

**STRUCTURAL STUDIES OF TERPENOID BIOSYNTHESIS AND
BACTERIAL CELL DIVISION**

A Dissertation

by

DONG YANG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2006

Major Subject: Biochemistry

**STRUCTURAL STUDIES OF TERPENOID BIOSYNTHESIS AND
BACTERIAL CELL DIVISION**

A Dissertation

by

DONG YANG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	James C. Sacchettini
Committee Members,	James C. Hu
	Patricia J. Liwang
	Terry L. Thomas
Head of Department,	Gregory D. Reinhart

August 2006

Major Subject: Biochemistry

ABSTRACT

Structural Studies of Terpenoid Biosynthesis and Bacterial Cell Division.

(August 2006)

Dong Yang, B.S., Zhe Jiang University

Chair of Advisory Committee: Dr. James C. Sacchettini

The objective of this work is to investigate the structures of two nucleotide binding proteins: mevalonate kinase (MVK) and FtsZ.

MVK is the key enzyme involved in terpenoid biosynthesis. In this study, we solved the crystal structures of the *M. jannaschii* MVK apoprotein, as well as the protein in complex with ligands. Its fold was analyzed and firmly established within the GHMP kinase family, in which homoserine kinase (HSK), phosphomevalonate kinase and galactokinase also belong. Structural analysis in combination with enzyme kinetics studies revealed the mechanism of this enzyme upon substrate binding, catalysis and inhibition. In particular, the phosphate-binding loop was found to be critically involved in the binding of nucleotides and terpenoids, via the interaction with a di-phosphate moiety from the ligand. An enzymatic reaction mechanism was constructed based on our structural data and it is consistent with kinetics studies from the literature. In this mechanism, the invariant residue Asp 155 functions as a general base that increases the nucleophilicity of the phosphoryl acceptor. Finally, a virtual screening study has been

performed to explore the ligand binding potential of MVK. Compounds predicted to bind MVK were tested and analyzed.

FtsZ is a prokaryotic homologue of tubulin that forms the apparatus for bacterial cell division. The structure of a crystal filament of the *M. tuberculosis* FtsZ complexed with GDP was described in this study. It shows an anti-parallel, left-handed double helical architecture. Compared with the straight crystal filament revealed earlier by other groups, the catalytic T7 loop in our structure is found to be outside the nucleotide binding site, indicating the GTPase is inactive. Furthermore, the buried surface area in our crystal filament is less, probably suggesting the helical FtsZ filament is less stable. We therefore proposed that the hydrolysis of GTP and the releasing of the γ -phosphate group will trigger the rearrangement of the FtsZ fiber, characterized by the exclusion of the T7 loop, which might lead to a less stable helical filament and would be the first step for the disassembly of FtsZ polymer.

DEDICATION

To my parents and sister, for their love and support. To my friends for their help and encouragement.

ACKNOWLEDGEMENTS

First of all, I am thankful to Dr. James Sacchettini, my advisor, for giving me opportunities to do research in his lab and also giving me valuable advice and guidance during these years. I am also thankful to other members of my advisory committee: Dr. Patricia Liwang, Dr. James Hu and Dr. Terry Thomas, for their suggestions and discussion on my work.

Second, I have to thank colleagues in Dr. Sacchettini's lab, particularly Dr. Lance Shipman, as he led me into the world of crystallography when I just finished my rotation and joined this lab. I thank other members or former members in our group, particularly Dr. Remo Perozzo, Dr. Christoph Eicken, Dr. Sheng Ye, Dr. Clare Smith, Dr. Chih-chin Huang, Dr. Yoonsang Cho, Dr. Sudharsan Sridharan, Ms. Stephanie Swanson, Ms. Arati Ramesh and Mr. Feng Wang, for their advice and help at various stages during my work.

My research also involves several collaborators. Particularly, I have to thank Dr. David Borhani (Abbott Bioresearch Center) and Ms. E. Lucile White (Southern Research Institute), for their initial work on FtsZ and their generosity in supplying me full-length FtsZ proteins. I have to thank Mr. Christos Savva for his kind help in performing EM studies on FtsZ filaments. I am also thankful to Drs. A. Ian Scott and Charles Roessner for providing us initial clones of MVK.

Finally, my work could not be completed without help from the supporting staff in this department, particularly the administrative assistants for Dr. Sacchettini: Ms. Karen Hodges and the late Dr. Linda Fisher.

NOMENCLATURE

ADP	Adenosine 5'-diphosphate
ADP β S	Adenosine 5'-(β thiol)-diphosphate
AMP	Adenosine 5'-monophosphate
AMPPCP	Adenosine 5'-(β , γ -methylene)-diphosphate
APS	Advanced Photon Source
ATP	Adenosine 5'-triphosphate
CAMD	Center for Advanced Microstructures and Devices
CDP	Cytidine 5'-diphosphate
CDP-ME	4-diphosphocytidyl-2-C-methyl-erythritol
CDP-ME2P	4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate
CMP	Cytidine 5'-monophosphate
CoA	Coenzyme A
CTP	Cytidine 5'-triphosphate
DEAE	Diethylaminoethyl
DMAPP	Dimethylallylpyrophosphate
DMSO	Dimethyl Sulfoxide
DTT	Dithiothreitol
dUTP	2'-deoxyuridine 5'-triphosphate
DXP	1-Deoxy-D-xylulose 5-phosphate
EM	Electron Microscopy
FPP	Farnesyl Pyrophosphate

GAP	GTPase Activating Protein
GDP	Guanosine 5'-diphosphate
GDP β S	Guanosine 5'-(β thiol)-diphosphate
GGPP	Geranylgeranyl Diphosphate
GMPPCP	Guanosine 5'-(β , γ -methylene)-diphosphonate
GMPPNP	Guanosine 5'-(β , γ -imido)-triphosphate
GPP	Geranyl Pyrophosphate
GTP	Guanosine 5'-triphosphate
GTP γ S	Guanosine 5'-(3-O-thio)triphosphate
HEPES	N-(2-hydroxyethyl)-piperazine-N'-2-ethanesulfonic Acid
HIDS	Hyperimmunoglobulinemia D Syndrome
HMB-PP	(E)-4-hydroxy-3-methyl-but-2-enylpyrophosphate
HMG	3-Hydroxy-3-methyl glutaryl
HSK	Homoserine Kinase
IPP	Isopentenyl Pyrophosphate
IPTG	Isopropyl-beta-D-thiogalactopyranoside
ITP	Inosine 5'-triphosphate
LB	Luria Bertani Broth
LDH	Lactate Dehydrogenase
LDL	Low Density Lipoprotein
MAD	Multiwavelength Anomalous Dispersion
MECDP	2-C-methyl-D-erythritol 2,4-cyclodiphosphate

MEP	Methylerythritol Phosphate
MIR	Multiple Isomorphous Replacement
MPD	2-methyl-2,4-pentanediol
MVK	Mevalonate Kinase
NAD ⁺	Nicotinamide Adenine Dinucleotide (oxidized form)
NADH	Nicotinamide Adenine Dinucleotide (reduced form)
NADPH	Nicotinamide Adenine Dinucleotide Phosphate (reduced form)
NCS	Non-crystallography symmetry
NDP	Any Diphosphonucleotide
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PIP	Di-iodobis-ethylenediamine-di-platinum II nitrate
PK	Pyruvate Kinase
PMSF	Phenylmethylsulfonyl Fluoride
RGS	Regulators of G Protein Signaling
SAD	Singlewavelength Anomalous Dispersion
SIR	Single Isomorphous Replacement
SDS	Sodium dodecyl sulphate
Se-Met	Selenium Methionine
SRE	Sterol Regulatory Element
Tris	Tris(hydroxymethyl) Aminomethene
UPPS	Undecaprenyl Pyrophosphate Synthetase

UTP	Uridine 5'-triphosphate
X	Any Amino Acid Residue

TABLE OF CONTENTS

	Page
ABSTRACT	iii
DEDICATION	v
ACKNOWLEDGEMENTS.....	vi
NOMENCLATURE.....	vii
TABLE OF CONTENTS.....	xi
LIST OF FIGURES.....	xiii
LIST OF TABLES.....	xvi
 CHAPTER	
I INTRODUCTION TO STRUCTURAL BIOLOGY	1
Scope of Structural Biology.....	1
Introduction to Crystallography	5
Theory of Structural Based Drug Design	20
Overview of Chapters	24
II BACKGROUND OF TERPENOID BIOSYNTHESIS AND MVK.....	27
Terpenoid Biosynthesis	27
MVK.....	41
Significance of This Work.....	47
III STRUCTURAL STUDY OF MVK.....	48
Methods	48
Results and Discussion.....	56
Conclusion.....	113

CHAPTER	Page
IV BACKGROUND OF CELL DIVISION AND FTSZ.....	116
Bacterial Cell Division	116
FtsZ	117
Significance of This Work.....	126
V STRUCTURAL STUDY OF A FTSZ CRYSTAL PROTOFILAMENT.	128
Methods	128
Results and Discussion	133
Conclusion	157
VI GENERAL CONCLUSION AND FUTURE DIRECTIONS	158
REFERENCES.....	163
VITA.....	175

LIST OF FIGURES

FIGURE	Page
1. Basics of Diffraction.....	8
2. Principle of MAD.....	12
3. Principle of Molecular Replacement.....	14
4. The Mevalonate Dependent Pathway for IPP Biosynthesis.....	33
5. The Non-mevalonate Pathway for IPP Biosynthesis.....	36
6. Prenyltransferase.....	38
7. The Reaction Mechanism of MVK.....	42
8. Multiple Sequence Alignment of MVKs from Various Sources.....	58
9. Anomalous Patterson Map.....	62
10. The Electron Density Maps before and after Density Modification.....	64
11. Rat and <i>M. jannaschii</i> MVK.....	66
12. Ramachandran Plot of MVK Crystal Structure.....	68
13. Fold Comparison of N-domains of GHMP Kinases.....	70
14. Fold Comparison of C-domains of GHMP Kinases.....	71
15. Superose 6 Size-exclusion Chromatography of MVK.....	73
16. Oligomer State of MVK and HSK Crystal Structure	74
17. Interactions between Structural Motifs.....	76
18. Active Site Pocket of GHMP Kinases.....	78
19. The Electron Density of Nucleotide Ligands.....	81

FIGURE	Page
20. Overall Structures.....	85
21. The Binding Mode of the Phosphate Groups.....	86
22. The Mode of Mg ²⁺ Coordination and the Comparison of the Rat and <i>M. jannaschii</i> Enzymes.....	89
23. The Binding Mode of the Nucleoside Base.....	90
24. The Phosphoryl Acceptor Binding Site.....	91
25. The Proposed Mechanism of the <i>M. jannaschii</i> MVK.....	96
26. The Terpenoids Binding Site.....	98
27. Two Virtual Screening Hits.....	105
28. The Structural Alignment of Tubulin and FtsZ.....	120
29. Composition and Constriction of the Z-ring	123
30. Structures of FtsZ in Complex with Accessory Proteins.....	125
31. Use of Native PAGE to Screen Heavy Atoms.....	132
32. Polymerization Assay of the Full-length FtsZ.....	136
33. EM Images of FtsZ Filaments.....	138
34. Anomalous Difference Map.....	143
35. Electron Density Map for FtsZ.....	144
36. The GDP Binding Site.....	147
37. The Double Helical Structure of the FtsZ Crystal Filament.....	149
38. The Difference in the Dimeric Interface between the Curved and Straight Dimer.....	151

FIGURE	Page
39. The Putative Mechanism for FtsZ Dynamics.....	154
40. The Potential Role of the C-terminal Tail.....	155

LIST OF TABLES

TABLE	Page
1. Representative Terpenoids.....	28
2. Data Collection Statistics of the MAD Experiment.....	54
3. Statistics of Data Collection and Structural Refinement.....	55
4. Selenium Sites and Their Statistics.....	63
5. Refinement Statistics of MVK Apoprotein Crystal Structure.....	65
6. List of MVK Secondary Structural Elements.....	67
7. Kinetic Studies on MVK.....	83
8. Virtual Screening Statistics.....	100
9. FLEXX Hits Chosen for Initial Validation.....	101
10. Virtual Screening Hits Exploring the Nucleoside Base Binding Pocket.....	107
11. Structural Activity Studies Based on HTS08391.....	110
12. Additional Virtual Screening Results.....	112
13. List of FtsZ (a) and Tubulin (b) Crystal Structures.....	119
14. Data Collection Statistics of FtsZ Crystals.....	135
15. Refinement Statistics of FtsZ.....	146

CHAPTER I

INTRODUCTION TO STRUCTURAL BIOLOGY

Scope of Structural Biology

If a cell could be viewed as a factory, the architecture of the assembly line and the design of each machine would be essential for understanding how this factory works. Since its birth in mid 20th century, structural biology, in combination with biochemical and genetic methods, has provided critical pieces of information for decoding biological secrets. For example, without the fiber diffraction data, Watson and Crick would not have been able to deduce the double helical model of DNA that, even after several generations, still inspires our imagination.

The earliest subjects studied by structural biology were enzymes or other ligand binding proteins. By co-crystallizing the ligand with the protein, one can understand the mechanism of substrate recognition and catalysis. Comparison of ligand bound and apoprotein structures also reveals protein conformational changes upon ligand binding and supports the concept of “induced fit”. One success story was the work done by Perutz three decades ago (1). By studying the crystal structures of the oxy and deoxy forms of hemoglobin, he discovered that the binding of oxygen led to the rearrangement of the protein quaternary structure by the movement of the iron relative to the porphyrin ring, which resulted in the breakage of a salt bridge. This structural observation

This dissertation follows the style and format of *Journal of Biological Chemistry*.

explained the basis for cooperativity of oxygen binding in hemoglobin.

Transition state analogs can sometimes be co-crystallized with the enzyme forming a tight “activated complex”, which provides more direct information regarding the catalysis. More advances in the aspects of “kinetic protein crystallography” have been made recently. For example, by using the “trigger-freeze” method (2), one might trap a reaction intermediate in the crystal. This approach, however, is limited by the relatively slow rate of cooling (about 0.1s), which sometimes fails to capture the intermediate. An alternative approach is time resolved crystallography that takes advantage of Laue diffraction (2). Events in the time frame of less than 1 ns could be captured by this method. The limitation of this approach, however, is the quality of the data collected. Therefore, time resolved crystallography and trapping methods can be used in complementation to better understand the kinetic questions at hand. Finally, advances in synchrotron radiation have pushed the diffraction limit of some crystals to beyond 1 Å, which permits the visualization of hydrogen atoms as well as multiple conformations of side chains (3). As a result, a plausible catalytic mechanism can be distinguished from its alternatives.

