PRODUCTION OF SINGLETON PROTEINS (EXPRESSION, PURIFICATION AND CRYSTALLIZATION) FOR STRUCTURAL GENOMICS

by

RAMIAH ANNAPOORANI

(Under the direction of Bi-Cheng Wang)

ABSTRACT

The main objective of this thesis was to prepare samples (soluble proteins and crystals) of selected ORFan proteins for structure determination by X-ray crystallography. ORFan proteins are the proteins having no homologous sequence with any other proteins in other known organisms. Hence for ORFan proteins, structure prediction is not reliable and the only possible way to identify the protein family, fold is the determination of their three-dimensional structures.

In this work, a select set of ORFan proteins from *Pyrococcus furiosus* (ORF PF0772), *Aeropyrum pernix* K1 (ORFs AP0305, AP0436, AP0371) and *Clostridium thermocellum* and (ORF Cthe_3042) have been expressed, purified and crystallized. Several crystallization techniques were employed to improve the quality of the diffraction of the crystals. Both seleneo-methionyl (Se-Met) incorporation and heavy atom soaking were used to provide initial phase information. Diffraction quality crystals were obtained for AP0305, AP0436, AP0371 and Cthe_3042. The structure of Cthe_3042 has been determined to 2.1Å resolution using Se-Met labeled protein.

INDEX WORDS: ORFan proteins, singleton, *Pyrococcus furiosus, Aeropyrum pernix*K1, *Clostridium thermocellum, expression, purification and crystallization.*

PRODUCTION OF SINGLETON PROTEINS (EXPRESSION, PURIFICATION AND CRYSTALLIZATION) FOR STRUCTURAL GENOMICS

by

RAMIAH ANNAPOORANI

M.S., MADURAI KAMARAJ UNIVERSITY, INDIA, 1994

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2007

PRODUCTION OF SINGLETON PROTEINS (EXPRESSION, PURIFICATION AND CRYSTALLIZATION) FOR STRUCTURAL GENOMICS

by

RAMIAH ANNAPOORANI

Major Professor:	Bi-Cheng Wang
Committee:	John Rose William Lanzilotta

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December, 2007

DEDICATION

I dedicate this to my beloved husband Dr. Rathinam Viswanathan who always guided me through difficult times during my stay at UGA. He took care of everything including cooking and provided food for me and my daughter.

I also dedicate this to my daughter Suvitha Viswanthan, who shared my frustrations and often asked me “mommy did your crystal diffract?” whenever I talked about my work to my husband. She asks me to show my crystal pictures to her and try to understand what I was doing. She also learned to take of care of herself and gave me plenty of time to stay late at lab.

Finally to my parents and to my in-laws who were very supportive through out my studies and in my life.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Dr. Bi-Cheng Wang for his support to complete my project. I would like to thank my committee members Dr. John Rose and Dr. William Lanzilotta for their suggestions and support.

I would like to thank Dr. John Ruble and Dr. Lirong Chen for their continuous help in data collection, and processing the data. I would also like to thank Dr. Peter Horiyani for his suggestions in my work.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

1.1 Purpose of Study

Structural Genomics (SG) is the building blocks of **i)** Bioinformatics, to choose the targets for SG, **ii)** Molecular biology, to clone and express the interested gene in a *E.coli* host and to produce recombinant protein, **iii)** Biochemistry, to purify and characterize the protein, to understand the function of the protein and **iv)** Structure determination either by means of Nuclear Magnetic Resonance (NMR) or by X-ray Crystallography techniques [1]. Determination of the three dimensional structure of a protein helps us to find its family, to discover more new protein fold [2]. The new structure can be used as template model for the protein whose structure has to be determined.

BLAST analysis (http://www.ncbi.nlm.nih.gov/BLAST/) of the genes, PF0772 from *Pyrococcus furiosus*, AP0305, AP0436, and AP0371 from *Aeropyrum pernix* K1, Cthe_3042 from *Clostridium thermocellum* showed them to be hypothetical proteins, having a sequence identity of less than 30% when compared to the structural template. These proteins do not have any sequence homology with any other proteins known in any organisms, and not have any assigned function. Therefore such proteins are referred to as unique proteins or otherwise known as ORFans [3-5]. Structure determination of ORFans is important since their structures are likely to represent a new fold [6]. Subsequent structure-based analyses might help in identifying the functions. In order to learn these; a select set of proteins from three organisms (*Pyrococcus furiosus* [7, 8], *Aeropyrum pernix* K1 [9] and *Clostridium thermocellum)* should be determined.

The questions can then be addressed whether these proteins have unique fold or an already known fold with new function? Hence, as part of on-going efforts to determine the X-ray crystal structures at the Southeast Collaboratory for Structural Genomics at the University of Georgia, Athens, this dissertation reports the crystallization and preliminary studies by structural studies on aforementioned ORFan proteins.

1.2 Back Ground

Structural Genomics

At present, there are more than 44,700 protein structures in the Protein Data Bank (PDB, http://www.rcsb.org/pdb). However, only 1054 of these structures correspond to a unique protein fold. Prediction of structure and function of a target protein is difficult when its sequence show lower, less than 30% in general, sequence identity to the known structural template [10]. One of the aims of Structural Genomic programs is to populate "protein fold space" by determining the structures of unique proteins of organisms. If the new determined structure has novel fold then this structure can be a starting model for a protein family with no known structure [11, 12]. If it does not have novel fold, and has already known fold, then it helps us to analyze what might be the cause for the sequence to diverge.

Structures of different proteins show structural or fold similarity in spite of low sequence homology between them. For e.g., the structure determination of hypothetical ORFan protein PF0725 suggests that it is CoA-binding protein which is inferred from the presence of CoA (coenzyme A) by electron density map analysis, and by the structure similarity to the *Thermus thermophilus* CoA binding protein TT 1466/TTHA1899 [13]. The structure determination of PF0899 hypothetical protein suggests it may be structural protein by having similar wedge-shaped domain as that found in capsid protein from bacteriophage HK97 [14]. Structure

determination of hypothetical ORFan protein from Thermotoga maritima TM0875 has significant structural homology with hypothetical protein YggU (PDB code 1n91) in spite of their sequence identity is only 7% [15]. Sperm-whale myoglobin and horse hemoglobin show structural similarity even though the identity between the two sequences is very low [16]. Similarly, the sequence similarity between actin and the ATPase fragment of heat shock protein is very low [17]. However, the structure determination of these proteins revealed to have structure similarity. Also, functions like, molecular mechanism, active site residues, the electron donor & acceptor pathway and transport of molecules etc., could be better explained by determining the structure of the protein [10, 18, 19].

Introduction of high-throughput technology in X-ray crystallography speeds up the process of structural determination [18, 20, 21]. According to Joint Centre for Structural Genomics, using high-throughput technology, as many as 130,000 proteins can be analyzed for crystallization per day from all the structural genomic centers. The high-throughput structural genomics allow us to select multiple targets at the same time to work in parallel, thus leaving handful of targets for crystallization, which requires protein in milligram quantities.

Apart from the advantages of determining the structures of a protein, there is also a opinion, that the determination of structures once had a great impact on understanding the structure of protein and infer function from the structure, now seemed to have only little effect on providing information, thus determination of more structures is not necessarily required as we expected [22]. Hence, it is not enough to work only on structural determination, SG should also focus on protein function.

ORFan Proteins

With the advent of whole genome sequencing techniques, complete list of Open Reading Frames (ORFs), coding regions of the genomes are available for several microorganisms [5]. These completely sequenced genomes also contain families with little or no known function [23]. These sequences are defined as singleton ORFans or orphan ORFs which has no detectable homologous sequence in the database [24]. Three types of ORFans are observed **i)** Singleton ORFans which is denoted as an ORF with no homology to any other protein **ii)** Paralogous ORFans which is described as an ORF with homology to proteins of the same genome only **iii)** Orthologous ORFans are the ORFs with homology only in closely related organisms only [5, 25]. The possible reasons for the existence of ORFans may be due to misannotation of genes as protein encoding genes or they may belong to superfamilies which are undetectable with current sequence comparison programs or may be existing only in a particular organism [3]. According to the length distributions in bacterial genomes, the short non coding ORFs were once considered as true ORFans i.e. sequence with 150 residues and lower than that [26]. As the evolution goes on, structure determination is highly conserved than sequence. The structure determination ORFan targets help to discover novel folds which are characteristic to unknown superfamilies.

In 1990, a universal phylogenic tree was proposed by Woese *et al*., consists of three domains: the Bacteria, the Archaea, and the Eukarya [27].

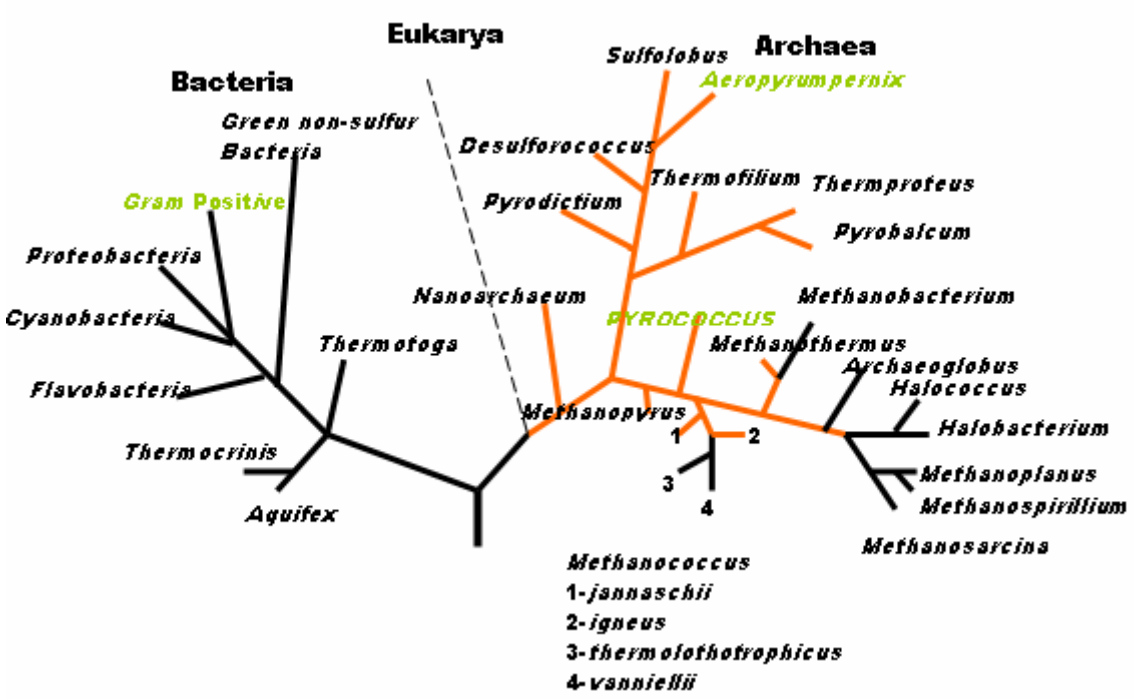


Figure 1.1 Universal Phylogenetic tree.
Modified from Stetter *et al*., 1992 and Woese *et al*., 1990

Archaea can be divided into two distinct kingdoms: the Crenarchaeota and the Euryarchaeota. The Euryarchaeota consists of methanogens (extreme halophiles, sulfate reducing species) and extreme thermophiles including pyrococcus (shown as red in Figure1.1). The Crenarchaeota consists of thermoacidophiles and sulfur dependant archaebacteria.

