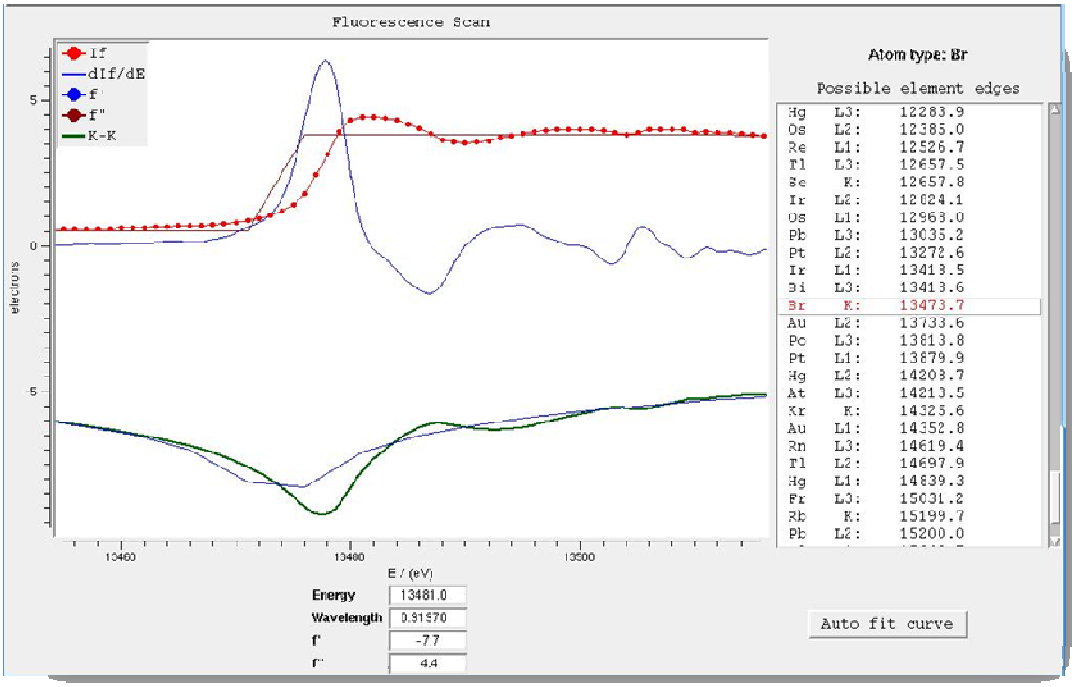
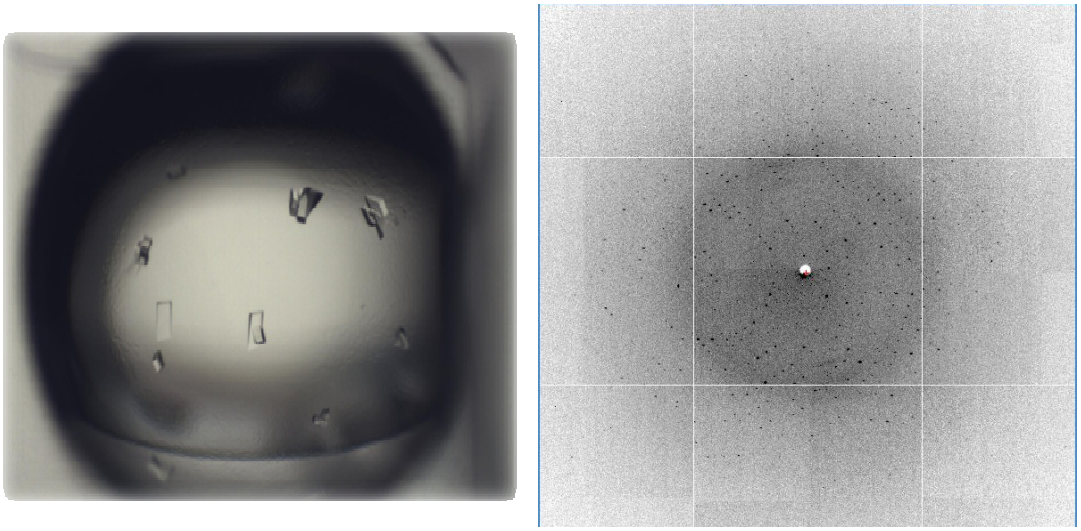
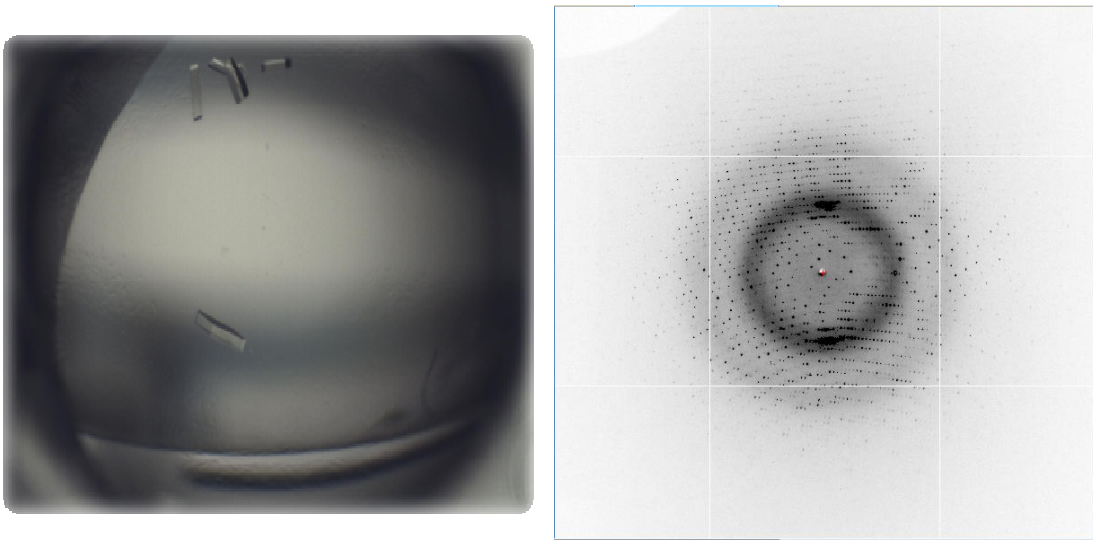


Figure 3: Rv0731c Crystals and X-ray diffraction images



a) Br soaking crystal for data collection and corresponding diffractoin image



b) Rv0731c native crystal for data collection and corresponding diffractoin image

Figure 4. Initial Structure solution of Protein Rv0731c