Many macromolecules are in fact drug targets. Recent breakthroughs in the field of structural biology have aided in the increased output of the number of crystal structures solved. As a result, much interest in the field has occurred in the pharmaceutical sector, which has led to successful target and inhibitor development. For example, protein kinases and their various roles in carcinogenesis are important targets for developing anti-tumor drugs. Studies on BCR-Abl kinase in complex with anti-

cancer drug Gleevec not only sheds light on how this compound is effective, but also provides clues for developing inhibitors against other protein kinase targets (4). Another example is the HIV protease. Structural studies on the binding mode of its inhibitors play significant roles in improving their efficacy (5). Another important drug is INH, a first line anti-tuberculosis drug. Crystallographic research reveals that it forms an adduct with the cofactor NADH and occupies the active site of the drug target, enzyme InhA, which blocks the synthesis of fatty acids in mycobacterium (6).

Although we can learn a great deal by studying individual macromolecules, a complete understanding of the cell factory is not possible without an intact picture of the assembly line. Not surprisingly, structural biology has progressed into the domain of large macromolecular complexes (7). Structures of the ribosome, signal recognition particle, ATP synthase, RNA polymerase, DNA polymerase, nucleosome, proteasome, chaperones, endosomal sorting complexes and also multi-enzyme complexes such as rubisco and pyruvate dehydrogenase have been reported. Many of these were solved by X-ray crystallography. Other methods, particularly single particle cryo-EM and electron tomography, are making an important contribution to this field (8). Fitting atomic crystal structures into low resolution EM density has been a standard approach to study the very large complexes that are refractory to crystallographic method alone. Success stories include the structural revelation of the bacteriophage T4 infection mechanism, the interaction by the ribosome with the signal recognition particle and also the ribosome protein-conducting channel complex, etc.

Another frontier of structural biology lies with integral membrane proteins which include channels, transporters, receptors and protein complexes involved in electron transfer. As much cellular processes take place at or on the membrane, the structural solution of these proteins is indispensable for addressing important biological questions. They also represent more than 70% of current drugs targets. The structural determination of membrane proteins is a tough problem for crystallographers. The first membrane crystal structures were of the photosynthetic reaction centers. Membrane proteins involved in respiration, such as succinate:quinone oxidoreductase, cytochrome *bc*₁ complex and cytochrome *c* oxidase (designated as complexes II, III and IV), followed suit. Crystal structures of ion channels emerged at the turn of this century. Structures of transporters were reported only years later, first ABC transporters and then lactose permease. Now membrane protein crystallography has been progressed into the era of membrane protein structural genomics, which advocates the rapid expression, purification and crystallization of a large number of membrane proteins from a variety of species.

Technology has always set the pace for structural biology. The development of better instrumentation for X-ray diffraction, NMR and EM has and still remains the driving force for this field. The vast improvement in computing power is another important factor. The progress in molecular biology, protein expression systems and robotic technologies has also made important contributions. As this dissertation will be focused on X-ray crystallography, we will now start to discuss the basic theories for

protein crystallization and crystal diffraction, which will then be followed by practical issues regarding protein phasing and structural refinement.

Introduction to Crystallography

Theory of protein crystallization

A protein crystal is a type of protein aggregate formed by orderly deposition of protein molecules. During crystallization, the protein is kept in a supersaturated state in which a two-stage process takes place: in the first stage (nucleation stage), a few protein molecules start to associate to form the nuclei that have the same intermolecular contacts as that of the final crystal; in the second stage (crystal growth stage), more protein molecules deposit onto the nuclei and visible crystals appear (9). This process is dependent upon many factors such as precipitants, salts (particularly divalent cation), pH, polyamines, chaotropes, linkers and detergent (10). Therefore, for a particular protein to be crystallized, a variety of crystallization conditions, generally organized as a sparse matrix for crystal screening, are tested before a condition is found.

In addition, kinetic parameters, including the mixing ratio between protein and precipitant, the method for establishing equilibrium, as well as the temperature are also significant during crystallization (11). One technique that is worth special consideration is the seeding method (12). Generally, the energy barrier for nucleation is higher than that for crystal growth. Therefore, the stage for nucleation requires a higher supersaturation level. In order to limit the number of nuclei and produce larger crystals, it is sometimes advantageous to separate these two stages, which can be accomplished by seeding preformed crystals or microcrystals into a new drop with a lower saturation

level. The conditions in the new drop are not sufficient for nucleation *per se* but are supportive for growth from the seeded nuclei.

Theory of crystal diffraction

X-rays are electromagnetic radiation with a very short wavelength occurring at a range from 0.1 to 1000 Å and can be described mathematically as follows.

$$A \exp[i\alpha], \quad (1.1)$$

where A is the amplitude of the wave while α is the phase angle. A wave can be represented as a vector in a two-dimensional axial system, with the length of the vector as the amplitude and the angle with the real axis as the phase angle. Adding waves is thereby the summation of vectors and can be calculated accordingly (Figure 1a). Depending on the difference in phase angles, the amplitude of the summated wave can be any values between 0 and the simple summation of the amplitudes of the original waves. For example, if two waves have the same amplitude A and are in phase (with the same phase angle), the amplitude of their summation will be $2A$, which is defined as constructive interference. If two waves are 180° out of phase, the amplitude of their summation will be zero, which is destructive interference.

When an incoming X-ray wave encounters an electron, it will be scattered as a photon is first absorbed by the electron and then re-emitted at the same wavelength. In a system with multiple electrons, each electron will scatter X-rays, and of the scattered X-ray waves should be summated according to the vector algebra described above. In a crystal, because of its periodicity, only certain scattered reflections are allowed to exist. This situation can be explained as follows (Figure 1b). Suppose we could cut a crystal

into a series of parallel reflection planes with space d , and there exists a parallel incident X-ray light. The X-ray beam has wavelength λ and an angle with the plane θ . Since there are almost infinite numbers of such planes in a crystal, it is easy to imagine the only condition for any scattered light is that all scattered light is exactly in phase, or because of the destructive interference they will eventually be cancelled out with each other. To fulfill this requirement, the path difference of incident and scattered light from each plane should be an integral of the incident wavelength λ , which is called the Bragg's law and is expressed in the following form.

$$2\sin\theta = n \lambda / d \quad (n \text{ is an integral.}) \quad (1. 2)$$

Bragg's law can be used to predict where the scattered light (called reflection) will appear in space. For a three dimensional crystal, the sets of reflection planes can be described by three integers h, k, l . The reflection generated by each set of planes is described by its structural factor $F(h, k, l)$.

For a crystal whose unit cell contains large numbers of electrons, the structural factor of each reflection is the Fourier transform of the electron density in the cell.

$$F(h, k, l) = V \int \int \int \rho(x, y, z) \exp [2\pi i(hx+ky+lz)] dx dy dz, \quad (1. 3)$$

where ρ is the electron density and V is the unit cell volume; x, y, z is the coordinates within the unit cell.

And because of the properties of the Fourier transform, one can also calculate the electron density, if the structural factor $F(h, k, l)$ is known.

$$\rho(x, y, z) = (1/V) \sum \sum \sum |F(h,k,l)| \exp [-2\pi i(hx+ky+lz)+i\alpha(h, k, l)] \quad (1. 4)$$

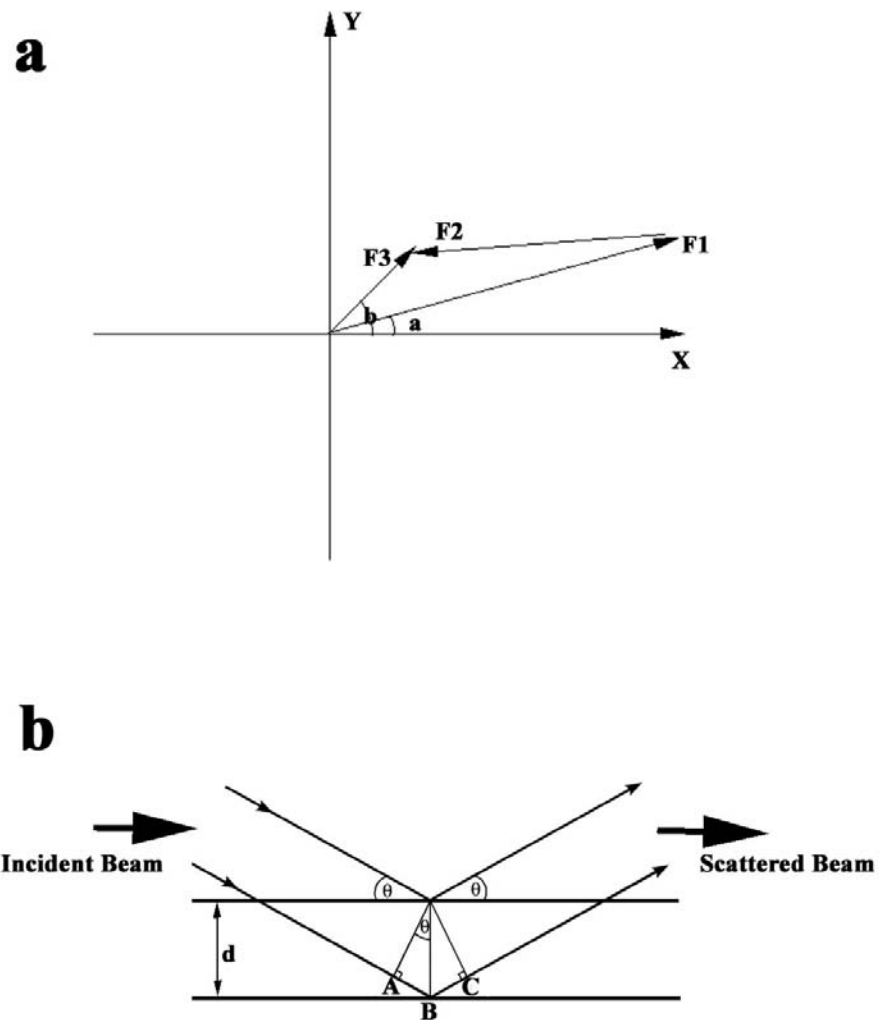


Figure 1 Basics of Diffraction.

a. The Basics of Vector Summation

Vector F1 and F2 are summated to generate vector F3. The phase angle of F1 is “a” and the phase angle of F3 is “b”.

b. Bragg’s Law

The extra distance of the second beam (AB+BC) should be an integer multiple of the wavelength, in order for the scattered beams to be constructively interfered.

($|F(h, k, l)|$ is the amplitude of the structural factor and can be directly converted from the measurable intensity of each reflection. $\alpha(h, k, l)$ is the phase angle of the structural factor. As $\alpha(h, k, l)$ can not be determined directly, it has to be inferred from either experimental phasing methods or a structural model.)

When the phase information is from the structural model, a 2Fo-Fc map is calculated, in which the amplitude is $2|F(h,k,l)_{obs}| - |F(h,k,l)_{cal}|$, which can decrease the possible bias of the map towards the model. ($|F(h,k,l)_{obs}|$ and $|F(h,k,l)_{cal}|$ are the amplitudes measured from the observation and that calculated from the model, respectively; $\alpha(h, k, l)_{cal}$ is the phase angle calculated from the model.) The equation is given in the following form:

$$\rho(x, y, z) = (1/V) \sum \sum \sum (2|F(h,k,l)_{obs}| - |F(h,k,l)_{cal}|) \exp [-2\pi i(hx+ky+lz) + i\alpha(h, k, l)_{cal}]. \quad (1.5)$$

Data collection and processing

In most crystallography labs, the X-ray radiation is generated using a copper based rotating anode that produces a specific wavelength near 1.54 Å. This approach is limited by low intensity and the inability to adjust wavelength. This problem is commonly overcome by synchrotron radiation that is generated from circulating electrons traveling near the speed of light. Synchrotron can produce X-rays at least two orders of magnitude stronger than that of conventional sources. Synchrotron radiation is also tunable that is it can produce radiation over a wide range of wavelengths, and a suitable wavelength can be selected using a monochromator.

The X-ray detector is the equipment used to record each reflection during the diffraction experiment. Image plates and CCD are the most commonly used devices. The data collection process is normally performed at low temperature using liquid nitrogen, and the crystal is protected by a cryoprotectant that prevents the formation of ice on the crystal. Data collection at low temperatures limits the diffusion of hydroxyl free radicals, which minimizes the radiation damages done to the crystal.

Crystals are characterized by their internal symmetry. Data processing is used to determine the symmetry of the crystal and assign the indices (h, k, l) of each reflection. This process is divided into two steps. First indexing is performed in which each frame is processed and the crystal is assigned to belong to one of the 14 Bravais lattice. The indices of the reflections are determined at this step and parameters, such as detector to crystal distance, crystal orientation, beam position, are also refined. The second step is scaling, in which reflections throughout the dataset are brought to the same scale and symmetrically equivalent reflections are averaged. Space groups can also be determined at this stage. The program HKL2000 (13) can accomplish both indexing and scaling. Other programs are also available, including MOSFLM (indexing) (14) and SCALA (scaling) (15).

Obtaining experimental phases

As discussed earlier, two strategies are available to determine phase angles. We will discuss the experimental phasing method here, and the molecular replacement method will be described in the next section.

The attachment of heavy atom(s) can remarkably affect the intensity of the reflections, and can be explored to infer the protein phase angles. There are two general approaches to obtain experimental phases from heavy atoms: by simply comparing the intensity difference between derivative and native crystals or by exploring the anomalous scattering.

MIR is the phasing method that adopts the first approach. Its strategy is to first calculate the amplitude contributed by heavy atoms ($|F_H|$) by subtracting the amplitude of the native crystal from that of the derivatized crystal. A special type of Fourier transform, called the Patterson function, is then used to calculate a Patterson map for the heavy atoms, using only the heavy atom amplitude $|F_H|$.

$$P(u, v, w) = (1/V) \sum |F_H|^2 \cos [2\pi(hu+kv+lw)] \quad (1.6)$$

The Patterson map is a vector map and the peaks on this map represent the vector pointing from one atom to another in the unit cell. The actual atom positions can be solved from the Patterson map provided their number is small, which is the case for heavy atoms. The locations of heavy atoms lead to their phase angles (α_H), and with the native and derivative amplitudes, can be used to solve the protein phase angles (α_P).