1.3 Hyperthermophiles and Thermophiles

The hyperthermophiles and thermophiles classified under extremophiles. Extremophiles are microorganisms which can thrive in extreme conditions of temperature, pH, and salinity [28]. The hyperthermophilic microorganisms were isolated from chimney smoker walls in the early 1980's by Stetter and coworkers [29]. These hyperthermophiles prefer to grow at the temperatures of 80ºC and above. *Pyolobus fumarii* found at the walls of deep-sea hydrothermal

vents can grow at the temperature as high as 113°C [30]. Most of the hyperthermophiles are found in deep-sea levels and are strictly anaerobes [31] except *Aeropyrum pernix* K1 and *Pyrobalcum* , are aerobic and can tolerate the oxygen up to 5% (v/v) [32, 33]. For our studies we have selected two hyperthermophilic archaeon namely *Pyrococcus furiosus* and *Aeropyrum pernix* K1.

Pyrococcus furiosus was first isolated from geothermally heated soils and hydrothermal vents off the coast of Vulcano, Italy, in 1986 by Stetter and his coworkers. It is spherical in shape 0.8 to 2.5μm in width and has monopoloar polytrichous flagellation [34]. Poole et al., deposited 2065 ORFs in 2002 [8] and out of this 19.0% (375) are conserved hypothetical and 6% as ORFans which are unique to pyrococcus furiosus.

Aeropyrum pernix K1 is an aerobic hyperthermophilic crenarchaeon. It was isolated from coastal solfataric thermal vent in Kodakara-jima Island in Kagoshima, Japan. It grows at the temperature of 90-95°C [33]. So far, 2700 ORFs are reported in this genome and out of this 49.4% are ORFans.

Clostridium thermocellum is an anaerobic, thermophilic celluloytic and ethanogenic bacterium, prefers to grow at the temperature range of 45-80ºC, which can convert cellulosic substrate into ethanol [35]. According to TIGR annotation summary there are totally 5744 genes available and about 51.7% (2934) are hypothetical ORFan genes. Presence of considerable number of ORFans in the above three organisms allowed us to pick targets for our studies.

CHAPTER 2

MATERIALS AND METHODS

2.1 Materials

Luria Bertani (LB), (Fischer Biotech.), Bacto Agar and Bacto yeast (Dickinson & Co.), SDS-PAGE molecular weight markers (Bio-Rad), Isopropyl-beta-D- thiogalactopyranoside (Biosynth), Para formaldehyde 16% solution (Electron microscopy sciences) and Dimethylamine-borane complex (DMAB), (Aldrich).

Media solution preparation: **LB** (1L): 20 g

Rich media (1L): 16g Bacto Agar, 10g Bacto yeast, 5g NaCl

1000x Metal Mix (1L): 0.1M $FeCl_3.6H_2O$, 1M $CaCl_2$, 1M $MnCl_2$, 1M $ZnSO_4 \cdot 7H_2O$ 0.2M $CoCl_2 \cdot 6H_2O$ and 0.1M H_3BO_3.

50x 5052 (1L): 0.5 % glycerol, 0.05% glucose, 0.2% α-lactose

20x NPS (1L): 0.5 M of $(NH_4)_2SO_4$, 1M of KH_2PO_4, 1M of Na_2HPO_4

17 Amino acid mixture (1L): 10 mg/mL of 17 amino acids except Cys, Tyr, Met

PA-0.5G (100mL): 92.2 mL of H_2O, 100µL of 1M $MgSO_4$, 10µL of 1000x Metal mix, 1.26 mL of 40% glucose, 5mL of 20x NPS, 400µL of 25mg/mL Methionine, 1mL of 17Amino acids mixture (10mg/mL each), 100µL of 100mg/mL Ampicillin

PASM-5052 (1L): 900mL of H_2O, 1.25mL of 1M $MgSO_4$, 1.25mL of 1000x Metal mix, 20mL of 50x 5052, 50mL of 20x NPS, 1mL of 100µM vitamin B12, 20mL of 17Amino acids mixture (10 mg/mL each), 400µL of 25 mg/mL Methionine, 5mLof 25 mg/mL Seleno-methionine, 1mL of 100 mg/mL Ampicillin.

For Ni-affinity purification:

10x Ni Phosphate buffer (1L): 0.02 M NaH_2PO_4, 0.18 M Na_2HPO_4, 2 M NaCl

Buffer A (1L): 100mL of 10x Ni Phosphate buffer, 1mL of β-Mercaptoethanol, 1mL of 0.2 M PMSF (phenylmethylsulfonyl fluoride), 5.0 % glycerol at pH 7.6.

Buffer B (1L): 100mL of 10x Ni Phosphate buffer, 500mM Imidazole, 1mL of 0.2M PMSF, 1mL of β-Mercaptoethanol, 5.0 % glycerol at pH 7.6. All the solutions were stored at 0-4°C.

2.2 Methods

Singleton proteins from *Pyrococcus furiosus (PF), Thermus Thermophilus (TT), Clostridium Thermocellum (Cthe), Aeropyrum pernix* K1 *(AP) were* cloned and purified. *Pyrococcus furiosus* cloned targets were obtained from Dr. Adams' lab, University of Georgia. *Thermus thermophilus, Clostridium thermocellum* and *Aeropyrum pernix* K1 targets were cloned at the Southeast Collaboratory for Structural Genomics (SECSG), UGA.

Expression and preparation of cell free extracts of PF Targets

The Open Reading frame (ORF) encoding PF protein was transformed into BL21-DE3 cells (Stratagene, LaJolla, CA). Step culture of 5 mL in LB media with cells carrying PF protein (e.g., PF0772) was grown aerobically with antibiotic (Ampicillin100µg/mL) for 16h at 37°C and transferred to 50 mL culture and then to 1L. The 1L culture was grown aerobically in an incubator/shaker at 37°C until OD_{600} ~ 0.6. Then the temperature was reduced to 20°C and isopropyl α-D-thiogalactopyranoside (IPTG, 1mM) was added to induce over-expression of the target PF cells. Incubation was continued at this temperature until OD_{600} ~2.0 (~5h). Cells were harvested by centrifugation (7500 xg, 15 min, 4°C). Approximately 2.0g of cells collected from 1L of *E.coli* culture were resuspended in 25 mL of phosphate buffer, 250µL of 0.2 M PMSF to inhibit the protease activity, 25µL of β-mercaptoethanol to prevent oxidative damage and 25µL

of 0.1mg/mL of DNase1. The resulting suspension was sonicated on ice using Branson Sonifier cell disrupter with a 0.5 inch probe tip for 2 min at 30sec intervals. Cell debris was removed by centrifugation (13,000 xg, 30 min, 4°C) and the cell extract (supernatant) was used immediately for further purification. All the purification process was carried out 0-4°C.

Ni-affinity purification of PF targets

A 5mL Hi-Trap affinity column (Qiagen, Valencia, CA) was charged with Nickel Chloride by following the protocol described below. The matrix was washed with 2 volumes of 1M NaOH followed by10 volumes of water, 2 volumes of EDTA followed by 10 volumes of water. The column was loaded with 2 volumes of 0.1 M of $NiCl_2$ and then allowed to stand for 10min. The column with bound Nickel was finally washed with 10 volumes of water. All the purifications were carried out using Akta Prime protein purification system (Amersham Biosciences). The Ni affinity column was then equilibrated in phosphate buffer A. The cell extract from *E.coli* was loaded onto the Nickel column and the bound 6x His-tagged PF protein was then eluted with buffer B. The fractions containing PF protein were collected, analyzed by SDS-PAGE and the relevant fractions were pooled.

His-Tag cleavage

After the Ni-affinity purification, the pooled fractions were treated with TEV protease (1-2 mL/L of culture) to remove 6x His-tag and dialyzed against buffer A at room temperature for 16h. The cleavage was analyzed using SDS-PAGE. Once the bound 6x His-tag was cleaved from the PF protein, it was subjected to heat treatment at 65°C for 30 min. to eliminate the unwanted *E.coli* protein. The 6x His-tag cleaved protein was again loaded onto Ni affinity column and the protein without His-tag was eluted in buffer A in flow through and the His-tag

alone was eluted with buffer B on a gradient run. The fractions were analyzed with SDS-PAGE and pooled.

Ion Exchange, Gel Filtration purification

Depending upon the purity of the protein from the above steps, further purification was carried out using either Ion exchange Q/P or Gel filtration column (S-75, Pharmacia) with 20mM HEPES/Tris, 100mM NaCl, 5% Glycerol, 5 mM EDTA adjust pH=7.6. The purity of the protein was checked with SDS-PAGE and the protein identity, molecular weight was further verified by Mass spectrometry (Bruker Autoflex) using MALDI TOF at the Department of Chemistry, UGA.

