Structural Determination of the Hypothetical Protein, TM1816 From *Thermatoga Maritima*

Nancy Fathali

ERULF Program
Pacific University
Stanford Linear Accelerator Center
Menlo Park, California

August 14, 2002

Prepared in partial fulfillment of the requirement of the Office of Science, DOE Energy
Research Undergraduate Laboratory Fellowship under the direction of Aina Cohen and
Paul Ellis in Structural Biology at Stanford Linear Accelerator Center.

Table of Contents

Abstract.....	1
Introduction.....	2
X-ray crystallography.....	2
Structure.....	3
X-ray diffraction.....	3
Thermatoga maritima.....	4
Materials and Methods.....	5
Results and Discussion.....	6
Acknowledgements.....	7
References.....	8

Structural Determination of the Hypothetical Protein, TM1816 From *Thermatoga Maritima*. Nancy Fathali (Pacific University, Forest Grove, OR, 97116) Aina Cohen and Paul Ellis (Stanford Linear Accelerator Center, Menlo Park, CA, 94025).

X-ray crystallography has altered the course of molecular biology. It is largely because of this method that we are able to visualize the structure of molecules in atomic detail. It's important to note that if one knew, in accurate detail, the structure of the active site, or other important regions of an enzyme, then one can fabricate molecules that would bind to those sites and interfere with the enzyme's activity. The present study is on a bacterium known as *Thermatoga maritima*. The function of this protein is unknown, to date. However, it is known that *T. maritima* metabolizes ~~many~~ simple and complex carbohydrates. The protein was crystallized, x-ray diffraction was performed, a pattern was obtained, the data was analyzed in order for an electron density map to be calculated and then a model was determined. Since it is a hypothetical protein, there isn't much that is known yet. The fact that *T. maritima* is related to other organisms mean that if every protein in this organism is to be solved, a library of sequences with corresponding structures can be obtained.

Introduction

X-ray Crystallography

X-ray crystallography has altered the course of molecular biology. It is largely because of this method that we are able to visualize the structure of molecules in atomic detail. The technology of macromolecular crystal growth is essential in x-ray diffraction analysis, thus making crystals of macromolecules the keystones of modern biological science. The objective is obtaining that single, perfect crystal, composed of regularly repeated arrangement of atoms. Upon obtainment of that perfect crystal, x-ray diffraction is used to quantify the relative quality of the crystal.

It is often that focus is given exclusively to the successes of x-ray crystallography without realizing the importance of macromolecular crystallization and its significance in other areas of study. Protein crystals are utilized for downstream processing and purification in the industrial production of enzymes, they are involved in pharmaceutical formulation, and they are often useful in various other areas of biophysical experimentation (McPherson. 1989). More significantly is their role in the broader field of crystal growth research. As McPherson said,

Modern techniques, such as atomic force microscopy, have shown that macromolecular crystals employ the same mechanisms for growth as do conventional crystals. Because of the much larger size of the molecules, however, the kinetics of growth are substantially attenuated. As a consequence, direct observation of nucleation and growth processes are possible that would otherwise elude detection with conventional crystals. Thus, macromolecular crystallization now serves as the best model system for general crystallization from solution.

Structure

It would be impossible to understand a macromolecule completely without knowledge of its structure. It provides information about the arrangement of the atoms and the changes in the atomic arrangements as the molecules carry out their biological or chemical function. It's important to note that if one, in accurate atomic detail, the structure of the active site, or other important regions of the enzyme, then one can fabricate molecules that would bind to those sites and interfere with the enzyme's activity. These drugs, or conventional molecules, can be very beneficial in treating diseases.

X-ray diffraction

Further enhancing the impact of crystallography has been the advent of synchrotron x-ray sources of ultra-high intensity coupled with sensitive and fast x-ray detectors (McPherson, 1989). These instruments have reduced the size of crystals required for analyses from millimeters to less than 50 μm in diameter, and have made data collection more accurate and rapid. Other improvements such as crystal cooling mean that no longer are dozens of crystals needed to obtain a set of diffraction data, instead one crystal will suffice for an entire structure determination.

A diffraction pattern is obtained by directing a beam of x-rays onto a crystal, producing a set of reflections. Each reflection corresponds to an intensity with an interplanar spacing (d). The d spacing is related to the x-ray scattering angle θ by Bragg's law, which is $n\lambda = 2d\sin\theta$, where λ the wavelength of the radiation. The smaller the value of d , the more detailed is the information that the diffracted wave conveys. However, the smaller the d , the greater the corresponding the Bragg scattering angle θ , the farther from

the primary x-ray beam the reflection falls in the diffraction pattern, and the weaker the data becomes.

In a diffraction pattern the spacing of spots depends on the size of the unit cell, while the intensity of the spots depends on the structure of the protein. For example, in a big protein one would observe a big unit cell and close diffraction spots. In a small protein, one would get a small unit cell and diffraction spots that are far apart. It's important to note that for crystals of two different proteins, which have identical unit cells and the same space group, the spot spacing would be identical. However, the intensities of the spots would be different.

Thermatoga maritima

The present study is on a protein known as *Thermatoga maritima*. This protein was obtained from the Joint Center for Structural Genomics (JCSG). This institute is a collaboration to determine the three dimensional structure of proteins that would be produced by the genetic code of various organisms.

T. maritima is a non-spore forming, rod-shaped bacterium belonging to the order Thermotogales. It is the one-thousandth eight-hundred sixteenth protein found in this organism. These bacteria thrive in very hot, high pressure environments such as thermal vents at the bottom of the sea.