Anomalous scattering is enhanced when the incident X-ray wavelength approaches the absorption edge of the element. Consequently, the scattered light will have an imaginary component, whose phase angle is 90° retarded from the normal component.

$$f = f^0 + f' + i f'' \quad (1.7)$$

(f is the overall atomic scattering factor. f^0 is for the “normal” scattering. For anomalous scattering, f' is the component that has same phase angle with the “normal” scattering, and f'' is the component that has phase angle 90° retarded from the “normal” scattering. i means f'' is the imaginary component.) With the presence of anomalous scattering, the amplitude of the Bijvoet pair ($|F(h, k, l)|$ and $|F(-h, -k, -l)|$) will be different. This difference, called anomalous difference, can provide information for phasing purposes.

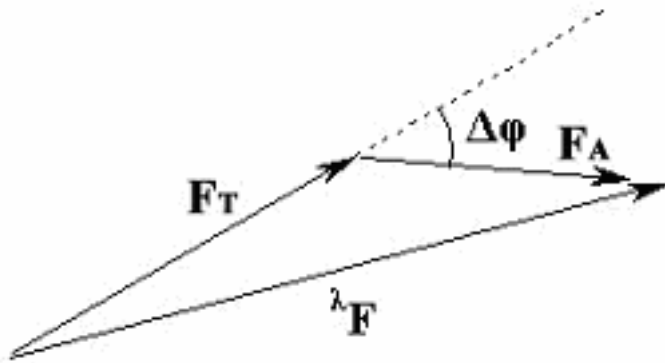


Figure 2 Principle of MAD.

F_T is the vector for normal scattering from all atoms; F_A is the vector for anomalous scattering; ${}^\lambda F$ is the vector for total scattering under a particular wavelength; and $\Delta\phi$ is the phase angle difference. During the MAD experiment, the amplitude of the total scattering ($|{}^\lambda F|$) is observed.

The most commonly used method for this purpose is MAD (Figure 2), in which datasets are collected at multiple wavelengths. Two wavelengths near the absorption edge are of particular importance. They are λ_1 (peak), where f'' has the highest value, and λ_2 (inflection), where it is at the absorption edge and f' has the most negative value.

Normally, one or two more wavelengths at least 100eV from the absorption edge are also needed as high or low energy remote sets. Data from a MAD experiment are sufficient to locate anomalous scatterers and calculate protein phase angles. Because only one crystal is needed for structural determination, MAD circumvents the problem of non-isomorphism and is very powerful for solving protein structures.

Many programs are available for solving structures. One such program is SOLVE (16). It provides an automatic approach to solve Patterson functions and is used for both MIR and MAD experiments. For MAD structural solution, it first analyzes the Patterson map calculated using $|F_A|$ (Figure 2) to find a few heavy atom positions. These partial heavy atom structures are used as the “seed” for further calculations, during which a difference Fourier map is made and top peaks are picked as additional heavy atom sites and added into the partial structure. The new partial structure is then refined and scored. The whole process is iterated until a complete heavy atom structure and protein phases are produced. Each solution is evaluated by Z-scores that estimate its agreement with Patterson functions, the cross-validation from the difference Fourier, the figure of merit and the boundary between solvent and macromolecule. In general, for two heavy atoms, a good solution should have an overall Z-score at least 5-10. Recently, direct methods, previously only used in small molecule structure determination, have been implemented (17) in programs such as SHELXD (18) and RANTAN (19) to locate heavy atoms. They have advantages over Patterson based methods when there are many heavy atoms (e.g. more than 20) in an asymmetric unit.

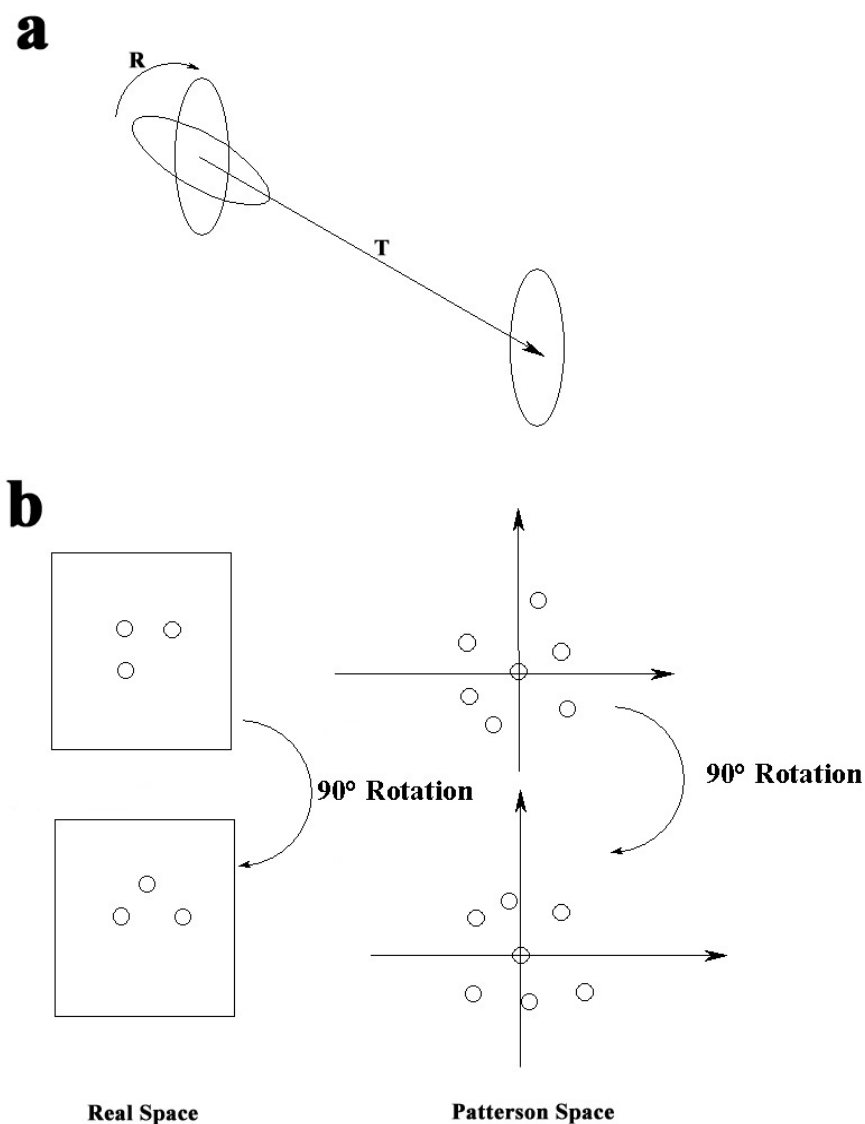


Figure 3 Principle of Molecular Replacement.

a. The Procedure of Molecular Replacement

The model (left) is first rotated (R) to generate proper orientation. It is then translated within the unit cell (T) to be located to the proper position (right).

b. The Property of the Patterson Map

A three atom molecule is shown in left (real space) and the intramolecular vectors derived from the molecule are shown in the right (Patterson space). As the molecule rotates 90° , the corresponding intramolecular vectors also rotate 90° .

Normally, the initial heavy atom sites are refined, so the protein phase angles calculated based on these sites will be more accurate. Traditionally, this step is performed based on a least-squares approach. More recent programs, in particular SHARP (20), implement a maximal likelihood based method, which avoids the problem of biases and the improper assumption regarding the native dataset. The latter is particularly important for MAD datasets, as there is no real “native” dataset from such experiments.

Phasing by molecular replacement

Available protein structures, if sufficiently similar, can be used as a model to provide phase information that, in combination with the experimental amplitudes, may be used to calculate the electron density map for the unknown protein. Normally, the space group and unit cell of the model structure are different from that of the unknown protein. The approach to correctly locate the model structure is called molecular replacement.

The position of a model is defined by its three rotational and three translational parameters. In order to reduce the computational cost, most molecular replacement methods, such as AMORE (21), CNS (22) and MOLREP (23), divide the process into separate rotational and translational search steps (Figure 3a). The first step is based on a property of the Patterson function in which the intramolecular Patterson vectors only depend on the orientation of the model and will rotate as the model rotates (Figure 3b). Therefore, by rotating the model, one can compare the Patterson function calculated

from the model with that from the data, and a significant overlap should indicate the correct orientation of the model.

During the translational search step, the correctly orientated model will move throughout the unit cell to find the maximal correlation coefficient that indicates the correct solution. The success of molecular replacement is mostly dependent on the quality of the model, both its similarity with the unknown structure and its completeness (the fraction of the asymmetric unit the model covers).

The above approach is limited by possible errors during the rotational search step. For a less perfect model, this traditional approach may fail to generate a clear solution. Although a simultaneous search of all six parameters may overcome these limitations, this type of search is computationally costly. One program that has partly solved this problem is the evolutionary based program EPMR (24). During an EPMR run, a population of trial solutions with six random parameters is first generated. The correlation coefficient of each solution is calculated and compared. Those with the highest correlation coefficient are the “fittest” solutions and are retained as “survivors”. These solutions will produce the next generation by introducing normally distributed “mutations” in their six parameters, and the process continues for many generations thereafter. For some poorer models, a true solution could only be found after running more than thirty generations.

Phase improvement (Density modification)

In many cases, the initial phases can be improved to generate a better electron density map that aids in simplifying the model building. One of the most commonly used

methods is solvent flattening, which is based on the knowledge that in an electron density map, there is greater density area in the region where the protein molecule occupies and less density area in the bulk solvent region. If the solvent-protein boundary can be identified, the electron density and the protein phase angles can be improved significantly simply by applying a low constant to the solvent region (25). This approach has been implemented into many programs such as DM (26), RESOLVE (27) and CNS. A related program SOLOMAN (28) also improves the electron density, but instead of applying a constant to the solvent region (flattening), it inverts the density in the solvent region by a flipping factor (flipping).

Another important phase improvement approach is based on the fact that there is often more than one molecule in an asymmetric unit, and some of these molecules might be identical (e.g. homodimers). Therefore, NCS operators can be identified and by averaging the electron density of the equivalent copies, density and phase angles can be improved (29). This approach is called molecular averaging and has also been implemented in programs DM, RESOLVE and CNS. The combination of solvent flattening (or flipping) with molecular averaging is very efficient at removing the phase ambiguity associated with experimental phasing. These methods are necessary for producing an interpretable electron density map when data from only one heavy atom derivative (SIR) or one wavelength of anomalous diffraction (SAD) are available.

If phases are from a structural model, the resulting electron density map may be in danger of being biased by the model, which might lead to erroneous conclusions. There are methods, such as omit maps (calculate a $2F_o - F_c$ map using a partial model

structure with the questionable part deleted), perturbation of the model (“shaking”) and map averaging, that substantially decrease the model bias. Shake ‘n’ wARP (30) is one program that implements several of these approaches in combination, including omitting part of structure, “shaking” the model, maximum likelihood refinement, real space placing and removing dummy-atoms from the density then averaging six maps derived from differently perturbed models. As a result, the final electron density map from shake ‘n’ wARP run have less phase bias and overall have a better clarity and contrast.

Model building and refinement

Model building is performed in a graphic program, such as SPOCK (31), O (32) and XtalView (33). For a map generated from experimental derived phases, manual fitting of structural elements into the electron density is performed, as no structural model is available. Normally, one will first generate secondary structure elements (α -helix, β -strand) composed of poly-alanine, using programs such as MOLEMAN (34) and fitting these elements into the density. The alanine residues can then be “mutated” into actual amino acids, once the protein sequence in that region is recognized. Sometimes before model building, a skeleton (bones) representation of the electron density is generated by programs such as MAPMAN (35), which can simplify the view of the map. Model building can also be partially done automatically using ArpWarp (36), RESOLVE or TEXTAL (37), though manual residue adjustments and regional rebuilding are still necessary to complete the model. If the map is generated from molecular replacement derived phases, one will have a mostly correct model to start with. Therefore, the model building process is focused on the rebuilding of the initial model.

Regions that do not fit the density are deleted and rebuilt, and side chains are “mutated” according to the actual sequence.

The protocol for model refinement is dependant on the abundance and quality of the data. If the resolution is very low (e. g. 4.0 Å or lower), the data to parameter ratio is low and rigid body refinement programs, such as CORELS (38), are more useful. In the normal resolution range, atomic coordinates and isotropic temperature factors are refined in the programs CNS or REFMAC (39). In these methods, stereochemical restrains are incorporated and the relative weight between the stereochemical and X-ray restrains can be adjusted according to the data and model quality. Generally, a higher weight is assigned to the X-ray restrain when the resolution is high and the opposite is true when the resolution is low. If the resolution is even higher (better than 1.5 Å), anisotropic temperature factor refinement or even unrestrained refinement may be performed, because of the high data to parameter ratio. Simulated annealing, available in programs CNS and PHENIX (40), can sometimes be performed to free the model from the local minimum.

The progress of model refinement is monitored by the R factor, which evaluates the difference between amplitudes observed from data and those calculated from the model. The equation is given as the following:

$$R = (\sum |F(h, k, l)_{\text{obs}}| - |F(h, k, l)_{\text{cal}}|) / (\sum |F(h, k, l)_{\text{obs}}|). \quad (1. 8)$$

In order to check the validity of each refinement step and avoid over-refinement, about 5-10% data are randomly selected as the free set before the first round of refinement (41). The rest data are assigned as the work set to be used for refinement. The R factor

calculated using the data from the work set is designated as R_{work} , while the R factor calculated based on the data from the free set is R_{free} . After each refinement step, both R_{work} and R_{free} should decrease. The divergence of these two R factors indicates over-refinement of the model and should invalidate the particular refinement step that leads to this divergence.

Theory of Structural Based Drug Design

Virtual screening

High throughput screening based on combinatorial libraries has already become the backbone for the pharmaceutical industry. The complimentary approach is virtual screening, in which the structure of the target protein is used for screening *in silico* over a large compound database. This approach has the benefit of lower costs, more compounds to screen and higher speeds. Successes include the designing of inhibitors against the carbonic anhydrase II, farnesyl transferase, DNA gyrase, estrogen receptor and thrombin (42). At its current technological level, virtual screening is not expected to correctly rank all possible ligands. Instead, the goal is to identify compounds worthwhile for further characterization or optimization. The overall process of virtual screening can be divided into four steps: i) preparation of a compound library and protein target; ii) molecular docking; iii) post processing; iv) *in vitro* testing for inhibition.

In the first step, the compound library is prepared, usually by various filters that limit the number of compounds for docking. One general filter is the “rule of five”. It states that a drug-like molecule should have molecular weight less than 500, logP value less than 5, less than 5 hydrogen donors and less than 10 hydrogen acceptors. This rule is

a good predictor of the bioavailability of the compound. Other filters are more specific and are based on the analysis of the substructures associated with known inhibitors, and the chemical stability or toxicity. Because many libraries only supply a two dimensional structure, the compounds have to be converted into three dimensions using programs such as CONCORD (42). Care also needs to be taken when properly assigning the tautomers and ionization state of the chemical.