SDS-PAGE

The 26-well and 10-well 4-20% Tris-HCl (Bio-Rad, Criterion) were used to run SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis). 10µL of the protein sample was mixed with 10µL of 4x Laemmli Buffer (160mM Tris-HCl pH=6.8, 50% glycerol, 8% SDS, 0.1% bromophenol blue) and heated at 100°C for 10 min. Samples were loaded onto the gel with running buffer 1x Tris/Glycine/SDS (Bio-Rad) at 150 volts for 80 min. The gel was stained with Coomassie dye (Fisher Biotech) for 10-30 min. and destained with buffer containing 5% ethanol, 7% acetic acid and then the gel image was recorded.

Purification of Seleno-methionine protein

The proteins purified using LB media yielded crystals were considered for purification in PASM 5052 supplemented with Seleno-methionine [36]. The labeling of methionine residues in native protein with Se-Met is a standard and an easy way to obtain phase information in structure determination. The cells (BL21-DE3) encoding the interested proteins grown 5mL of LB media with ampicillin as an antibiotic (100µg/mL). This was then step cultured into 50mL of PA- 0.5 G media and incubated for 8h at 37°C. The culture was finally transferred to 1L PASM-

5052 auto inducing media containing 25mg/mL of Se-Met and incubated at 37°C with shaking for 3h. Then the temperature was lowered to 15-20°C for 20-24h until OD_{600} ~ 1.5. The cells were harvested and purified by Ni-affinity, TEV cleavage of 6x His-tag, Ion exchange and Gel filtration as described earlier.

2.3 Crystallization of PF targets

In order to identify an optimal crystal growth condition, initial crystallization screening was carried out in sitting drop vapor-diffusion process using Greiner Crystalquick plate [37]. The crystallization was set up using Honey bee robot (Cartesian Technologies) that uses 384 different conditions from 8 matrix screens: Crystal Screen, Crystal Screen II, MemFac, PegIon, Crystal Screen Cryo from Hampton Research, Wizard I, and Wizard II from Emerald Biostructures, Inc. This technique requires 200nL (nano liter) of the protein. After the setup of crystallization, the plates were incubated in Crystal farm (Bruker Nonius, Discovery Partners Intl.) at 18°C and the images were captured by a built-in camera. The droplet containing the purified protein, buffer and precipitant were allowed to equilibrate with a large reservoir containing same buffer and precipitant at higher concentration. Initially, the drop contains an insufficient concentration of precipitant for crystallization. Upon equilibration over a period of time, the water from the drop vaporizes and transferred to the reservoir which eventually increases the precipitant concentration and facilitates crystallization. Once the crystals observed with the sitting drop vapor-diffusion, the conditions were then optimized using modified microbatch under oil method in 72 well Nunc plates [38]. Modified microbatch was carried out using Douglas instruments ORYX-6 robot by mixing an equal volume (0.5μL each) of the protein and the precipitant which was then covered with oil (80% of paraffin and 20% of silicone oil). In addition to these techniques, hanging drop using 24 well-plate was also performed for

crystallization. Each hanging drop was prepared by mixing 0.5µL of the protein with 0.5µL of reservoir solution on a microscope plastic cover slip (Fisher Scientific) and vacuum greased to the well containing 300µL of the reservoir solution. The hanging drop employs the same principle as the sitting drop vapor diffusion. This technique differs from sitting drop in a manner that the drop appears as hanging manner. Both sitting and hanging drop were performed in airtight condition. Additive screen was also performed which mostly consists of small organic solvent molecules and inorganic salts which promotes crystallization [39]. Out of the 10 targets purified, only PF0772 was studied in detail. Crystals of PF0772 appeared in sitting drop conditions were optimized both in modified microbatch and in hanging drop methods. The crystals with size larger than 100 microns were mounted and screened for diffraction using Cu-rotating anode X-ray generator. The crystals of PF0772 were also grown in a slow cooling method. The crystals appeared this experiment were big in size. Other PF targets were purified as described earlier and screened for crystallization.

2.4 Reductive Methylation of Lysine.

When the crystallized protein diffracts poorly, the reductive methylation of lysine was carried out to modify the protein surface which probably would improve the diffraction quality of the crystals [40]. 20µL of freshly prepared 1M DMAB was added to the protein maintaining a minimum concentration of 10 mg/mL which was followed by the addition of 40µL of freshly prepared 1M formaldehyde and incubated for 2h in the dark with shaking at 4°C. This process was repeated two more times with final addition of 10µl of DMAB with continuous shaking for 18h in the dark. Then buffer exchange was carried out with 20mM HEPES/Tris, 100mM NaCl and the protein concentration was measured using UV-Vis spectrometer (Biomate, Thermo Spectronic). Theoretically, the protein concentration after reductive methylation would be half

(5mg/mL) of the original concentration (10mg/mL) and the similar trend was observed. The methylated protein was then screened for crystallization.

2.5 Mass spectrometry Analysis of the protein

The crystal of the interested protein or the protein itself was mixed with loading dye and boiled for 10min. The sample was then loaded onto the gel and SDS-PAGE was run with running buffer (1x Tris/Glycine/SDS) at 150 volts for 80 min. The gel was then stained with Commassie dye for 10-30min. and destained. The gel band of the interested protein at the margin of the detectable stain was excised and placed in a 500µL microfuge tube. The cut gel piece was rehydrated with 100µL of 50mM NH_4HCO_3 at 37 °C for 10min with shaking at 800 rpm and the resulting solution was discarded. The gel was then dehydrated with 200µL of 50% acetonitrile in 50mM NH_4HCO_3 for 15 min at 37°C with shaking. The solution was discarded and this process was repeated thrice. Then the gel slice was dried with vacuum evaporator (Precision, Jouan) with no heat for 15 min and the dried pieces were then resuspended in 15µL of 10ng/µL freshly prepared trypsin and incubated at 37°C for 16h with shaking. The peptides were then extracted by washing with 15µL of 50mM NH_4HCO_3 and twice with 15µL of 75% acetonitrile, 0.5% TFA. The supernatants were collected and pooled, then concentrated to 4-5µL using vacuum evaporator without heat. 1µL of 5% TFA was added to the concentrated sample. Using C-18 tip (Glygen Corp.) pre-washed with 0.05%TFA, the sample was bound to C-18 cartridge by pipetting up and down and finally washed with 0.05% TFA. The bound sample was eluted with 1µL of α-cyano-4-hydroxycinnamic acid MALDI matrix which was dissolved in 50% acetonitrile, 0.05%TFA. The sample was spotted on the MALDI plate by pipetting up and down 10 times which releases the peptides slowly from C-18 cartridge. The sample was then analyzed by mass spectrometry.

2.6 Expression and Purification of AP targets

Cells containing *Aeropyrum pernix* K1 target proteins were purified in LB media and those yielded good diffracting crystals were then purified in Se-Met media. The purification involved Ni-affinity column, His-tag cleavage, Ion exchange column and Gel filtration (S- 75). The protein purity was analyzed with SDS-PAGE and the molecular weight was verified using mass spectrometry (MALDI TOF).

2.7 Crystallization of AP targets

Initial screening for crystallization of these proteins was done as described earlier for *Pyrococcus furiosus*. Microbatch and hanging drop methods were also employed here. Out of several targets purified only three of the AP targets AP0305, AP0346, AP0371 produced crystals with diffracting quality. AP0305 is 11 kDa protein with 93 amino acids and has one methionine. Optimization of AP0305 in modified microbatch and hanging drop out were produced crystals in 68 different conditions. AP0436 is 17 kDa protein with 4 methionines. The optimization of AP0436 was done by placing the tray at 15°C for 2 days and placed at 4°C until the crystals appeared. AP0371 is 11.8 kDa protein with 109 amino acid residues. Crystals obtained from these proteins were mounted and screened for diffraction at home source Cu-rotating anode X-ray generator.

2.8 Preparation of Xenon derivatives

The crystals from AP0305 were used to prepare Xenon derivatives using Cryo-Xe-Siter (Rigaku Co.,) shown in Figure 2.1 [41]. The crystal was mounted on the loop and placed on the magnetic pin holder. The cryocup was half-filled with the cryoprotectant and fitted into a copper holder which can be pulled in or pushed away from the chamber by cryocup positioner. Initially the cryocup was pushed inside the chamber and the crystal was placed over the cryocup. The

handle was the clamped to close the upper chamber. The crystal was then exposed to Xenon gas for specified period of time (5, 10, 20, 25, 30 min). The dewar filled with liquid nitrogen holding the vial filled with tetra fluoromethane was placed inside the lower chamber. Once the crystal was exposed to the Xenon gas, handle holding the magnetic pin was lifted and cryocup was pushed away. The crystal loop was pushed down from the magnetic pin holder to sit on the vial at the lower chamber. The dewar was then removed and crystal was screened for diffraction.

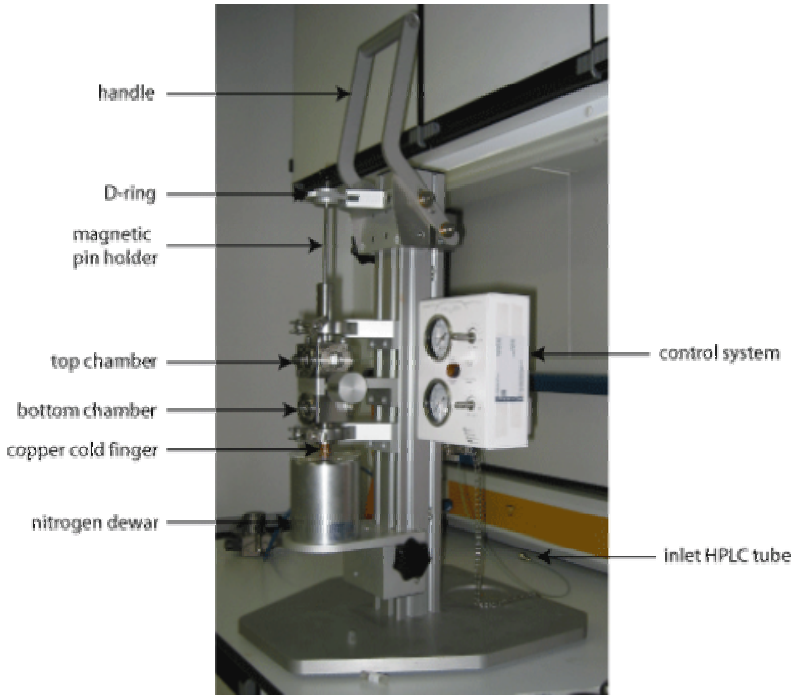


Figure 2.1 Cryo-Xe-Siter.