It is also known that *T. maritima* metabolizes many simple and complex carbohydrates including glucose, sucrose, starch, cellulose, and xylan. Both cellulose and xylan, through conversion to fuels (such as H_2), have great potential as renewable carbon and energy sources.

Materials and Methods

2.1 Materials

Thermatoga maritima was crystallized by University California San Diego and the Scripps Institute

2.2 Preparation for sequencing

A crystal of *Thermatoga maritima* was received and data was collected.

2.21 Blu-Ice: Used to obtain two images at 90° apart.

2.22 Mosflm: Used to autoindex and give a cell—found to be orthorhombic with a $p2_12_12_1$ symmetry. Ran strategy to find spacegroup and determine where to collect data.

2.23 Blu-Ice: Three data sets at 90° was collected. Collected absorption spectra at selenium.

2.24 Mosflm: Autoindexed, refined cell, and integrated. This was performed three times.

2.25 Scale: Obtained an average intensity (B-factor, scale factor). Noticed that there was a systematic absence therefore it had a three screw axis with a $p2_12_12_1$ symmetry. Gave a list of magnitudes but no phase.

2.26 Solve: Worked out the Se positions (1 strong, 1 weak), calculated the phases, calculated the map, and autotraced the map. This gave a partial map by looking at the sequence and deciding whether it was a reasonable match.

2.3 Sequencing of *Thermatoga maritima*

2.31 Xfit: Rebuilt the side chains and backbone based on: stereochemistry, polarity, and other properties of amino acids.

- 2.32 CNS: Automatically refines the geometry so that it becomes structurally adequate. Performs reciprocal space refinement which includes use of Fourier transforms. Recalculates map.
- 2.33 Xfit: Builds more into structure. Ran CNS and X-fit until an adequate R-factor was obtained.

Results and Discussion

A ribbon diagram of *Thermatoga maritima* was obtained. The structure displayed an alpha-beta-alpha motif. This is very typical of proteins. The beta sheets also displayed parallel and anti parallel characteristics.

Observations of the genome sequence show similarities between the thermophilic Archaea and Eubacteria. The theory that *Thermatoga* is one of the closest living descendants of the first life on Earth is also supported by the fact that every other species studied so far, including human beings, has some of the exact same DNA found inside *Thermatoga*.

Since it is a hypothetical protein, there still isn't much that we know about *Thermatoga maritima*. Hypothetical meaning that it is in the genome, but it hasn't been observed in the organism; therefore, it isn't sure if the organism actually creates it.

The fact that *T. maritima* is related to everything is why it was chose, because the structure is preserved in other organisms. Therefore, if every protein in this organism is to be solved, then a library of sequences with corresponding structures can be obtained. Thus, one can take any arbitrary sequence and find out what it does. This method can prove to be extremely valuable in the future.

Acknowledgements

I thank the United States Department of Energy – Office of Science for giving me the opportunity to participate at the Stanford Linear Accelerator Center for structural biology research.

My gratitude also goes to my mentors, Aina Cohen and Paul Ellis. The Structural Biology group at SSRL- Ashley Deacon, Mitch Miller, Tim McPhillips, and Linda Brinen.

A special thanks to Helen Quinn and Al Ashley for giving me the opportunity to return to the Stanford Linear Accelerator Center.

The research described in this paper was performed at the Stanford Linear Accelerator Center, a national scientific user facility sponsored by the United States Department of Energy's Office of Biological and Environmental Research and located in Menlo Park, California.

References

- Adams, M., Ellis, G., Gover, S., Naylor, C., & Phillips, C., "Crystallographic Study of Coenzyme, Coenzyme Analogue and Substrate Binding in 6-Phosphogluconate Dehydrogenase: Implications for NADP Specificity and the Enzyme Mechanism." Structure, vol. 2, July 15, 1994. 651-668.
- Chowdhury, E., Nagata, S., & Misona, H., "3-Hydroxyisobutyrate Dehydrogenase from *Pseudomonas putida* E23: Purification and Characterization." Bioscience Biotechnology Biochemistry, vol. 60, June 11, 1996. 2043-2047.
- Hawes, J., Crabb, D., Chan, R., Rougraff, P., & Harris, R., "Chemical Modification and Site-Directed Mutagenesis Studies of Rat 3-Hydroxyisobutyrate Dehydrogenase." Biochemistry, vol. 34, December 28, 1994. 4231-4237.
- Hawes, J., Harper, E., Crabb, D., & Harris, R., "Structural and Mechanistic Similarities of 6-Phosphogluconate and 3-Hydroxyisobutyrate Dehydrogenases Reveal a New Enzyme Family, the 3-Hydroxyacid Dehydrogenases." Biochemistry, vol. 389, April 29, 1996. 263-267.
- Karsten, W., Chooback, L., & Cook, P., "Glutamate 190 Is a General Acid Catalyst in the 6-Phosphogluconate-Dehydrogenase-Catalyzed Reaction." Biochemistry, vol. 37, May 29, 1998. 15691-15697.
- Worrall, E., Gassain, S., Cox, D., Sugden, M., & Palmer, N., "3-Hydroxyisobutyrate Dehydrogenase, an Impurity in Commercial 3-Hydroxyisobutyrate Dehydrogenase." Biochemistry, vol. 241, October 17, 1986. 297-300.
- Zhang, L., Chooback, L., & Cook, P., "Lysine 183 Is the General Base in the 6-Phosphogluconate-Dehydrogenase-Catalyzed Reaction." Biochemistry, vol. 38, February 23, 1999. 11231-11238.