In some cases, the homology model of a protein can be used as the target; however, the most commonly used target is an experimentally derived structure, normally from X-ray diffraction or NMR. There is a quality requirement for these structures. For example, X-ray crystal structures should have an R factor lower than 0.25, resolution higher than 2.5 Å and good overall geometry (43). In addition, the active site chosen for docking should not have high temperature factors. Water molecules in the crystal structure are generally deleted, unless they play important roles in the protein structure or ligand binding (42).

The ideal approach for docking is to treat both protein and ligand as flexible entities. However, because of the limitation of computing power, most programs only explore the flexibility of the ligand. One common approach is the incremental construction algorithm implemented in the program FLEXX (44). During docking by FLEXX, the ligand is first divided into small fragments. A base fragment is first selected and docked into the active site as a rigid body. The remaining fragments are added to the base fragment to incrementally build the ligand. This approach is capable of exploring

the flexibility of the ligand in reasonable computing hours and has shown to reproduce about 70% of experimentally determined protein-ligand complex structures (44).

The docking results are monitored by scoring functions that predict how well the ligand binds in a particular docked pose. Scoring functions can be based on a force field, such as AMBER or CHARMM, which provide the advantage of accuracy but are generally slow for calculation (42). Other scoring functions are empirically based. They use an additive approximation and are faster (42). The scoring function provided by FLEXX is one example. However, because they are derived from binding affinity data from a small set (about 100) of protein-ligand complexes, bias towards these datasets is inevitable (42). To address this issue, knowledge based functions such as PMF or Drugscore have recently been developed. They are derived from the structures of protein-ligand complexes. Because it does not rely on binding affinity data, these functions are derived from a much larger data pool and are less biased (42).

These scoring functions are not perfect. To minimize the number of false positives, consensus scoring is generally performed in the post processing step (42). For example, during the FLEXX docking, in addition to the build-in FLEXX scoring function, a series of other scoring functions, including ChemScore, G_Score, D_Score and PMF_Score, can also be produced. A consensus score is given on a scale of 1 to 5 after evaluating all of these scores. A high consensus score suggests the agreement among different scoring functions, indicating the correctness of the scoring and the reliability of the docking result. A low consensus score generally indicates the

inadequacy of the scoring procedure that has been used and, as a result, the corresponding docking result should at best be treated with extreme caution.

***De novo* drug design and fragment based screening**

Both high throughput screening and virtual screening are based on a large library of drug-like molecules. One common problem is that many of these compounds have relatively large molecular weights (ranged from 250 to 600 Da) and contain functional groups that do not directly contribute to the binding of the target molecule. As summarized from the “rule of five”, larger molecules are generally associated with poor bioavailability. The current lead optimization process routinely increases the molecular weight on average about 80 Da (45, 46). Therefore, the hits from either high throughput screening or virtual screening may not necessarily lead to drug candidates.

Methods have been developed to directly design drug molecules, based on the binding properties of individual fragments. Computer based *de novo* design basically places fragments into the binding pocket of the target protein (45). Subsequent optimization links different fragments to build the potential ligand. Programs implemented with this approach include LUDI, LEAPFROG, SPROUT, etc. This field is still under development. The difficulties include the inaccuracy of docking and the poor “synthesizability” of the resulting compounds.

A more experimentally oriented approach is fragment based screening (46, 47). Instead of “drug-like molecules”, fragments whose molecular weights are within the range of 100 to 350 Da are tested. This approach is heavily dependent on structural biology techniques, such as NMR and X-ray crystallography, because fragments only

exhibit low binding affinities (mM to 30 μ M) and are difficult to detect using bioassays (46). The structures of the target protein in complex with the fragments provide the actual binding mode of the fragment, and drug leads can then be identified by either evolution from the fragment or assembling several fragments together. Compared with high throughput screening, the fragment based method is focused on the ligand efficiency, meaning a high value for the average free energy for binding per atom (46). In contrast to hundreds of thousands compounds tested in high throughput screening, only hundreds or at most a few thousands fragments need to be tested in the fragment based approach. As structural information is directly available for the fragment based method, fewer synthesis attempts might be needed for optimization, which can also be cost effective.

Overview of Chapters

Our general interest is to investigate the structure and function of biologically important proteins using X-ray crystallography. Proteins that bind nucleotides play important roles in biological systems. They either function as enzymes using the bound nucleotide as the substrate, such as kinases, or as regulatory or structural proteins in which hydrolysis of the bound nucleotide acts as a “switch” for protein conformation and, lastly functionality, such as G proteins, tubulin, actin, etc. This dissertation is focused on two nucleotide binding proteins: MVK and FtsZ, with the former as an example of a kinase and the latter as a case of a structural protein.

Chapter I already presented an overview of structural biology, particularly the discussion on X-ray crystallography and structural based drug design. It has provided an

explanation of the techniques used and prepares the reader for the description of results generated using these methods in the following chapters.

Chapter II and III are focused on MVK. Chapter II gives the background information about terpenoid synthesis and the properties of MVK. Chapter III discusses the structures of both apo and ligand bound forms of MVK. It also provides a detailed description of the processes of MAD phasing on MVK crystals. The goal of this chapter is to establish the fold of MVK within the GHMP kinase family and also to establish its mechanism of substrate binding, catalysis and inhibition. To achieve this, kinetic studies are also incorporated. The final part of this chapter is virtual screening and hit validation using MVK as a target protein. Some initial hits have been identified and attempts to optimize them are summarized. The limitation of the current virtual screening procedure is also discussed.

Chapter IV and V are focused on FtsZ. Chapter IV introduces the background of bacterial cell division and the role that FtsZ plays. Chapter V describes the investigation into the polymerization of FtsZ. It starts from the characterization, both by EM and biochemical methods, of the *M. tuberculosis* FtsZ polymerization. The main part of this chapter is the description of a crystal structure of FtsZ at 4.0 Å that reveals a double helical crystal filament. The significance of this macromolecular assembly is discussed. Finally, a working model for FtsZ mechanism based on the conformational changes upon GTP hydrolysis is suggested. The process of molecular replacement phasing and structural refinement at a low resolution is also discussed in detail.

Chapter VI is the final chapter. It provides a brief summary on the research described by this dissertation as well as the future directions that can still be pursued.

CHAPTER II

BACKGROUND OF TERPENOID BIOSYNTHESIS AND MVK

Terpenoid Biosynthesis

Terpenoids

Terpenoids are the most diversified chemicals. More than 20,000 terpenoids have been characterized and hundreds more are reported every year (48). Important biologically active compounds, such as sterols, retinol and dolichol belong to the terpenoid family. Other biologically important molecules, such as ubiquinone, chlorophyll and Ras protein (49) are modified by terpenoids, which is necessary for their membrane association and proper function. Many drugs, including Taxol (50) and Artemisinin (51), are derived from terpenoids. Natural rubber, as a widely used material, is also a terpenoid. (Table 1)

Terpenoids are produced from the polymerization of a five carbon unit – IPP. The successive head to tail polymerization of IPP forms GPP, FPP and GGPP. The further cyclization creates a family of highly diversified molecules. Many of these are natural products that are produced from plants and insects and probably have their original roles as defensive substances or pheromones. Two FPP molecules can also undergo head to head condensation to form a C₃₀ compound that is then converted into squalene. The cyclization of squalene is the first step that precedes a series of reactions leading to the formation of cholesterol.

Table 1 Representative Terpenoids.

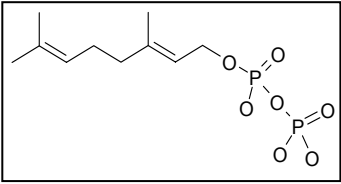
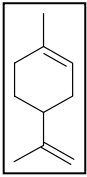
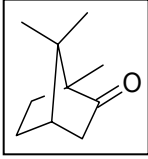
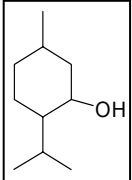
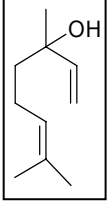
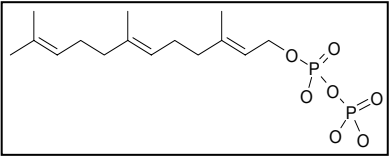
Structure	Name	Comments
	GPP	Precursor of Monoterpenes
	Limonene	Oil from Citrus Rind
	Camphor	Oil from Camphor wood
	Menthol	Main Component of Peppermint Oil
	Linalool	Main Component of Indian Basil Oil
	FPP	Precursor of Sesquiterpenes and steroids

Table 1 Continued.

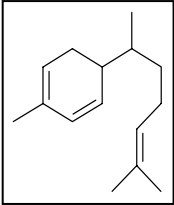
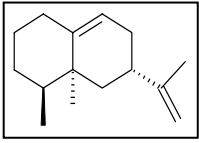
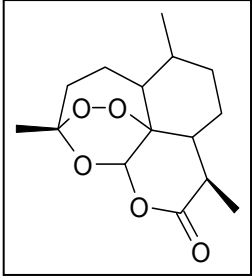
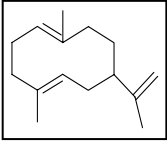
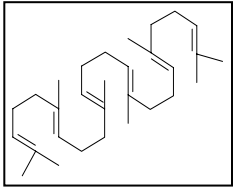
Structure	Name	Comments
	Zingiberene	Essential Oil from Ginger
	5-epi-Aristolochene	Precursor of antifungal phytoalexin capsidiol
	Artemisinin	Anti-Malaria Drug
	Germacrene A	Precursor of Many Sesquiterpenes
	Squalene	Precursor of Steroids

Table 1 Continued.

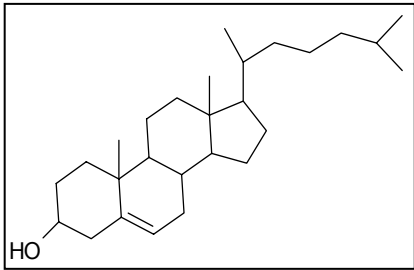
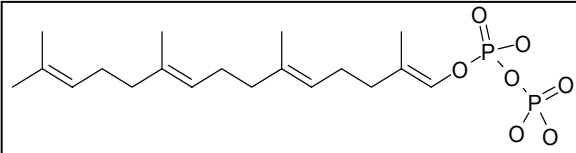
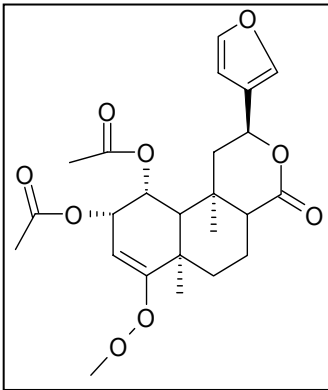
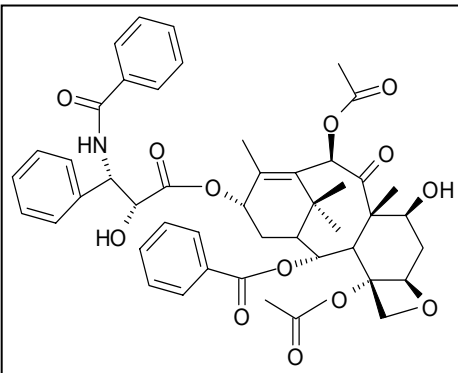
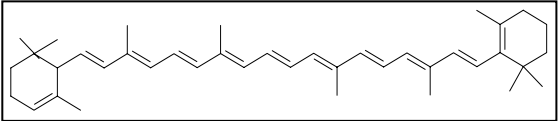
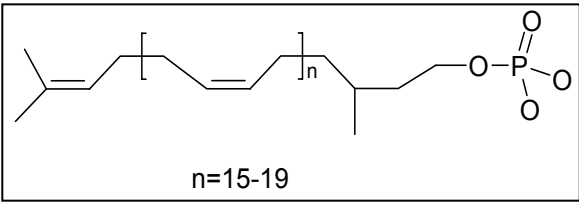
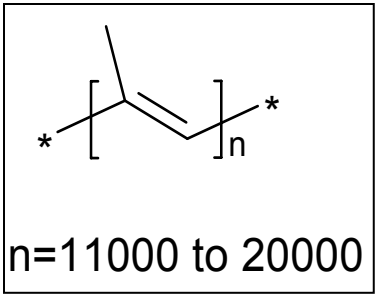
Structure	Name	Comments
 <p>The structure shows the characteristic four-ring steroid nucleus of cholesterol. It features a hydroxyl group (-OH) at the 3-position, a double bond between carbons 5 and 6, and a branched hydrocarbon side chain at the 17-position.</p>	Cholesterol	Eukaryotic Membrane Stabilization
 <p>The structure represents geranylgeranyl pyrophosphate (GGPP), a linear tetraterpene precursor. It consists of a long hydrocarbon chain with four methyl branches and two double bonds, terminated by a pyrophosphate group (-O-P(=O)(O-)-O-P(=O)(O-)-O-).</p>	GGPP	Precursor of Diterpenes
 <p>The structure shows Salvinorin C, a complex bicyclic sesquiterpene. It features a decalin core with multiple oxygen-containing functional groups, including a furan ring, a methoxy group, and several ester and ether linkages.</p>	Salvinorin C	Extracted from Hallucinogenic Mexican Mint
 <p>The structure depicts Taxol (paclitaxel), a complex polycyclic diterpene. It has a complex tetracyclic core with numerous oxygen-containing functional groups, including hydroxyl groups, ester groups, and a phenyl ring.</p>	Taxol	Anti-Tumor Drug

Table 1 Continued.

Structure	Name	Comments
	all-trans- β -Carotene	Precursor of Vitamin A
 <p style="text-align: center;">n=15-19</p>	Dolichol Phosphate	Transport Oligosaccharide
 <p style="text-align: center;">n=11000 to 20000</p>	Natural Rubber	Material

The biosynthesis of IPP is present in all three domains of life. The mevalonate dependent pathway is present mostly in eukaryote and archaea (52). It uses acetyl-CoA its starting block. The non-mevalonate pathway was discovered in the 1990S as the alternative pathway used by most bacteria and plant plastids (53). The non-mevalonate pathway uses pyruvate and glyceraldehydes 3-phosphate as the carbon source. Some gram positive bacteria, however, use the mevalonate pathway instead (52).

The mevalonate pathway

The first step of this pathway is the formation of acetoacetyl-CoA by the condensation of two molecules of acetyl-CoA, catalyzed by acetoacetyl-CoA thiolase. HMG-CoA is then produced from acetyl-CoA and acetoacetyl-CoA by HMG-CoA synthase. The next step is to reduce HMG-CoA to mevalonate by HMG-CoA reductase. Mevalonate is converted to 5-phosphomevalonate by MVK, then to 5-pyrophosphomevalonate by phosphomevalonate kinase. The last step is a decarboxylation reaction catalyzed by pyrophosphomevalonate decarboxylase, which forms IPP from 5-pyrophosphomevalonate (Figure 4). IPP can then be converted to DMAPP by IPP isomerase.