2.9 Expression and Purification of Cthe targets

The expression and purification of the protein was carried out in LB media and the proteins with diffracting quality crystals were purified in Se-Met. The protein purification strategy involved the use of Ni affinity, ion exchange and gel filtration. Cthe_3042 was studied in detail. This protein is 20.5 kDa with 181 residues, 2 cysteines, and 5 methionines. Since this

protein has totally 7 sulfur, the sulfur phasing also was tried with native protein. In addition to these targets, *Thermus thermophilus* (2 proteins) and *Archaeoglobus fulgidus* (6 proteins) target proteins were purified and crystallized as other target proteins.

2.10 Heavy atom soaking of the crystal

Heavy atom soaking of the crystals were carried out using Class A metal which consists of Alkali metals, alkaline earth metals, lanthanides and Class B metals like Pt, Hg, Au, Ag, Ir, Os. Heavy atom derivatives were prepared by soaking 1-2 mM of freshly prepared heavy atom solution to the crystals in mother liquor and the soaking was done over a period of time which ranges from 2 min. to 15 days. The soaked crystals were mounted at different time intervals and flash frozen in liquid nitrogen. The crystals were screened for diffraction and the best diffracted crystals were then used for data collection.

2.11 Data collection

Crystals were harvested using mounting pins [42] and quickly immersed in mother liquor either with 25% glycerol or Lithum sulfate as their cryoprotectant. The crystal was then flash frozen with liquid nitrogen. The crystals screened first at home source, the best diffracting crystals were sent to Southeast Regional Collaborative Access Team (SER-CAT). The collected data was indexed, integrated and scaled using HKL2000 [43] and D*trek.

CHAPTER 3

RESULTS

3.1 Target selection and BLAST analysis

Several numbers of ORFan targets from different organisms are selected for our studies towards structure determination and they are listed in Table 3.1. The selected targets are checked with BLAST for any homologous for these proteins (http://www.ncbi.nlm.nih.gov/BLAST/). pBLAST is a computational analysis or algorithm to compare the protein sequence at hand with a non redundant sequence database. The results of pBLAST are evaluated based on E value. When the protein sequence returned by pBLAST has an E value less than or equal to 0.001 then the protein found by algorithm is considered to be homologous to the template. If no protein with an E value below threshold is found, then the protein is considered to be unique. This criterion is approximately equivalent to 30% sequence homology. Using this method we chose 5 proteins that can be considered unique and their function was not known. Table 3.2 shows the results from BLAST analysis of crystallized protein with good diffraction.

Table 3.1 Total number of targets selected

A	B	C	D	E	F
PF1620	PF1468	AP0491	PF0247	AP0305	Cthe3042
PF1384	PF1225	AP0466	AP0471	AP0436	
AP0409	PF0997	AP0340	AP0955	AP0371	
AP0129	PF0867	AP0819	AP0087	PF0772	
AP0335	PF1786	AP0898	AP1103		
AP0356	AF1001	AP0683	AP0317		
AP0383	AF0854	AP0247	AF1053		
AF0774	Cthe1877	AP0270	AF1058		
AF0703	Cthe2022	AP0294			
Cthe1876	Cthe2844	AP0220			
Cthe3088	Cthe1771	AP0223			
	AP0238				

PF: *Pyrococcus furiosus;* **AP:** *Aeropyrum pernixK1;* **AF***: Archaeoglobus fulgidus;*
Cthe*: Clostridium thermocellum*

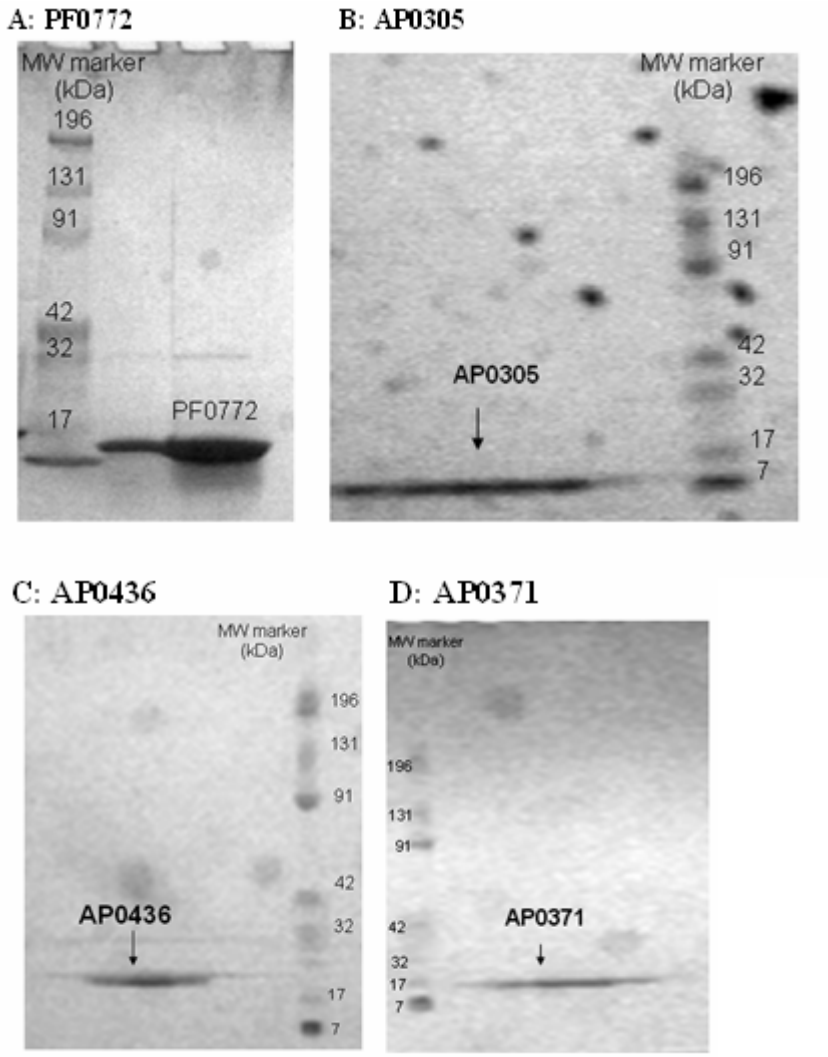
A: Proteins failed at various stages of purification
B&C: Proteins purified and screened for crystallization
D: Proteins produced crystal
E: Protein crystals with diffraction 3.5 Å and higher resolution
F: Protein crystal data phased and structure to be solved

Table 3.2 BLAST search for the target proteins against PDB structures

Protein	E value
PF0772	0.52
AP0305	0.90
AP0436	0.37
AP0371	1.90
Cthe_3042	0.97

3.2 Purification results

Using Ni-affinity and the Gel filtration columns, the proteins (PF0772, AP0305, AP0436, AP0371, and Cthe_3042) were purified to homogeneity and then concentrated for crystallization. The SDS-PAGE of the final fractions from gel filtration column S-75 is shown in Figure 3.1A-E.



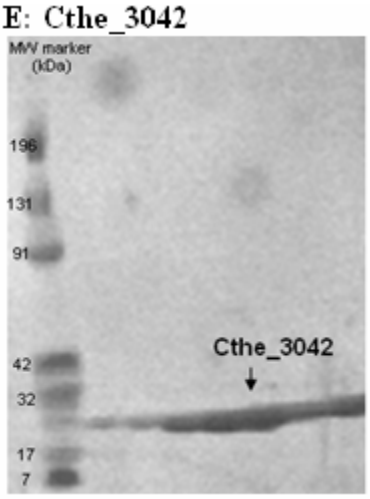


Figure 3.1 SDS-PAGE for the gel filtration column fractions of the proteins A-E

3.3 Mass spectrometry analysis results

Mass spectrometry analysis of the crystals was carried out as a quality control measurement to confirm that the purified protein corresponds to the target. As an example the MALDI analysis of AP0305 is shown here. The Schematic diagram explains the methodology of sample preparation for MALDI analysis (Figure 3.2). AP0305 protein was trypsin digested and analyzed in MALDI. The result from the Matrix Science is shown below. From the results it could be inferred that the crystallized protein corresponds to AP0305 (Figure 3.3).

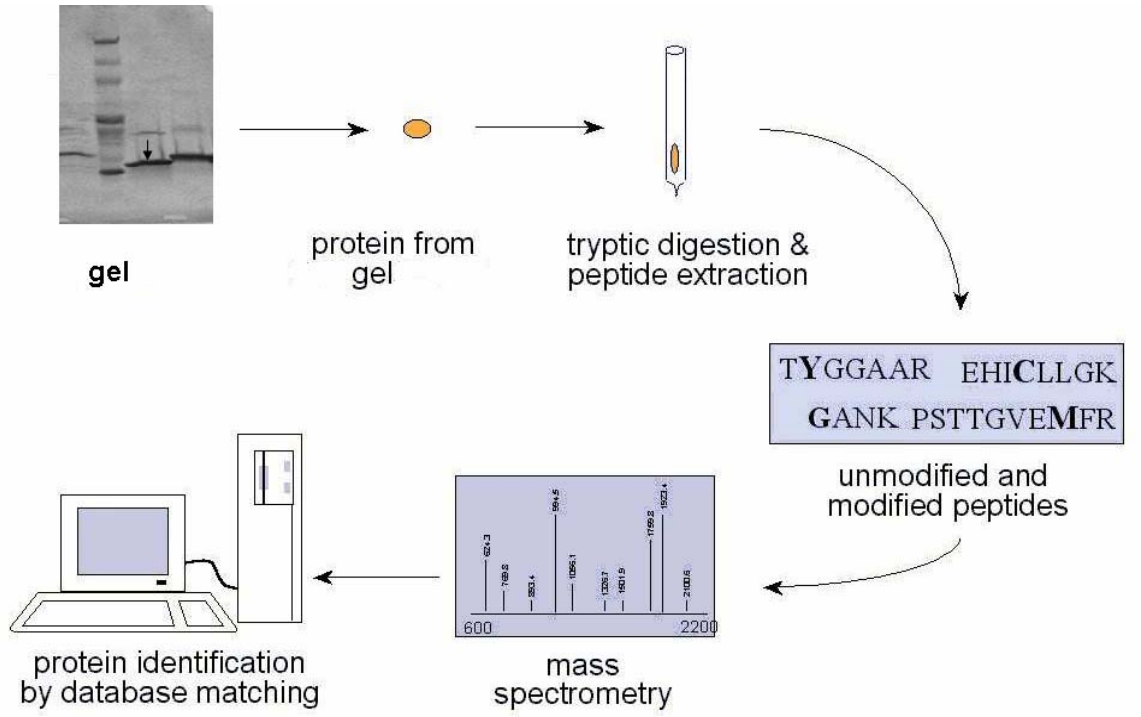


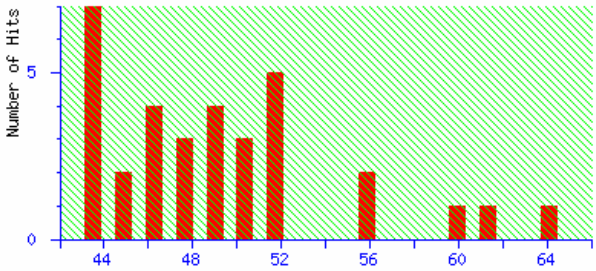
Figure 3.2 Schematic diagram of sample preparation for MALDI

Mascot Search Results

```
User            : poorani
Email           : poorani@chem.uga.edu
Search title    :
Database        : MSDB 20060831 (3239079 sequences; 1079594700 residues)
Timestamp       : 12 Jun 2007 at 22:19:12 GMT
Top Score       : 64 for E72689, hypothetical protein APE0933 - Aeropyrum pernix (strain K1)
```

Probability Based Mowse Score

Protein score is -10*Log(P), where P is the probability that the observed match is a random event.
Protein scores greater than 78 are significant (p<0.05).



Concise Protein Summary Report

Format As | Concise Protein Summary | Help

Significance threshold p< 0.05 | Max. number of hits AUTO

Re-Search All | Search Unmatched

1. E72689 **Mass:** 10858 **Score:** 64 **Expect:** 1.3 **Queries matched:** 5
hypothetical protein APE0933 - Aeropyrum pernix (strain K1)

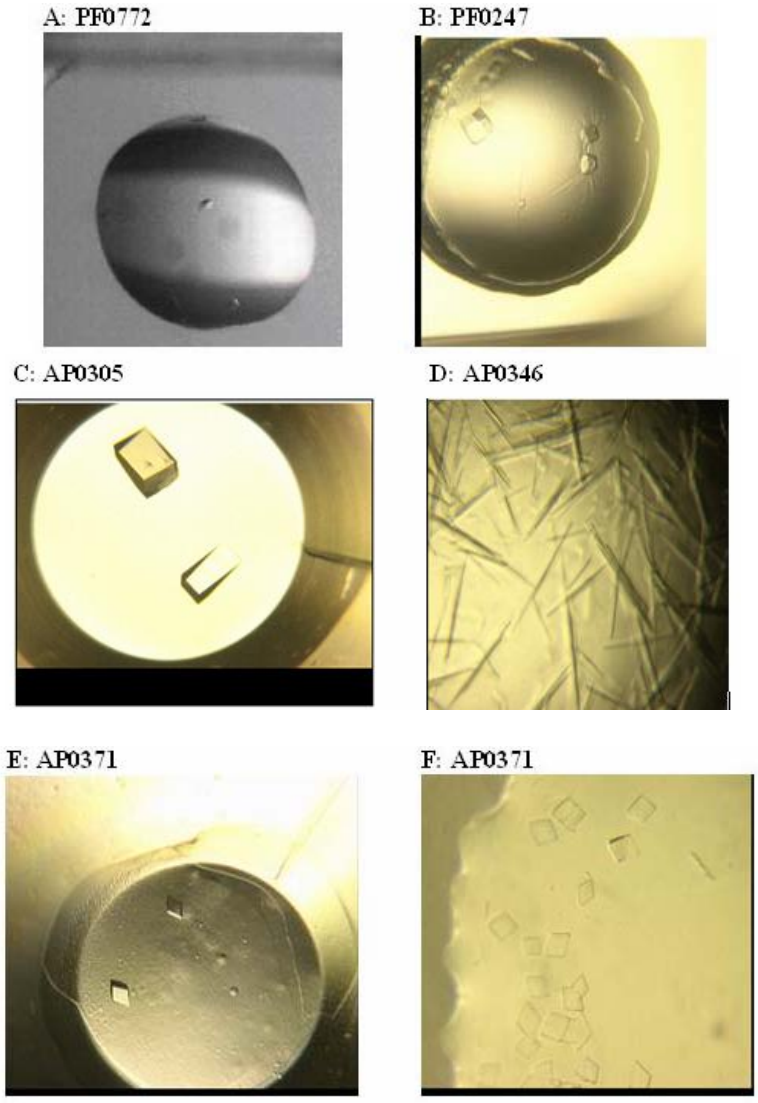
Search Parameters

```
Type of search          : Peptide Mass Fingerprint
Enzyme                  : Trypsin
Mass values             : Monoisotopic
Protein Mass            : Unrestricted
Peptide Mass Tolerance  : ± 1.2 Da
Peptide Charge State    : 1+
Max Missed Cleavages    : 1
Number of queries       : 12
```

Figure 3.3 MALDI results

3.4 Crystallization of proteins at 18ºC

The proteins purified from both LB and Se-Met (except AP0305) set up for crystallization using sitting drop vapor diffusion method and allowed to crystallize at 18°C. The conditions from the initial hit were then optimized using modified micro batch and hanging drop vapor diffusion. The crystals formed at 18°C are shown in Figure 3.4 A-G.



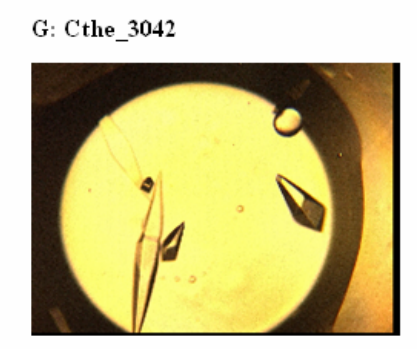


Figure 3.4 Crystals formed at initial stage A-G.

Figure 3.4A: Crystal of PF0772 from WI-40 (Se-Met)
Figure 3.4B: Crystal of PF0247 from Hampton Crystal Screen II-11 (Se-Met)
Figure 3.4C: Crystal of AP0305 from Hampton Crystal Screen I-17 (LB)
Figure 3.4D: Crystal of AP0436 from Hampton Crystal Screen I-18 (LB &Se-Met)
Figure 3.4E: Crystal of AP0371 from MP I-23 (LB &Se-Met)
Figure 3.4F: Crystal of Ap0371 Wizard I-5
Figure 3.4G: Crystal Cthe_3042 from Hampton Crystal Screen with Additive screen (LB&Se-Met)

The crystallization conditions under which crystals are formed and the cryo protection employed for mounting the crystals are shown in Table 3.3. The crystals were screened at APS and their diffraction limit is shown in Table 3.4

Table 3.3 Crystallization conditions

Protein	Crystallization condition	Cryo protection	Crystallization period & appearance of the crystal
PF0772	a) Wizard-40: 100mM MES/NaOH, pH 6, 10% v/v isopropanol, 200mM $CaAc_2$ b) Hampton Crystal Screen I-24: 100mM Na Acetate/HCl pH 4.6, 20% v/v isopropanol, 200mM $CaCl_2$ c) Cryo-30: 1500mM $(NH_4)_2SO_4$, 25%v/v glycerol	30% Li_2SO_4	3-4 days, Single crystal
PF0247	Hampton Crystal Screen I-11: 100mM Na Acetate/HCl pH 4.6, 100mM $CoCl_2$, 1000mM 6-hexanediol	Direct	7days Single crystal
AP0305	Hampton Crystal Screen I-17: 100mM Tris-HCl pH 8.5, 30% v/v PEG 4000, 200mM Li_2SO_4	Direct	1 day Single crystal
AP0436	Hampton Crystal Screen I-18: 100mM Na cacodylate/HCl, pH 6.5, 20% v/v, PEG 8000, 200mM Mg acetate	25% Ethylene glycol	3 days Cluster crystal
AP0371	a) MPI-23: 100mM CAPS/NaOH, pH 9.6, 30% v/v PEG 400, 100mM NaCl, 100mM Li_2SO_4 b) WizardI-5: 100mM CAPS/NaOH, pH 10.5, 30% v/v PEG 400	30% glycerol	5 days Single crystal
Cthe_3042	Hampton Crystal Screen I-17: 100mM Tris-HCl pH 8.5, 30% v/v PEG 4000, 200mM Li_2SO_4	10% Paratone- N oil, 90% mineral oil	1 day Single crystal

Table 3.4 Diffraction limit of proteins

Protein	Diffraction Limit (Å)
PF0772	**8**
PF0247	**10**
AP0305	**1.08-1.77**
AP0436	**2.9**
AP0371	**1.73**
Cthe_3042	**2.0-2.25**

3.5 Improvement of crystal size and diffraction limit

In order to improve the diffraction limit of the crystals of PF0772 and to improve the crystal size of AP0436 produced from sitting drop vapor diffusion method, modified microbatch under oil and hanging drop vapor diffusion methods, slow cooling crystallization and methylation of lysine residues were carried out and explained below.

Slow cooling crystallization of PF0772 and AP0436

The crystals of PF0772 diffracted to 8Å and various optimization techniques were employed to improve the crystal size diffraction limit. Temperature dependence of PF0772 i.e., protein precipitation at 4°C and at -20°C, disappearance of the precipitation at 65ºC, suggested to try an alternate method for crystallization: carry out the slow cooling crystallization to produce better diffracting crystals. The crystallization was setup with Crystal screen I-24 condition using modified microbatch method. The plate was initially placed at 39°C, and the crystallization temperature was reduced to 24°C, 18°C, 15°C, and finally to 4°C there by controlled crystallization was carried out (Figure 3.5A). The crystal previously diffracted to 8Å, was improved to 3.5Å by this method.