Among enzymes involved in the mevalonate pathway, HMG-CoA reductase is the best characterized (54). Type I HMG-CoA reductases are distributed in eukaryotes and archae and have an N-terminal transmembrane anchoring domain (54). Type II enzymes are distributed mostly in bacteria. They lack the transmembrane domain but have a unique C-terminal flap domain instead (54). Type I HMG-CoA reductases are the target for statins, a class of popular drugs used for cholesterol control. All statin drugs

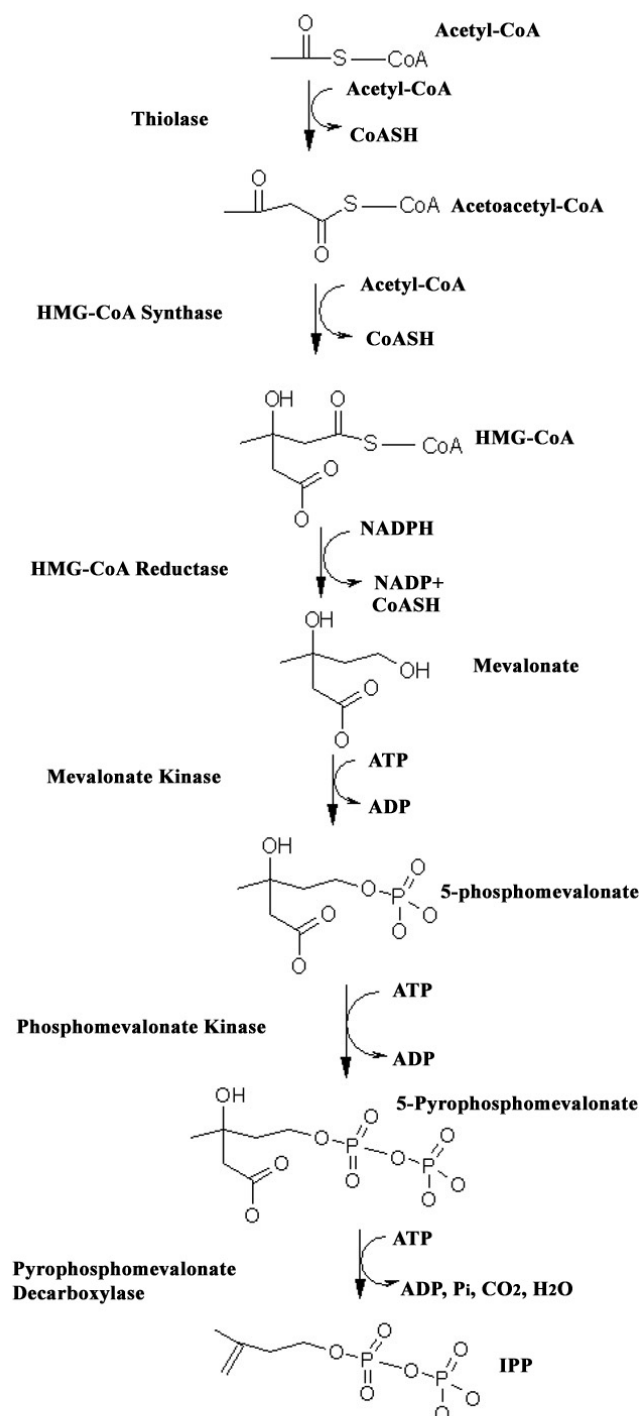


Figure 4 The Mevalonate Dependent Pathway for IPP Biosynthesis.

are composed of a HMG-like moiety and a hydrophobic ringed moiety. Structures of statin and human HMG-CoA reductase complexes show that the HMG-like moiety binds to the HMG binding pocket (55). The hydrophobic moiety of statins binds to a shallow hydrophobic groove that is part of the CoA binding site. This hydrophobic groove is flexible, as it undergoes reorganization according to the hydrophobic moiety of the drug, which may explain the properties of different drugs. As the enzyme for the limiting step, HMG-CoA reductase is regulated by a variety of factors. For example, insulin can stimulate its dephosphorylation, which promotes its activity. Glucagon has the opposite effect (56). The activity of HMG-CoA reductase is also regulated by its synthesis and degradation. As a feedback control mechanism, cholesterol is able to reduce the expression of HMG-CoA reductase as well as to promote its degradation (56).

MVK will be reviewed later. Phosphomevalonate kinases can be classified into two distinct families. The majority of this enzyme, distributed in plants, fungi and some bacteria, belongs to the well characterized GHMP family, as with MVK. The second group of the phosphomevalonate kinase, distributed in animals, belongs to the P-loop kinases family. Pyrophosphomevalonate decarboxylase catalyzes the conversion of phosphomevalonate to IPP. The reaction involves the transferring of the phosphate group from ATP to the 3'-hydroxyl group of 5'-pyrophosphomevalonate, which is followed by the decarboxylation and the leaving of the phosphate group. It is still unclear about the exact mechanism of this reaction (57).

The non-mevalonate pathway

The first step in the non-mevalonate pathway is the condensation of one pyruvate with one glyceraldehydes 3-phosphate to form DXP, catalyzed by the thiamine dependent enzyme DXP synthase. DXP is then converted to MEP by DXP reductoisomerase. A CMP moiety is transferred to MEP to form CDP-ME by CDP-ME synthase. CDP-ME is then phosphorylated by CDP-ME kinase, producing CDP-ME2P. Cyclization of CDP-ME2P by MECDP synthase forms MECDP. The ring opening reaction on CDP-ME2P is then catalyzed by HMB-PP synthase. The product, HMB-PP, is reduced by HMB-PP reductase, which produces IPP and DMAPP (Figure 5).

The first enzyme in this pathway, DXP synthase, shares homology with transketolases as well as the E1 component of pyruvate dehydrogenase. These proteins are thiamine dependent enzymes that transfer a 2-carbon unit. Recent studies suggest that ketoclofomazone, a breakdown product of herbicide clomazone, is an inhibitor of DXP synthase (58). The second enzyme DXP reductoisomerase catalyzes a rearrangement reaction in which the C-4 carbon migrates to the C-2 atom. One potential drug targeting DXP reductoisomerase is the herbicide fosmidomycin that inhibits this enzyme by mimicking its substrate (59). Recently, fosmidomycin was found to show good activities against *Plasmodium* spp., the pathogenic agent of malaria (60).

CDP-ME synthase, CDP-ME kinase and MECDP synthase have all been characterized structurally (61, 62, 63). The structure of MECDP synthase shows a homotrimer, with the substrate binding site at the interface between two monomers. Interestingly, in its central cavity there is an additional density that resembles a GPP

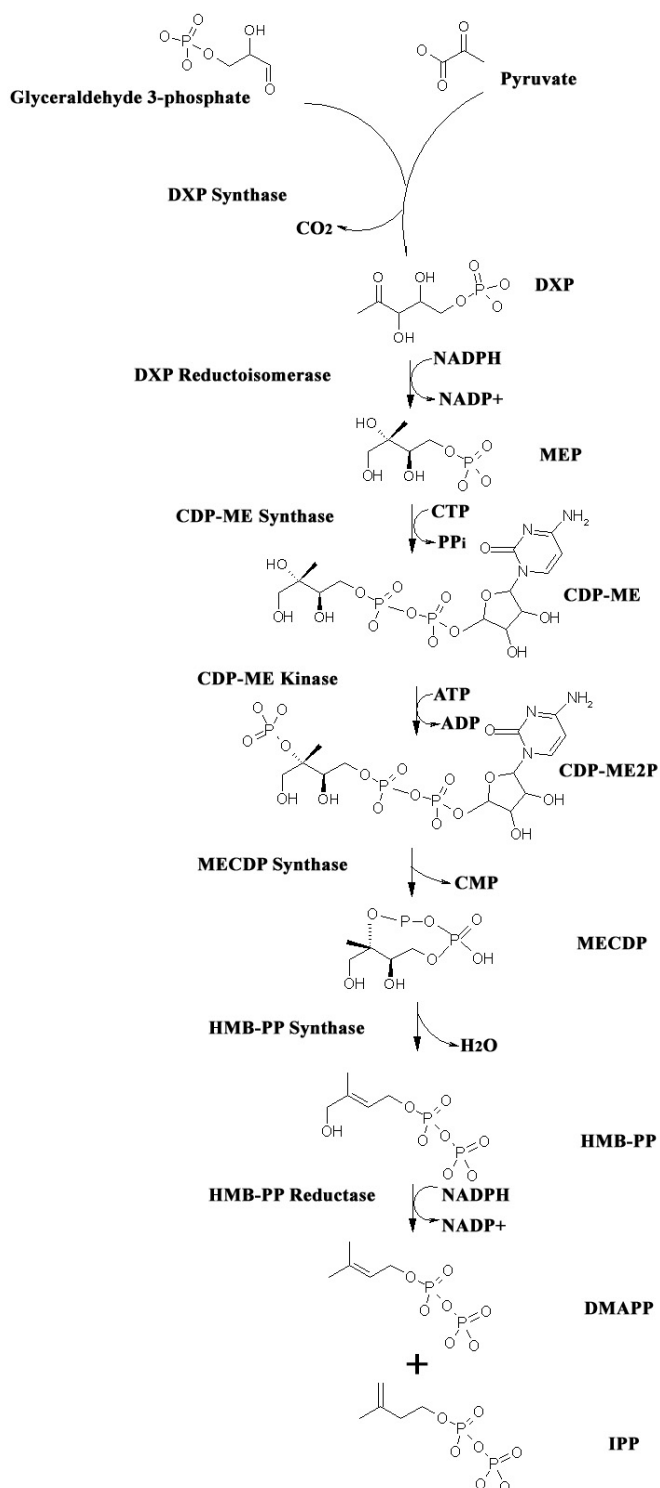


Figure 5 The Non-Mevalonate Pathway for IPP Biosynthesis.

molecule. It is possible that GPP plays some regulatory roles in the enzyme activity, as also occurs in MVK. HMB-PP synthase (GcpE in *E. coli*) converts MECDP to HMB-PP. Two cysteine residues are indicated as catalytic residues (64). The formation of a disulfide bond between these two residues is presumed to be essential for catalysis. HMB-PP reductase (LytB in *E. coli*) is a NADPH dependent enzyme and catalyzes the reduction of HMB-PP (64). This step generates both IPP and DMAPP, in about 5:1 ratio (64). A carbocation intermediate was proposed to be involved in the reaction and responsible for forming both products (64).

Chain elongation

Beginning with DMAPP, successive additions of IPP unit creates a family of polyprenyl diphosphate, including GPP, FPP and GGPP. The chain elongation is catalyzed by prenyltransferases (also known as polyprenyl diphosphate synthase). The majority of those are E-prenyl diphosphate synthase that create *trans* products (65). The E-type enzymes all contain a conserved aspartate-rich motif DDXXD that binds the substrate via bridging Mg^{2+} ions (66). Other prenyltransferases belong to the Z-type that catalyzes the formation of the *cis* double bond. These enzymes do not have sequence homology with E-prenyltransferases. DDXXD motifs have not been identified in Z-type enzymes. Instead, they may use an aspartate and a glutamate to bind polyprenyl pyrophosphate and IPP, via a bridging Mg^{2+} ion, as demonstrated by the crystal structure of UPPS (67).

The mechanism for chain elongation is an electrophilic alkylation (68) (Figure 6a). First, the leaving of the pyrophosphate group from DMAPP or polyprenyl

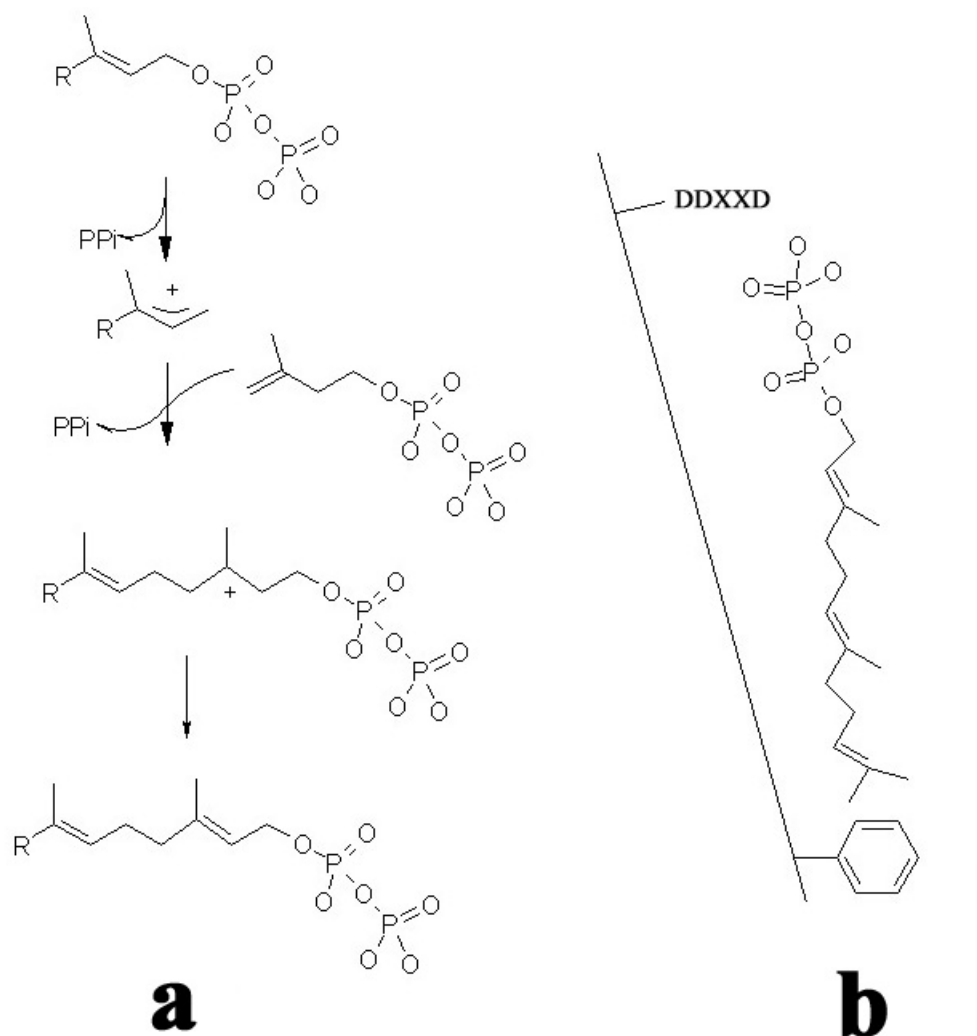


Figure 6 Prenyltransferase.

a. The Mechanism of the Prenyl-transfer Reaction

b. Chain Length Control Mechanism of Prenyltransferase

The terpenoid chain binds to a long tunnel and its length is limited by the distance between the DDXXD motif and a bulky residue (represented by a phenylalanine side chain in this figure).

pyrophosphate creates a carbocation that makes an attack on the C3-C4 double bond of IPP. The carbocationic intermediate is then eliminated, creating a double bond at C2 position. The chain-length of the final product is determined by the position of a bulk residue within the active pocket, as the “molecular ruler” (65, 68) (Figure 6b). In E-prenyltransferases, a bulk residue located before the first DDxxD motif forms the floor of the pocket. Enzymes that give different product lengths usually have a bulky residue at different positions. For example, FPP synthase has a phenylalanine at the fifth position before its first DDXXD motif, which prevents the formation of products longer than fifteen carbons (66). The heptaprenyl pyrophosphate synthase has an isoleucine located at the eighth position before its first DDXXD motif, which therefore produces a longer product with 35 carbons (69).