Crystals of AP0436 from the initial hit in Hampton Crystal Screen I-18 appeared as clusters. These crystals were not useful for diffraction screening. In order to obtain single crystals, the condition was optimized in modified microbatch under oil and placed at 4ºC for 3-4 days, then it was moved to 15°C for a day or two until the crystals appear. The plate was then moved back to 4ºC. The crystals grown by this method were mostly single crystals (Figure 3.5B). The single crystals were stuck to the plate and attempts to mount the crystal resulted in breaking of the crystal into pieces. Using the suggestion from Douglas Instruments Ltd., the plastic (Nunc plate) besides the crystal was pressed with a needle and the plastic was deformed which released the crystal.

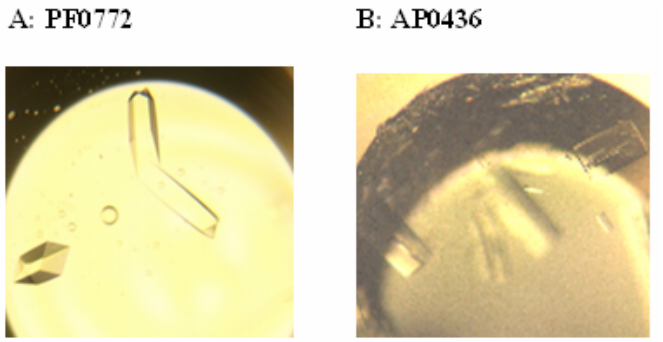


Figure 3.5 Crystals from slow cooling A-B.

Figure 3.5A: Crystal of PF0772 from WI-40 obtained from slow cooling
Figure 3.5B: Crystal of AP0436 from Hampton Crystal Screen obtained from slow cooling

Crystallization of Methylated PF0772

Methylation of lysine residues was carried out for further improvement in the diffraction to yield a resolution higher than 3.5Å which was obtained through slow cooling. The crystallization of methylated PF0772 was setup in Hampton Crystal screen I-24 using both the modified microbatch and the hanging drop method. The microbatch method did not produce any

crystals. However, the hanging drop method produced crystals with different morphology (Figure 3.6). The crystal from methylated protein did not diffract.

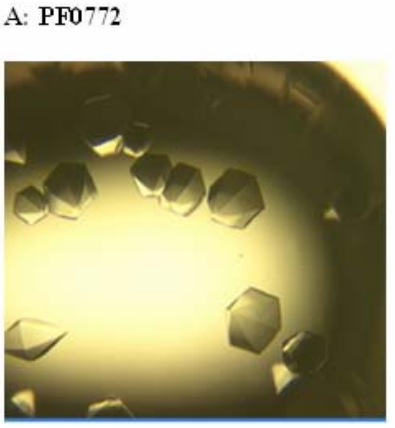


Figure 3.6 Crystals of Methylated PF0772**.**

Figure 3.6: Crystals of Methylated PF0772 from Hampton Crystal Screen I-24

3.6 Crystals soaked with heavy atoms

The structure determination requires heavy atom signals for phase information. Usually Selenium atom signals from Se-Met derivatives crystals are employed. In cases where this method could not be employed, the heavy atom soaking of the native crystals were carried out. In our studies, the heavy atom soaking of the native crystals were carried out due the following reasons: a) the absence of anomalous signals of Selenium from the data collected for Se-Met crystals of AP0436 and AP0371(data not shown), b) unavailability of Se-Met crystals of AP0305 and c) low quantity of Se-Met protein from Cthe-3042. Crystals of AP0305, AP0436, AP0371 and Cthe-3042 were soaked in several heavy atoms as listed in Table 3.5. The heavy atom soaked crystals are shown in Figure 3.7A-B.

Table 3.5 Heavy atoms used for soaking

Protein	Heavy atoms used (2-5mM)	Soaking time (h)
AP0305	K_2PtCl_4, $K_2Pt(NO_2)_4$, $HgCl_2$, $KAu(CN)_2$, $(H_2NCH_2CH_2NH_2)PtCl_2$, $KAuCl_4$, K_2IrCl_6, HgC_6H_5COOCl, KI, $PrCl_3$, $La(NO_3)_3$, $HgC_6H_5SO_3Cl$, $NdCl_3{\cdot}6H_2O$, $LaCl_3$,$UO_2(CH_3COO)_2{\cdot}2H_2O$, XENON GAS	2-24
AP0436	K_2PtCl_4, $UO_2(CH_3COO)_2{\cdot}2H_2O$, K_2PtCl_6	2-6
AP0371	K_2PtCl_4, KI, $UO_2(CH_3COO)_2{\cdot}2H_2O$, $La(NO_3)$, CsCl, $CaCl_2$	2-15
Cthe_3042	$KAuCl_4$	168- 200 (7-14 days)

K_2PtCl_4: Potassium tetrachloro Platinate (II)
$K_2Pt(NO_2)_4$: Potassium tetranitro Platinate (II)
$HgCl_2$: Mercuric Chloride
$KAu(CN)_2$: Potassium Auro cyante
$(H_2NCH_2CH_2NH_2)PtCl_2$: Dichloro ethylenediamine Platinum (II)
$KAuCl_4$: Potassium tetrachloro Aurate (III)
K_2IrCl_6: Potassium hexachloro Iridate (IV)
HgC_6H_5COOCl: parachloromercury-benzoic acid (Hg(II))
$HgC_6H_5SO_3Cl$: parachloromercury-phenyl sulphonic acid
KI: Potassium Iodide
$PrCl_3$: Praseodymium (III) Chloride
$NdCl_3{\cdot}6H_2O$: Neodymium (III) Chloride
$La(NO_3)_3$: Lanthanum Nitrate
$LaCl_3$: Lanthanum Chloride
$UO_2(CH_3COO)_2{\cdot}2H_2O$: Uranyl Acetate
K_2PtCl_6: Potassium hexachloro platinate (IV)
CsCl: Cesium Chloride
$CaCl_2$: Calcium Chloride

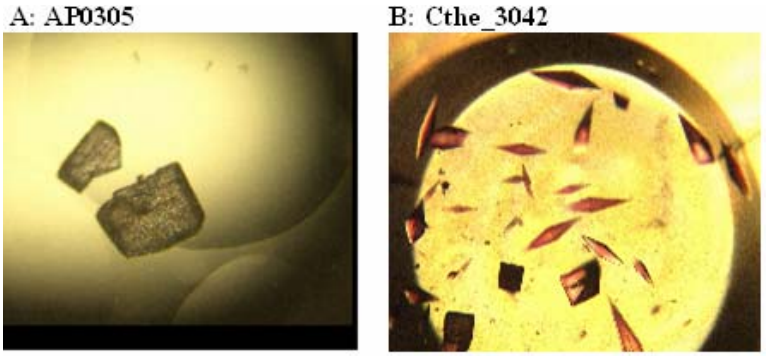


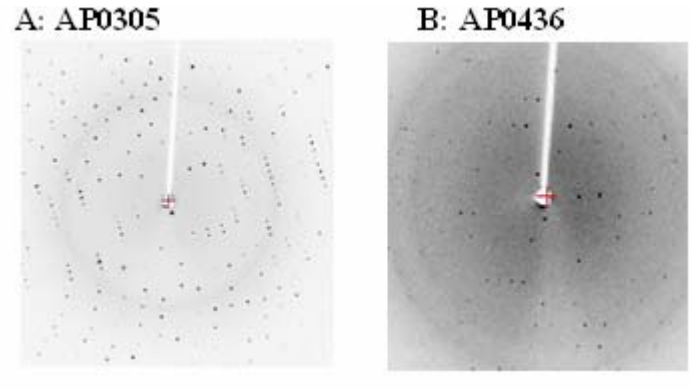
Figure 3.7 Heavy atom soaked crystals A-B.

Figure3.7A: Crystals of AP0305 soaked in KAu $(CN)_2$
Figure 3.7B: Crystals of Cthe_3042 in $KAuCl_4$

3.7 Diffraction and Data processing

Diffraction pattern of the protein crystals

The heavy atom crystals were harvested with cryoprotection and screened for diffraction. The diffraction limit of PF0772 was very low and could not be continued further. The data were collected for all the heavy atom soaked crystals of AP0305 but only KI soaked data are shown, whereas for other targets: Pt soaked crystals of AP0436, Cs soaked AP0371, Au soaked Cthe_3042 were only useful for data collection. The crystals soaked with other heavy atoms were either dissolved or cracked. The diffraction patterns of all the other crystals are shown in Figure 3.8A-D.



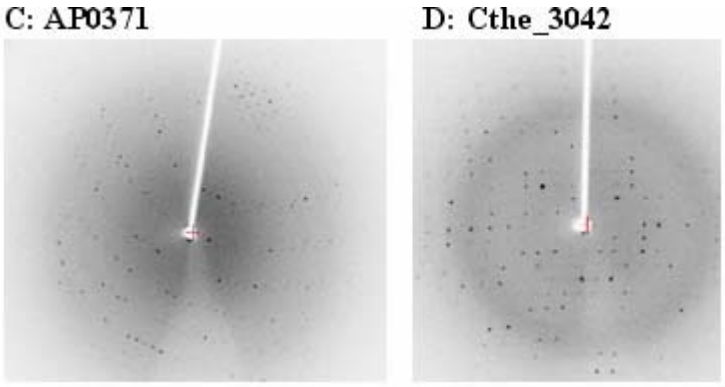


Figure 3.8 Diffraction pattern of heavy atom soaked crystals A-D.

Figure 3.8A: KI soaked AP0305
Figure 3.8B: K_2PtCl_4 soaked AP0436
Figure 3.8C: CsCl soaked AP0371
Figure 3.8D: $KAuCl_4$ soaked Cthe_3042

Data collection and Processing

The crystals screened for diffraction at home source, were mounted in a Hampton Research loop and flash frozen in liquid nitrogen and sent APS. The data were collected at 22-ID of the SER-CAT beam line, APS (Argonne National Laboratory) using Mar CCD300 and at BM using Mar CCD225. The data then were processed and scaled using HKL2000 and D* TREK. The parameters inferred from processing the data is given in Table 3.6.

Table 3.6 Crystal Data parameters

ORF	Diffraction (Å)	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	Space Group	Matthew's Coefficient $Å^3 Da^{-1}$	Solvent Content %
AP0305	**1.37**	**30.0**	**46.96**	**59.44**	**90**	**90**	**90**	**$P2_12_12$**	**1.98**	**37.9**
AP0436	**2.9**	**30.43**	**113.18**	**76.67**	**90**	**90**	**90**	**C222**	**1.97**	**37.44**
AP0371	**1.73**	**25.57**	**55.14**	**82.75**	**90**	**97.79**	**90**	**$P2_1$**	**2.47**	**50.11**
Cthe_3042	**2.15**	**40.76**	**40.76**	**194.93**	**90**	**90**	**90**	**P42**	**1.95**	**37**

From the statistics of the processed data shown in Table 3.7-3.10, the Ras (shown in red) found to be very low which indicates that there was no anomalous signal corresponds to the heavy atom. Hence the structure determination was not possible. The data statistics for all the protein are listed below:

Table 3.7 Data statistics for KI soaked AP0305

Res.Shell	Rsym-Shell		Rfree	nRfree	<AnoI/SigI>a		<AnoI/SigI>c		Ras	
36.72										
to 5.30	0.0649	0.0649	0.0559	22	1.39	1.39	1.55	1.55	0.89	0.89
to 4.14	0.0643	0.0638	0.0586	36	1.33	1.28	1.42	1.23	0.93	1.04
to 3.58	0.0664	0.0709	0.0575	52	1.30	1.25	1.43	1.47	0.91	0.85
to 3.23	0.0694	0.0836	0.0603	76	1.32	1.36	1.51	1.82	0.88	0.75
to 2.99	0.0731	0.1075	0.0632	94	1.33	1.38	1.49	1.36	0.90	1.02
to 2.80	0.0762	0.1248	0.0660	113	1.33	1.30	1.49	1.55	0.89	0.84
to 2.66	0.0793	0.1595	0.0673	122	1.33	1.34	1.49	1.49	0.89	0.90
to 2.54	0.0818	0.1789	0.0693	143	1.34	1.43	1.47	1.23	0.89	1.16
to 2.42	0.0835	0.1997	0.0711	164	1.36	1.51	1.46	1.32	0.91	1.14

Table 3.8 Data statistics for Pt soaked AP0436

Res.Shell	Rsym-Shell		Rfree	nRfree	<AnoI/SigI>a		<AnoI/SigI>c		Ras	
56.64										
to 4.87	0.0536	0.0536	0.0594	29	1.56	1.56	1.68	1.68	0.93	0.93
to 3.82	0.0552	0.0562	0.0620	54	1.50	1.44	1.80	1.97	0.83	0.73
to 3.32	0.0612	0.0758	0.0662	84	1.61	1.80	1.83	1.88	0.88	0.96
to 3.01	0.0669	0.1038	0.0732	122	1.67	1.82	1.86	1.90	0.90	0.96
to 2.78	0.0727	0.1387	0.0819	160	1.77	2.15	1.90	2.09	0.93	1.03
to 2.62	0.0781	0.1722	0.0890	197	1.84	2.16	1.96	2.38	0.94	0.91
to 2.48	0.0813	0.1731	0.0913	226	1.86	1.95	2.01	2.47	0.92	0.79
to 2.36	0.0831	0.1717	0.0938	259	1.86	1.93	1.98	1.70	0.94	1.13
to 2.25	0.0839	0.1838	0.0948	287	1.86	1.84	1.98	2.15	0.94	0.86
to 2.05	0.0844	0.2407	0.0952	307	1.87	2.04	1.98	2.28	0.95	0.89

Table 3.9 Data statistics for Au soaked Cthe_3042 (7days)

Res.Shell	Rsym-Shell	Rfree	nRfree	<AnoI/SigI>a	<AnoI/SigI>c	Ras
41.05						
to 4.92	0.0631 0.0631	0.0641	81	3.52 3.52	2.97 2.97	1.18 1.18
to 3.89	0.0643 0.0658	0.0634	166	3.15 2.78	2.65 2.36	1.19 1.18
to 3.39	0.0722 0.1187	0.0698	236	3.27 3.51	2.84 3.23	1.15 1.09
to 3.07	0.0772 0.1795	0.0741	310	3.31 3.44	2.90 3.07	1.14 1.12
to 2.85	0.0822 0.3444	0.0781	382	3.36 3.53	3.01 3.04	1.12 1.04
to 2.68	0.0874 0.5033	0.0830	458	3.38 3.51	3.10 3.50	1.09 1.01
to 2.54	0.0925 0.7706	0.0872	531	3.40 3.50	3.17 3.59	1.07 0.98
to 2.42	0.0974 0.8957	0.0907	595	3.41 3.46	3.22 3.60	1.06 0.96
to 2.31	0.1012 0.9426	0.0934	659	3.44 3.67	3.25 3.76	1.06 0.98
to 2.13	0.1034 0.9103	0.0947	712	3.47 3.87	3.27 3.78	1.06 1.02

Table 3.10 Data statistics for Au soaked Cthe_3042 (15 days)

Res.Shell	Rsym-Shell		Rfree	nRfree	<AnoI/SigI>a		<AnoI/SigI>c		Ras	
48.20										
to 7.03	0.1066	0.1066	0.1065	16	6.92	6.92	2.86	2.86	2.42	2.42
to 5.44	0.1167	0.1647	0.1137	28	8.09	8.91	2.90	2.90	2.79	3.07
to 4.70	0.1237	0.1652	0.1191	40	7.03	5.43	2.67	2.28	2.63	2.38
to 4.23	0.1301	0.1900	0.1236	52	6.63	5.68	2.89	3.49	2.29	1.62
to 3.90	0.1375	0.2577	0.1365	66	6.06	4.29	3.00	3.31	2.02	1.30
to 3.67	0.1440	0.3180	01440	80	5.47	2.91	2.91	2.55	1.88	1.14
to 3.46	0.1507	0.4355	0.1492	93	5.45	2.91	2.96	3.36	1.75	1.14
to 3.31	0.1550	0.4897	0.1555	111	5.01	2.30	2.86	2.01	1.71	1.39
to 3.14	0.1580	0.4875	0.1589	126	4.78	3.07	2.79	2.20	1.70	1.30
to 2.90	0.1605	06426	0.1602	143	4.76	4.48	2.80	3.45	1.06	0.98

As seen in the table 3.9 and 3.10 the Ras value for Cthe_3042 seemed to increase from 1.18 (7 days soaking in Au) to 2.79(15 days soaking). Attempts were made to produce this protein in Se-Met media. The quantity of the protein obtained from 1L culture of Se-Met was not sufficient for crystallization. Hence 15L culture was grown to produce enough protein. The crystals obtained from Se-Met media were diffracted to 2.15Å and the initial phasing (Figure 3.9) was done by SGXPRO& pipeline and the structure was solved. Refinement and Validation are to be done as future work.

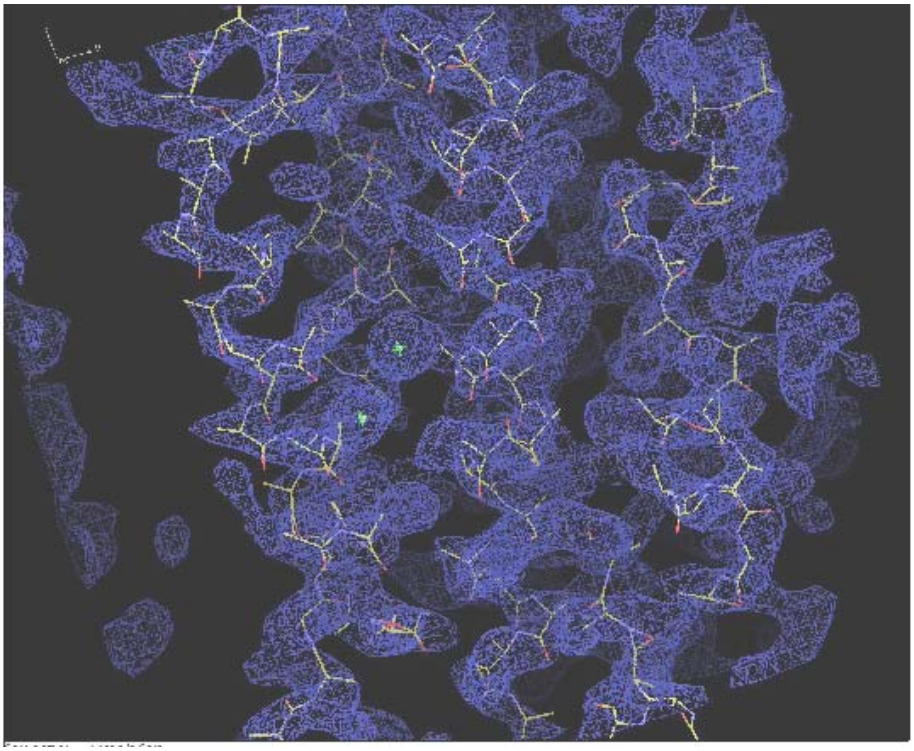


Figure 3.9 Initial phasing of Cthe_3042.

CHAPTER 4

DISCUSSION

In this thesis we have performed crystallization of many ORFan proteins from *Pyrococcus furiosus*, *Aeropyrum pernix*K1, and *Clostridium thermocellum*. From the crystallization results of PF0772, it is inferred that this protein produces diffracting quality (3.5 Å) crystals, however to obtain the structure we need better diffracting crystals. In order to get the better diffracting crystal we have carried out slow cooling crystallization and methylation of the lysine residues to modify the surface of the protein. The above methods of crystallization did not give better results. Hence, either mutation of the protein or recloning can be performed. The PF0772 clone used in our studies has 6x His-tag which could not be cleaved because it has no cleavage site for 6x His-tag. Cleaving of His-tag might give better crystal and hence a possibility for structure determination.

AP0305 gene has been cloned from *Aeropyrum pernix* and expressed in *E.coli* host to produce sufficient quantities of recombinant proteins to initiate structural analysis. AP0305 is an ORFan protein which was confirmed from the BLAST analysis. This protein was a good target for structure determination by the fact that the resolution of X-ray diffraction of this crystal was very good (~1Å). Since there is no methionine other than the "initiator" methoinine available in AP0305, we could not prepare seleno-methinione protein. Heavy atom soaking of the crystal with different time intervals and Xenon gas trapping methods to obtain phase information were

carried out. Since we could not get any phase information from above methods, "*Ab initio*" method calculation can be carried out to derive phase information.