Protein prenyltransferases is another class of prenyltransferase and transfers a polyprenyl moiety (FPP or GGPP) to the thiolate group of a cysteine residue within the CAAX motif of the target protein (70). Ras farnesyltransferase (Ras FTase) is one example. It transfers a farnesyl moiety to the protein Ras, a small G-protein that is important for signal transduction and carcinogenesis. As the farnesylation of Ras is essential for its function, inhibitors against Ras FTase have shown inhibition against tumor cell growth and may possess the potential to be developed into anti-tumor drugs (71).

Cyclization

Most terpenoids are polycyclic, produced by cyclization of linear polyprenyl diphosphates, which involves three steps (48): i) the creation of a carbocation; ii) the

rearrangement of the carbocation to create the prototype of the product; iii) the quenching of the carbocation. Cyclases use different strategies to complete each step. For example, the carbocation can be created by releasing the pyrophosphate group, protonation of a double bond or protonation of an epoxide. It is generally quenched by deprotonation and formation of a double bond. The product diversity is usually determined by how the substrate folds in the active pocket, rearrangement of the carbocation and the final location where the carbocation is quenched. In cyclases, catalytic residues are aligned in a particular manner that one type of carbocation intermediate is stabilized (72, 73).

One example for cyclization is the synthesis of steroids. Steroids are a family of chemicals that share a common core structure, composed of four rings (three 6-member rings and one 5-member ring). Cholesterol is the most important member of this family. It is an essential component of eukaryotic membranes and also plays important roles in the development of cardiovascular diseases. In bacteria, hopenes (with four 6-member rings and one 5-member ring) are produced instead of steroids. Squalene is the precursor of both steroids and hopenes. In bacteria, squalene-hopene cyclase initiates the cyclization by protonating the C2-C3 double bond of squalene and creating a carbocation (74). In eukaryotes, squalene monooxygenase first oxygenates the C2-C3 double bond, creating 2, 3-oxidosqualene. The protonation of the 2, 3-epoxide creates the carbocation and initiates the cyclization, which is catalyzed by lanosterol synthase (74). The bacteria squalene-hopene cyclase has been well studied (74, 75). The crystal structure shows that it is a homodimers with three cavities: two small water-filled

cavities and one large hydrophobic centre cavity. The substrate binds to the centre cavity. Conserved residues aligned around these cavities contribute catalysis by stabilizing carbocations intermediates at various stages. Mutations on these residues will produce alternative products with fewer rings, which indicate the premature quenching of the carbocation (74).

MVK

Enzymatic properties

MVK catalyzes the phosphorylation of mevalonate to phosphomevalonate. Enzymes from mammalian, bacterial and archaea sources have been characterized (76, 77, 78, 79). The size of MVK is around 40 KD. Size-exclusion chromatography indicated it mostly forms a homodimer in solution (76, 77, 79). In *E. faecalis*, however, MVK exists predominantly as a monomer (78).

The ordered sequential reaction mechanism of MVK was established by a series of studies in the hog enzyme (80). Product MgADP has been shown competing with ATP. The second product, 5-phosphomevalonate, was found to be a noncompetitive inhibitor when either MgATP or mevalonate is the variable substrate. The product inhibition pattern, combined with studies on terpenoid inhibition (81), suggested that mevalonate is the first substrate to bind and mevalonate-5'-phosphate is the first product to be released (Figure 7).

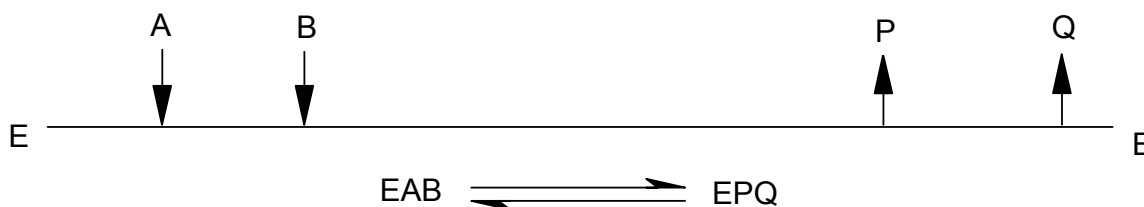


Figure 7 The Reaction Mechanism of MVK.

E is MVK, A is mevalonate, B is ATP, P is 5-phosphomevalonate and Q is ADP.

The activity of MVK is dependent on divalent cations. Mg^{2+} , Mn^{2+} , Zn^{2+} , Ca^{2+} and Co^{2+} were found to support the enzyme activity (80). Studies on hog enzymes showed that when metal ion concentration was lower than 1 mM, Mn^{2+} was the best activator. Mg^{2+} was found to show the best stimulatory effect when the concentration was higher than 10 mM. Enzyme activity was also stimulated by reducing agent. 1 mM dithiothreitol or 10 mM β -mercaptoethanol was sufficient to prevent the loss of activity by oxidation. It was also found that the addition of thiol group binding agents, such as p-hydroxymercuribenzoate, inhibited the enzyme. These data suggested that there is an important cysteine residue involved in enzyme activity (80).

Although ATP is probably the preferred phosphoryl donor, as demonstrated by studies in the yeast enzyme (81), alternative nucleotides, such as ITP, GTP, CTP and UTP, are also used by MVK. Studies in the *E. faecalis* enzyme suggested that the K_m values for all four nucleotides are similar (78). The hog enzyme, however, was reported

to only use ATP and IPP as phosphoryl donors (82). Therefore, MVKs from different species seem to have a degree of variability in the selectivity of nucleotides.

MVK is also inhibited by downstream terpenoid products, such as GPP, FPP, GGPP, dolichol monophosphate, etc (81, 83). Terpenoids are competitive with ATP and uncompetitive with mevalonate. This probably suggests a feedback control mechanism at the post-translational level. A recent study also showed that pyrophosphomevalonate, the product of phosphomevalonate kinase, inhibits the *S. pneumoniae* enzyme as a noncompetitive inhibitor (84). This compound does not inhibit the human enzyme, which suggests that it may be developed into an anti-bacteria agent.

General mechanisms for phosphoryl transfer

Most kinases form a ternary complex in which the phosphoryl donor and acceptor are in the “in-line” geometry. Therefore, the transfer of the phosphoryl group takes place directly between two substrates and the reaction mechanism is sequential (85). The exact mode for phosphoryl transfer is still subject to debate. Two possibilities have been proposed. In the associative mechanism (S_N2), a pentagonal bipyrimidal transition state is formed at the γ -phosphorus. Both the phosphoryl acceptor (the nucleophile) and the leaving group (ADP) are at the apical positions, while the other three γ -phosphate oxygens are in the equatorial positions (86). In the dissociative transition state (S_N1), the bond between the β - and γ -phosphate breaks first to produce a trigonal planar metaphosphate transition state that is then captured by the phosphoryl acceptor (86). Experimental evidence so far is not conclusive for either of the mechanisms (86).

Almost all kinases need metals, mostly Mg^{2+} , for catalysis. Generally, Mg^{2+} is coordinated by the β - and γ - or the α -, β -, and γ -phosphate groups, as well as protein residues and ordered water molecules. The coordination of Mg^{2+} polarizes the bond between the β - and γ -phosphate and makes the γ -phosphorus atom more electrophilic (86). It also orients the γ -phosphate so it is “in-line” with the phosphoryl acceptor as required by the reaction (86, 87). Kinases also use several mechanisms to stabilize the transition state (86). In both the P-loop kinases and GHMP kinases, the γ -phosphate group is located in adjacent with the positive charged N-terminus of an α -helix, which may neutralize the negative charges accumulated on the γ -phosphorus during the phosphoryl transfer (86, 88). A conserved positively charged residue (e. g. lysine) is often found near the β - and γ -phosphate groups, which may play a similar role (86, 88).

Catalytic basis of MVK

Since this dissertation is partly focused on MVK, its catalytic mechanism is investigated in this study and will be discussed in detail in the following chapter. Here only a brief review on this subject is presented.

Mutagenesis studies on MVKs, particularly those from the human and rat, established the basis for catalysis. Although many conserved residues have been found involved in the enzyme activity, two are most prominent based on the profound activity deficiency caused by their mutations. The first one is a fully conserved aspartate residue (Asp 204 in the human protein). Mutation of this residue to either an asparagine or alanine decreases V_m for more than 10,000 fold (89). As will be discussed later, a general base is used in MVK as the catalytic residue. Therefore, this conserved aspartate

residue has been suggested to play this role. The other one is the second serine residue in the invariant GLGSSA motif (equivalent to Ser 146 in the human enzyme). Its mutation to an alanine decreases V_{\max} for more than 1000 fold and also weakens the binding of Mg^{2+} -ATP (90). This residue was suggested as a potential ligand for Mg^{2+} .

Regulation and cellular localization of MVK

In addition to the feedback control by terpenoids, MVK is regulated at the transcriptional level, coordinated with HMG-CoA reductase. For example, diets with lovastatin (an statin class HMG-CoA reductase inhibitor) causes the level of MVK protein and mRNA to increase, while diets supplemented with 5% cholesterol causes both protein and mRNA level to decrease (91). The basis for transcriptional regulation is in the promoter region containing two Sp1-binding sites and one SRE. SRE also exist in the genes of HMG-CoA reductase and LDL receptor, indicating it as a major regulatory element in response to the sterol level (91). The fact that MVK is regulated both transcriptionally and post-translationally suggests this enzyme is a regulatory point, second to HMG-CoA reductase, in the mevalonate dependent IPP biosynthesis.

Studies in the hog and rat MVK suggest that this enzyme is located in the peroxisomes. This organelle also contains other enzymes involved in terpenoids biosynthesis, including HMG-CoA reductase, HMG-CoA synthase, acetoacetyl-CoA thiolase, phosphomevalonate kinase, pyrophosphomevalonate decarboxylase and FPP synthase, indicating the compartmentation of the terpenoid biosynthesis, in particular the mevalonate dependent pathway, in the peroxisome (92).

Deficiency of MVK

Mutations in MVK cause deficiency in IPP biosynthesis, which leads to diseases with the symptom varied from mild to life-threatening (91, 93, 94, 95). Mevalonic aciduria is caused by the total loss of MVK activity. Patients with mevalonic aciduria have psychomotor retardation, hypotonia, ataxia, recurrent episode of fever and even development delay, cataracts and diarrhea. The victims of the most severe form of mevalonic aciduria often die early in infancy. A much more benign symptom is HIDS. HIDS patients mainly suffer from episodes of fever, but lack the more serious developmental problems.

Mutations that cause these diseases have been mapped onto MVK gene. One of the most common mutations is V377I, as the cause for many HIDS cases (91). Interestingly, protein that carries this mutation shows about 35% residue activity when being expressed in *E. coli* but there are almost no detectable MVK proteins in fibroblasts from the patient, which probably suggests that this mutation mainly affects the protein stability *in vivo* (91). Other mutations include H20P, T243I, I268T, A334T, etc (91, 94, 95).

What is the mechanism for these mutations to cause the disease? The recent published rat MVK crystal structure provides some clues to this question (96). Residues His 20 and Ala 334 are located within the active site and their mutations probably directly disrupt catalysis or substrate binding. Residue I268 is in a helix that is involved in dimer formation. Probably its mutation will disrupt the dimer and makes the protein less stable. More interesting is the residue Val 377. It is far from the active site and does

not have any apparent role in protein dimerization. Modeling an isoleucine into its position does not generate any clash with neighboring residues (96). Therefore, it is still unclear how this mutation decreases the stability of the protein.

Significance of This Work

MVK is a key enzyme in the terpenoid biosynthesis and an important regulatory point in the mevalonate dependent pathway. Its importance has been demonstrated by the diseases caused by its mutation. It has also been shown that MVK is indispensable for the growth of many gram positive bacteria (52), indicating it could be a target for drug designing. To study MVK, we expect to understand the structural basis for its substrate binding, catalysis and inhibition. We would also like to compare its structure with other members of the GHMP family in order to delineate the structural conservation and variation within this family. In the next chapter, we will present a study of the *M. jannaschii* MVK that analyzes this enzyme from both structural and biochemical perspectives. Our results shed light on the structural mechanism of this important enzyme.

CHAPTER III

STRUCTURAL STUDY OF MVK

Methods

Protein expression and purification

The gene for the *M. jannaschii* MVK was initially cloned into the pET28a vector with an N terminal his-tag and a T7 promoter. The construct was provided by our collaborator A I. Scott (76) and was used to transform B834(DE3) *E. coli* strain. Cells were grown in LB media containing 100 µg/ml kanamycin at 37 °C until OD₆₀₀ was around 0.8 when the gene expression was induced by IPTG at 1 mM concentration. Cells were then grown at 18 °C for 20 hours, harvested and frozen at -20 °C.

In order to perform a MAD experiment, the selenomethionine substituted protein was produced. Cells were grown in LB media containing 100 µg/ml kanamycin at 37 °C until OD₆₀₀ was around 0.8. Cells were harvested, washed with M9 minimal media (with 100 µg/ml kanamycin and 50 µg/ml 19 amino acids except methionine) twice, and then grown in this media for 1 hour at 37 °C. Selenomethionine and IPTG were then added at 50 µg/ml and 1 mM concentration, respectively. After this step, the procedure was the same as the native protein production.