AP0436 has four methionines and considered as ORFan protein based on BLAST analysis. The protein was purified both in LB and Se-Met media. The native crystals were tried for sulfur phasing since it has four methionine residues. The crystals from both native and Se-Met media were sensitive to radiation damage, as is apparent from the comparison of images taken at the beginning and end of a data set collection. Also there was no heavy atom signal which can be inferred from low Ras value (anomalous signal).

The Crystals of AP0371 were obtained from native and Se-Met media. Determination of the structure of this protein was not possible due to the absence of heavy atom site in heavy atom soaked native crystals which is indicated by low Ras values.

Cthe_ 3042 has molecular weight of about 20.7kDa. The recombinant protein was purified from native and Se-Met media. The crystal from native media consists of seven sulfur atom which was very promising for sulfur phasing theoretically, however experimentally it was not possible since there was no signal for sulfur atom. Heavy atom soaking of Cthe_3042 with $KAuCl_4$ at different time intervals provided an evidence for the presence of heavy atom signal which can be inferred by an increase in Ras value.

The purification of Cthe_3042 in Se-Met required 15L of the culture to obtain sufficient amount of protein. Crystals obtained form the selenomethionie media were used for data collection. The crystal was diffracted to 2.15Å and the structure was solved. For future work the Refinement and Validation of Cthe_3042 structure has to be performed.

REFERENCES

1 Kim, S. H. (1998) Shining a light on structural genomics. Nature Structural Biology. **5**, 643-645

2 Wang, Z. X. (1998) A re-estimation for the total numbers of protein folds and superfamilies. Protein Eng. **11**, 621-626

3 Fischer, D. and Eisenberg, D. (1999) Finding families for genomic ORFans. Bioinformatics (Oxford, England). **15**, 759-762

4 Fischer, D. (1999) Rational structural genomics: affirmative action for ORFans and the growth in our structural knowledge. Protein engineering. **12**, 1029-1030

5 Siew, N. and Fischer, D. (2003) Twenty thousand ORFan microbial protein families for the biologist? Structure. **11**, 7-9

6 Burley, S. K. (2000) An overview of structural genomics. Nature Structural Biology. **7**, 932

7 Robb, F. T., Maeder, D. L., Brown, J. R., DiRuggiero, J., Stump, M. D., Yeh, R. K., Weiss, R. B. and Dunn, D. M. (2001) Genomic sequence of hyperthermophile, Pyrococcus furiosus: implications for physiology and enzymology. Methods Enzymol. **330**, 134-157

8 Poole, F. L., 2nd, Gerwe, B. A., Hopkins, R. C., Schut, G. J., Weinberg, M. V., Jenney, F. E., Jr. and Adams, M. W. (2005) Defining genes in the genome of the hyperthermophilic archaeon Pyrococcus furiosus: implications for all microbial genomes. Journal of bacteriology. **187**, 7325-7332

9 Yamazaki, S., Yamazaki, J., Nishijima, K., Otsuka, R., Mise, M., Ishikawa, H., Sasaki, K., Tago, S. and Isono, K. (2006) Proteome analysis of an aerobic hyperthermophilic crenarchaeon, Aeropyrum pernix K1. Mol Cell Proteomics. **5**, 811-823

10 Marsden, R. L., Lee, D., Maibaum, M., Yeats, C. and Orengo, C. A. (2006) Comprehensive genome analysis of 203 genomes provides structural genomics with new insights into protein family space. Nucl. Acids Res. **34**, 1066-1080

11 Brenner, S. E. and Levitt, M. (2000) Expectations from structural genomics. PRS. **9**, 197-200

12 Å ali, A. (1998) 100,000 protein structures for the biologist. Nature Structural Biology. **5**, 1029

13 Hiyama, T., Zhao, M., Kitago, Y., Yao, M., Sekine, S.-i., Terada, T., Kuroishi, C., Liu, Z.-J., Rose, J., Kuramitsu, S., Shirouzu, M., Watanabe, N., Yokoyama, S., Tanaka, I. and Wang, B.-C. (2006) Structural basis of CoA recognition by the Pyrococcus single-domain CoA-binding proteins. Journal of Structural and Functional Genomics. **7**, 119-129

14 Clancy Kelley, L.-L., Dillard, B. D., Tempel, W., Chen, L., Shaw, N., Lee, D., Newton, M. G., Sugar, F. J., Jenney, F. E., Lee, H. S., Shah, C., Poole, F. L., Adams, M. W. W., Richardson, J. S., Richardson, D. C., Liu, Z.-J., Wang, B.-C. and Rose, J. (2007) Structure of the hypothetical protein PF0899 from Pyrococcus furiosus at 1.85 Å resolution. Acta Crystallographica Section F. **63**, 549-552

15 Constantina Bakolitsa, R. S. D. M. L. S. B. J. M. C. X. D. A. M. D. M.-A. E. S. E. R. F. A. G. C. G. S. K. (2004) Crystal structure of an orphan protein (TM0875) from <I>Thermotoga maritima</I> at 2.00-Å resolution reveals a new fold. Proteins: Structure, Function, and Bioinformatics. **56**, 607-610

16 Brenner, S. E. (2001) A tour of structural genomics. Nature reviews. **2**, 801-809

17 Flaherty, K. M., McKay, D. B., Kabsch, W. and Holmes, K. C. (1991) Similarity of the Three-Dimensional Structures of Actin and the ATPase Fragment of a 70-kDa Heat Shock Cognate Protein. Proceedings of the National Academy of Sciences. **88**, 5041-5045

18 Jung, J. W. and Lee, W. (2004) Structure-based functional discovery of proteins: Structural proteomics. J Biochem Mol Biol. **37**, 28-34

19 Wang, Y., Bryant, S., Tatusov, R. and Tatusova, T. (2000) Links from Genome Proteins to Known 3-D Structures. Genome Res. **10**, 1643-1647

20 Nicholas O'Toole, M. G. Z. O. W. M. M. C. (2004) The structural genomics experimental pipeline: Insights from global target lists. Proteins: Structure, Function, and Bioinformatics. **56**, 201-210

21 Bray, J. E., Marsden, R. L., Rison, S. C. G., Savchenko, A., Edwards, A. M., Thornton, J. M. and Orengo, C. A. (2004) A practical and robust sequence search strategy for structural genomics target selection. Bioinformatics (Oxford, England). **20**, 2288-2295

22 Petsko, G. (2007) An idea whose time has gone. Genome Biology. **8**, 107

23 Siew, N. and Fischer, D. (2003) Analysis of singleton ORFans in fully sequenced microbial genomes. Proteins. **53**, 241-251

24 Siew, N. and Fischer, D. (2003) Unravelling the ORFan puzzle. Comparative and Functional Genomics. **4**, 432-441

25 Yin, Y. and Fischer, D. (2006) On the origin of microbial ORFans: quantifying the strength of the evidence for viral lateral transfer. BMC Evolutionary Biology. **6**, 63

26 Skovgaard, M., Jensen, L. J., Brunak, S., Ussery, D. and Krogh, A. (2001) On the total number of genes and their length distribution in complete microbial genomes. Trends in Genetics. **17**, 425-428

27 Woese, C. R., Kandler, O. and Wheelis, M. L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences of the United States of America. **87**, 4576-4579

28 Kristjánsson, J. K. and Hreggvidsson, G. O. (1995) Ecology and habitats of extremophiles. World Journal of Microbiology and Biotechnology. **11**, 17-25

29 Stetter, K. O. (1996) Hyperthermophilic procaryotes. FEMS Microbiol. Rev. **18**, 149-158

30 Madigan, M. T. and Narrs, B. L. (1997) Extremophiles. Scientific American. **276**, 82

31 Adams, M. W. W. (1999) The biochemical diversity of life near and above 100 °C in marine environments. Journal of applied microbiology(Print). **85**, 108-117

32 Volkl, P., Huber, R., Drobner, E., Rachel, R., Burggraf, S., Trincone, A. and Stetter, K. O. (1993) Pyrobaculum aerophilum sp. nov., a novel nitrate-reducing hyperthermophilic archaeum. Appl. Environ. Microbiol. **59**, 2918-2926

33 Sako, Y., Nomura, N., Uchida, A., Ishida, Y., Morii, H., Koga, Y., Hoaki, T. and Maruyama, T. (1996) Aeropyrum pernix gen. nov., sp. nov., a novel aerobic hyperthermophilic Archaeon growing at temperatures up to 100 °C. International journal of systematic bacteriology. **46**, 1070-1077

34 Fiala, G. and Stetter, K. O. (1986) Pyrococcus-Furiosus Sp-Nov Represents a Novel Genus of Marine Heterotrophic Archaebacteria Growing Optimally at 100-Degrees C. Archives of Microbiology. **145**, 56-61

35 Tyurin, M. V., Desai, S. G. and Lynd, L. R. (2004) Electrotransformation of Clostridium thermocellum. Appl. Environ. Microbiol. **70**, 883-890

36 Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. **41**, 207–234

37 Liu, Z. J., Tempel, W., Ng, J. D., Lin, D., Shah, A. K., Chen, L., Horanyi, P. S., Habel, J. E., Kataeva, I. A. and Xu, H. (2005) The high-throughput protein-to-structure pipeline at SECSG. Acta Cryst. **61**, 679–684

38 Chayen, N. E. (1997) The role of oil in macromolecular crystallization. Structure. **5**, 1269-1274

39 Jancarik, J. and Kim, S. H. (1991) Sparse matrix sampling: a screening method for the crystallization of macromolecules. J. Appl. Crystallogr. **24**, 409–411

40 Rypniewski, W. R., Holden, H. M. and Rayment, I. (1993) Structural consequences of reductive methylation of lysine residues in hen egg white lysozyme: An x-ray analysis at 1.8-.ANG. resolution. Biochemistry. **32**, 9851-9858

41 Sauer, O., Schmidt, A. and Kratky, C. (1997) Freeze-Trapping Isomorphous Xenon Derivatives of Protein Crystals. Journal of Applied Crystallography. **30**, 476-486

42 Teng, T. Y. (1990) Mounting of crystals for macromolecular crystallography in a free-standing thin film. Journal of Applied Crystallography. **23**, 387-391

43 Gewirth, D. (1994) The HKL Manual: an Oscillation Data Processing Suite for Macromolecular Crystallography. Yale University, New Haven, CT, USA