To purify the protein, cell pellets were thawed and re-suspended in 20 mM Tris-HCl (pH 7.5), 500 mM NaCl, 1 mM PMSF, 20 mM imidazole and lysed with a French press. The addition of PMSF is mandatory, because MVK has the tendency to be cleaved by proteases into two fragments (about 28 KDa and 7 KDa, respectively). Since *M.*

jannaschii MVK is a thermostable protein, with optimal activity at around 70 °C, the cell extract was heated to this temperature for 15 minutes, which denatured most *E. coli* proteins but fully retained MVK activity (76). After centrifugation at 16,000 G for 30 minutes, the supernatant was passed through a 0.45 µm filter and then applied to a nickel affinity column (AP Biotech) that was eluted with an imidazole gradient. MVK was eluted from the column when imidazole concentration reached 180 mM. Fractions were checked with SDS-PAGE and showed a major band with purity level estimated to be above 95%. Its position is consistent with the predicted molecular weight of the His-tagged MVK (37340 Da). Fractions with MVK were pooled and dialyzed with the protein storage buffer (5mM Tris, 1mM DTT, pH 7.5). Finally, MVK was concentrated to around 15mg/ml and stored at -20 °C. MVK shows stable activity for at least one year under this storage condition. The yield of MVK is about 10 mg protein per liter culture.

Assay of MVK activity

(RS)-mevalonate was initially purchased as mevalonic acid lactone. To prepare a stock solution with 50 mM of mevalonate, 65 mg mevalonic acid lactone was dissolved in 3 M KOH and heated to 37 °C for 1 hour to completely hydrolyze the lactone. The pH was then adjusted to 7.0 by 0.2 M HCl and the final volume was 10 ml. A PK/LDH coupling assay system was used to measure MVK activity. The measurement was performed in a Cary 100 spectrometer. To measure the K_m of ATP, a reaction was performed in a 0.5 ml quazi cuvette containing 0.1 M $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$ buffer (pH 8.0), 250 µM NADH, 1 mM phosphoenol pyruvate, 5 mM MgCl_2 , 2.5 mM (RS)-mevalonate,

40 unit PK and 37.5 unit LDH, 0.55 μg MVK and 10 μM to 500 μM ATP. The background oxidation of NADH was monitored at 340 nm for 2 minutes before the reaction was initiated by the addition of MVK. The reaction was followed at 340 nm for another 2 minutes and the slope was recorded as the initial velocity. To measure the K_m of (RS)-mevalonate, 2.5 mM ATP and 10 μM to 500 μM (RS)-mevalonate was used instead. To test MVK activities on alternative nucleotides, 500 μM of GTP, CTP or UTP as well as 400 unit of PK were used in each reaction.

In order to measure the inhibition by putative inhibitors, 0.55 μg of MVK was first incubated with 40 nM to 200 μM of the inhibitor on ice for 5 minutes. The mixture was then added into the cuvette that contains all other assay components mentioned above. (RS)-mevalonate concentration was fixed at 2.5 mM. ATP concentration was fixed at 500 μM during the measurement of the IC_{50} and varied from 10 μM to 500 μM when global curve fitting method (see below) was used to determine the K_i (for ADP β S). To validate the hits from virtual screening, compounds were first solubilized in DMSO to make 1 mM stock solutions. During the reaction, 5 μl of the stock solution was included into the 0.5 ml reaction. For the blank, a reaction with 1% DMSO was performed. To rule out the possible inhibition to coupling enzymes, a 0.5 ml reaction containing 0.1 M $\text{KH}_2\text{PO}_4/\text{K}_2\text{HPO}_4$ buffer (pH 8.0), 250 μM NADH, 1 mM phosphoenol pyruvate, 5 mM MgCl_2 , 0.073 unit pyruvate kinase, 37.5 unit lactate dehydrogenase, 50 nM ADP and 10 μM inhibitors was performed.

Determination of kinetic values

The IC₅₀ of terpenoid inhibitors was determined by measuring reaction velocity under the inhibitor concentration at a range of 200 μM to 40 nM. The data were fit using following equation, with the non-linear regression program Kaleidgraph:

$$V = T/(1+[I]/IC_{50}) \quad (3.1)$$

where [I] is the inhibitor's concentration; V is the initial velocity; T is the top plateau for the curve. Since terpenoids are competitive inhibitors against ATP, their inhibition constant K_i can be calculated from IC₅₀ values using the following equation:

$$K_i = IC_{50}/(1+K_m/[ATP]) \quad (3.2)$$

where K_m is the Michaelis constant for ATP, while [ATP] is the concentration of ATP.

K_m values for ATP and mevalonate were determined by plotting the initial velocity with the substrate concentration. The data were fit with the Michaelis-Menten equation:

$$V = V_m [S]/(K_m+[S]) \quad (3.3)$$

where V_m is the maximal velocity and [S] is the concentration of the substrate that varies.

The inhibition of ADPβS was studied by performing the reaction under various concentrations of ADPβS, as described above. The nature of competitive inhibition was revealed by the Lineweaver and Burk plot. The inhibition constant K_i was determined with global curve fitting in the program GraphPad Prism, using the following equation:

$$V = V_m[ATP]/(K_m+[I]K_m/K_i+[ATP]). \quad (3.4)$$

Crystallization

Crystallization was accomplished by the hanging drop vapor diffusion method. MVK apoprotein crystals were grown by mixing 2 μ l of the His-tagged protein with an equal volume of the reservoir solution and equilibrating against a 400 μ l reservoir. The crystallization condition contains 28 to 30% dioxane, 20 mM to 50 mM imidazole and 10 mM DTT. To improve the crystal quality, seeding was performed by transferring preformed crystals into new drops with the crystallization condition of 22% to 26% dioxane, 20 mM imidazole, 10 mM DTT and 10 mM $MgCl_2$. To produce crystals complexed with the nucleotide ligand, MVK was first incubated with 10 mM $MgCl_2$, 10 mM DTT, 0.5 to 2 mM nucleotide ligand and 2.5 mM mevalonate and then used for setting up the crystallization experiment, as described above. Apoprotein crystals were seeded into these drops. To produce MVK crystals complexed with terpenoid ligands, terpenoids stock solutions (about 2 mM concentration, in ethanol, as purchased from Sigma) were first dried in a SpeedVac low vacuum system and then resuspended to 20 mM in aqueous solution containing 0.6% n-octyl- β -D-glucopyranoside and 5 mM tris (pH 7.5). MVK was then incubated with 1.5 mM terpenoids, 10 mM DTT and 2.5 mM mevalonate. Crystals were grown using the seeding method, as described above, except that 10 mM EDTA instead of $MgCl_2$ was used in the crystallization condition, to prevent terpenoids from being precipitated by $MgCl_2$. The details of developing the seeding protocol will be discussed in the result section.

Crystallographic methods

Data were collected at 121 K. Cryoprotectant was prepared by mixing 2.5 μl MPD with 7.5 μl crystallization solution taken from the same reservoir well. The MAD experiment with a selenomethionine substituted apoprotein crystal was performed at the beamline 14BMD at APS in Argonne National Laboratory. The selenium absorption edge was first determined to be at 12.6618 Kev by fluorescence scanning of the crystal. A four-wavelength dataset (peak: 0.9792 \AA ; inflection: 0.9794 \AA ; high-energy remote: 0.9611 \AA and low-energy remote: 0.9919 \AA) was collected from that crystal with one-degree oscillation widths through a range of 360° for each wavelength (Table 2). Data for ligand bound crystals were collected at beamlines 14BMD, 14BMC, 19ID, 19BMD at APS and the protein crystallography beam line at CAMD. One-degree oscillation widths were collected through a range of 180° or when 95% completeness was reached (Table 3).

Data were processed using HKL2000. All crystals belong to the primitive monoclinic space group $P2_1$ with similar cell dimensions, though a 3 \AA reduction in the **a** axis is present in the ADP β S crystal. Matthews coefficient (97) calculation suggested only one molecule per asymmetric unit with the specific volume 2.2 $\text{\AA}^3/\text{Da}$. The solvent fraction is about 0.43. The apoprotein structure was determined in program SOLVE, followed by phase improvement with RESOLVE. The model was manually built in program SPOCK and O. The structure was refined in CNS. To determine the ligand bound crystals structures, the apoprotein structure was used as the model for refinement against the diffraction data in CNS or REFMAC. Phases were improved by the bias

removal program Shake 'n' wARP and the model was rebuilt in program XtalView and refined in CNS or REFMAC. The structure was validated by PROCHECK (98).

Table 2 Data Collection Statistics of the MAD Experiment.

$R_{\text{sym}} = \sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i I_i(\text{hkl})$, where $I_i(\text{hkl})$ is the intensity of an individual observation and $\langle I(\text{hkl}) \rangle$ is the mean intensity of reflection (hkl).

	Peak	Inflection	High Energy Remote	Low Energy Remote
Space Group	P2(1)	P2(1)	P2(1)	P2(1)
Cell Dimensions a, b, c (Å) α, β, γ (°)	56.14, 44.03, 64.85 90, 102.71, 90	56.14, 44.03, 64.86 90, 102.71, 90	56.14, 44.03, 64.85 90, 102.71, 90	56.16, 44.05, 64.86 90, 102.71, 90
Wavelength (Å)	0.9792	0.9794	0.9611	0.9919
Resolution (Å)	2.4	2.4	2.4	2.4
R_{sym}	0.063(0.126)	0.068(0.122)	0.061(0.126)	0.062(0.111)
I/σ(I)	20.9(6.5)	18.6(5.2)	21.2(6.7)	21.2(6.7)
Completeness (%)	94.5(72.1)	87.1(47.8)	96.2(79.2)	94.7(69.2)

Table 3 Statistics of Data Collection and Structural Refinement.

	MVK-AMPPCP-Mg ²⁺	MVK-UTP-Mg ²⁺	MVK-ADPβS-Mg ²⁺	MVK-ADP-Mg ²⁺
Data Collection				
Space Group	P2(1)	P2(1)	P2(1)	P2(1)
Unit Cell Parameters a, b, c (Å) α, β, γ (°)	55.887, 43.971, 64.446 90.0, 102.357, 90.0	55.849, 43.806, 65.081 90.0, 102.638, 90.0	52.777, 43.847, 66.249 90.0, 102.471, 90.0	55.998, 43.900, 65.020 90.0, 102.41, 90.0
Wavelength (Å)	0.9795	0.9793	1.3776	0.9789
Resolution Range (Å)	50.0-1.71	50.0-1.75	50.0-2.12	50.0-2.44
R _{sym}	0.045 (0.587)	0.046 (0.316)	0.106(0.564)	0.051(0.173)
Completeness (%)	99.6 (98.3)	97.9(90.3)	71.9(28.0)	99.4(95.3)
I/σ(I)	30.5(2.6)	37.9(3.9)	10.4(0.6)	28.1(9.2)
Refinement				
No. of Atoms/Residues				
Protein	317	317	317	317
Ligand	1	1	1	1
Metal	1	1	1	1
Water	114	172	42	95
Maximum Resolution (Å)	2.00	2.00	2.30	2.44
R _{work}	0.204	0.193	0.215	0.219
R _{free}	0.251	0.248	0.282	0.267
Average B factor (Å ²)	26.5	23.4	50.8	27.2
r.m.s.d from Ideal Geometry				
Bond Distance (Å)	0.0095	0.011	0.010	0.019
Bond Angles (Å)	1.57	1.77	1.53	1.68
Ramachandran Plot (%)				
Favored	92.0	92.3	87.1	91.3
Allowed	7.3	7.0	12.5	7.7
Generously Allowed	0.7	0.3	0.3	1.0
Disallowed	0.0	0.3	0.0	0.0

Methods for virtual screening

For the FLEXX docking, compound databases were downloaded from the Maybridge, NCI and ZINC websites. Docking was performed on four processors of SGI MIPS R12000 via the SYBYL7.0 graphic interface. For LUDI docking, the program as well as the library was accessible through INSIGHT. The active site for docking purpose was created by drawing a sphere around either the carbon atom between β and γ phosphorus of the AMPPCP ligand or the C_a atom of Ser 112, with a radius of 15Å. This covers the entire ATP binding site, as well as the potential mevalonate binding site.

The criteria for compound extraction were primarily based on docking scores (the FLEXX or LUDI scores). Compound scored within top 1% were extracted. Consensus scores, if available, were also considered. To determine the consensus score (CScore), docking scores were first calculated for each docked pose with several scoring functions during the docking process in FLEXX. The consensus among these functions was then evaluated in SYBYL. Compounds were excluded from visual inspection if their CScores were lower than 3 and from *in vitro* testing if they were lower than 4 (except for compound SPB01501). Visual inspection evaluated the binding modes and structures of docked compounds. In general, only one compound from those with similar the structures and docking scores was selected for validation.

Results and Discussion

Sequence analysis

MVK sequences from different species were studied by multiple sequence alignment, which revealed several strictly conserved residues (Figure 8). Mutation

studies had already shown that some of these residues, such as Glu 19, Ser 145, Ser 146, Glu 193, Asp 204 and Ala 334 (99) in the human enzyme and Lys 13 (100) in the rat enzyme are important for the enzyme activity. Among them, D204A decreases the V_m nearly 10,000-fold. In the *M. jannaschii* MVK, Glu 14, Ser 111, Ser 112, Glu 144, Asp 155, Ala 276 and Lys 8 are equivalent to the above conserved residues in the human and rat proteins, respectively. These results help to establish the probable mechanism of the enzyme and facilitate our structural analysis of the active site cavity.

Three motifs, designated as motif I, II and III, have been identified (76). They are also conserved among other GHMP proteins (101). Motif I is close to the N-terminus of the protein, including the residues from Pro 6 to His 15 in the *M. jannaschii* enzyme. It has the sequence PXXIXXGEH. Motif II corresponds to the region between Pro 104 to Ser 114 of the *M. jannaschii* protein with an invariant GLGSSA sequence and is generally recognized as the signature motif for GHMP kinases. It forms the phosphate-binding loop. Motif III is characterized by residues Lys 272 to Cys 281 of the *M. jannaschii* protein and contains an invariant sequence TGAGGGG.

Seeding improves the quality of MVK crystals

The MVK apoprotein crystals which formed directly were either tiny single crystals (less than 0.05 mm) or larger multiple crystals that were unsuitable for diffraction. Optimization trials by grid or additive screens did not improve the crystal. A seeding protocol was then tried by introducing preformed crystals into a new drop. Interestingly, massive nucleation was triggered as a result and a large number of small single crystals were observed. Therefore, we decided to focus on decreasing the

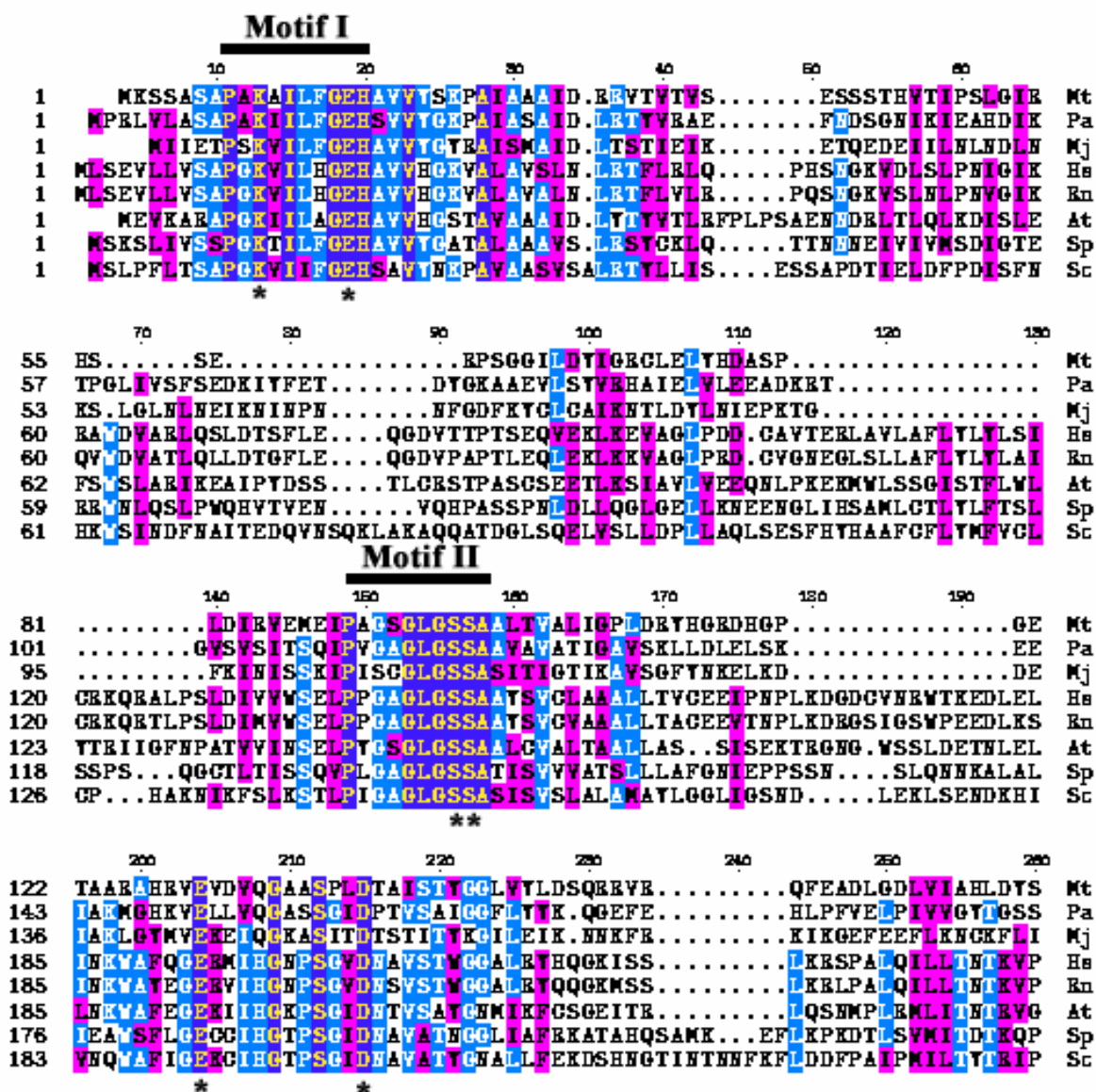


Figure 8 Multiple Sequence Alignment of MVKs from Various Sources.

Three conserved motifs are highlighted. The important residues are marked by asterisks. Sequences are color coded as *blue* (invariant), *cyan* (identical residues) and *magenta* (similar residues). Organisms in which these sequences are generated are coded as follows: Mt (*Methanobacterium thermoautotrophicum*), Pa (*Pyrococcus abyssi*), Mj (*Methanococcus jannaschii*), Hs (*Homo sapiens*), Rn (*Rattus norvegicus*), At (*Arabidopsis thaliana*), Sp (*Schizosaccharomyces pombe*) and Sc (*Saccharomyces cerevisiae*).

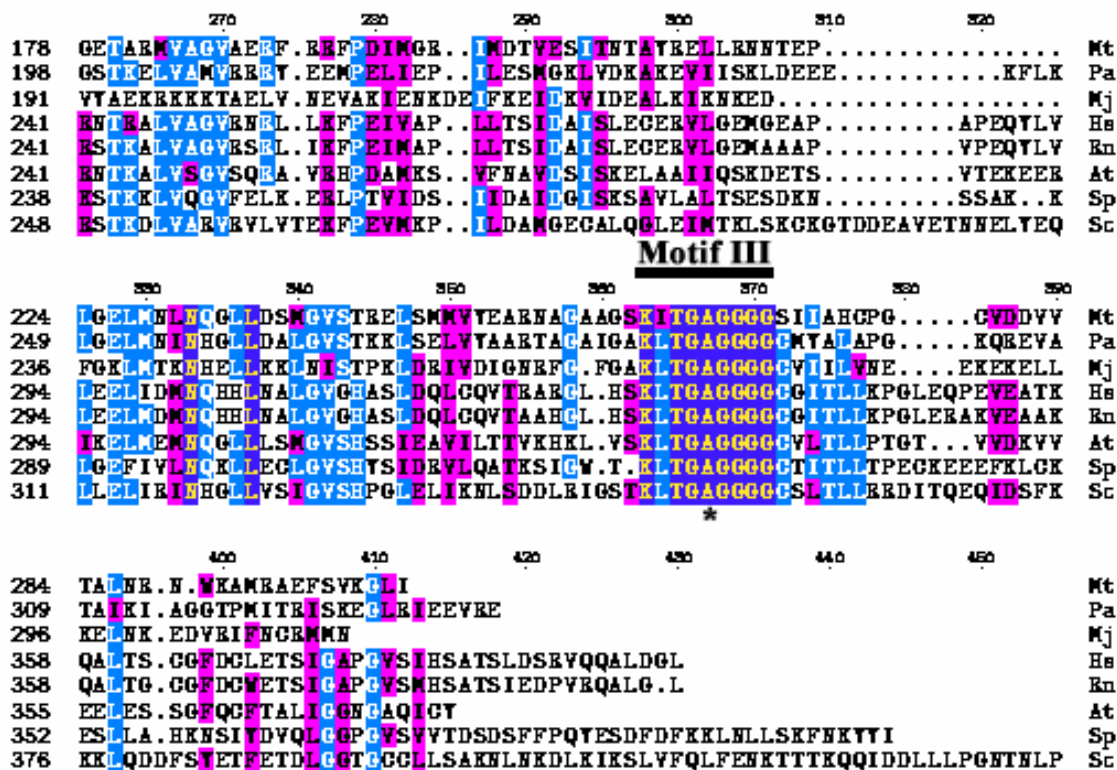


Figure 8 Continued.

nucleation. During the previous optimization process, Mg^{2+} was found to prevent crystal growth. Therefore, 10 mM $MgCl_2$ was included in the drop in which the seeds were to be transplanted. The dioxane concentration in the new drop was also lowered to between 22% and 26%. The nucleation did decrease and the fewer crystals grown from the seeds were much bigger (typically more than 0.1 mm). Diffraction test indicated that these

crystals diffracted to 3 to 4 Å using our home X-ray source. Their quality was independent of the morphology of the seeded crystal, which can be rationalized that they were actually grown out of invisible nuclei, probably released from the trunk of the seeded crystal as they underwent disintegration in the new crystallization drop. Basically this procedure can be regarded as a simplified form of the microseeding.

Having been incubated with the ligand, MVK seemed to have a lower propensity to crystallize. Therefore, we developed a heterogeneous seeding protocol by transplanting apoprotein crystals into drops containing MVK incubated with ligands. Compared with apoprotein crystals, crystals complexed with ligands were generally smaller (less than 0.1 mm) and more likely to be plane-like. Nevertheless, they had good diffraction quality at synchrotron radiation sources.

Apoprotein structural solution

The *M. jannaschii* HSK has about 20% sequence identity with the *M. jannaschii* MVK. The PDB coordinates of the HSK were downloaded from the protein data bank and used as the model for molecular replacement trial to solve the MVK structure, before we could collect the MAD dataset. Solutions from both rotational and translational search steps only had a low ratio for signal over noise, which indicated the failure of molecular replacement. After solving the MVK structure by MAD technique, it was found that the relative orientations between the N- and C-domain in MVK and HSK are apparently different, which probably makes HSK an unsuitable model.

The MAD data were first checked for the presence of anomalous scatters by calculating an anomalous difference Patterson map based on the peak dataset. Peaks due

to the presence of anomalous scatterers can be identified on the Harker section ($y=0.5$) (Figure 9). MAD phasing in SOLVE, using four wavelength data at the resolution range from 20 to 2.4 Å, located all six selenium sites in one asymmetric unit. The overall figure of merit is 0.65. The statistics of heavy atom sites indicate that the structural solution is correct (Table 4). The protein phases were improved with the electron density modification program RESOLVE. The map generated with the RESOLVE modified phases was less noisy. Furthermore, the density for protein chains was improved significantly. For example, the region around the loop Val 191 to Glu 194 was not well resolved in the map before density modification (Figure 10a). There was a break between residues Ala 193 and Glu 194. In addition, the side chain of Tyr 192 was largely missing. After the density modification, these problems were corrected (Figure 10b). The modified map was used for all subsequent model building.

Model refinement was performed in CNS, after up to 70% of residues were built. The first several rounds of refinement were simulated annealing, in which molecular dynamics in torsion angles were applied with a starting temperature of 2500 K. The model building process was alternated with the simulated annealing process, until the model building was complete, when the R factor was about 0.35 and R_{free} factor about 0.44. Further rounds of refinement, including minimization and individual B factor refinement, brought these two R factors to about 0.30 and 0.35, respectively. Water

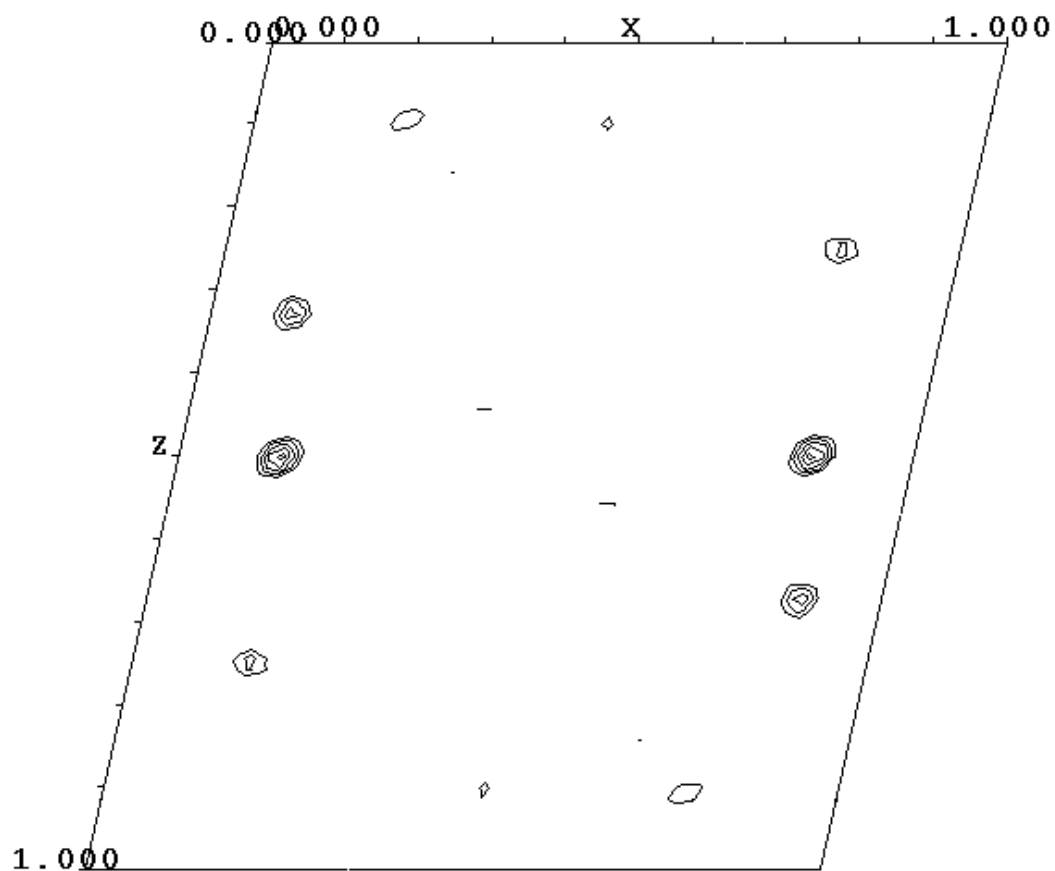


Figure 9 Anomalous Patterson Map.

Shown is the Harker section ($y=0.5$) of the anomalous difference Patterson map based on the peak data set. The map was calculated by xfft in XtalView. The resolution range used for calculation is from 3 to 6 Å and the outlier filter was set at 30% of average. The first contour level is at 3σ and each contour line thereafter represents 1σ in peak intensity.

Table 4 Selenium Sites and Their Statistics.

FOM is the mean value of the cosine of the error in phase angles.

Height/Sigma is the peak height from the cross-validation of the difference Fourier.

X	Y	Z	Occupancy	B Factor	Height/Sigma
0.431	0.006	0.251	0.912	24.7	17.8
0.419	0.083	0.126	0.927	33.8	15.2
0.458	0.028	0.423	0.926	60.0	10.9
0.639	0.486	0.349	0.802	60.0	9.0
0.472	0.932	0.295	0.745	27.5	14.1
0.043	0.785	0.381	0.381	28.9	7.9
FOM		0.65			
Overall Z Score		30.79			

molecules were then added by the automatic water picking method in CNS and individual occupancies of water molecules were then refined, in combination with more model rebuilding and structural minimization, which gradually brought the R factor down to about 0.22 and R_{free} to about 0.30. Two more additional densities were identified as dioxane molecules, the precipitant for crystallization, and were fit accordingly. Their group occupancies were refined to 1.06 and 0.78, respectively. Finally, the weighting factors for X-ray term weight (W_a) and for B factor refinement were also optimized. Final rounds of refinement using optimized weighting factors produced a structure with the R factor 0.20 and R_{free} factor 0.28 (Table 5).

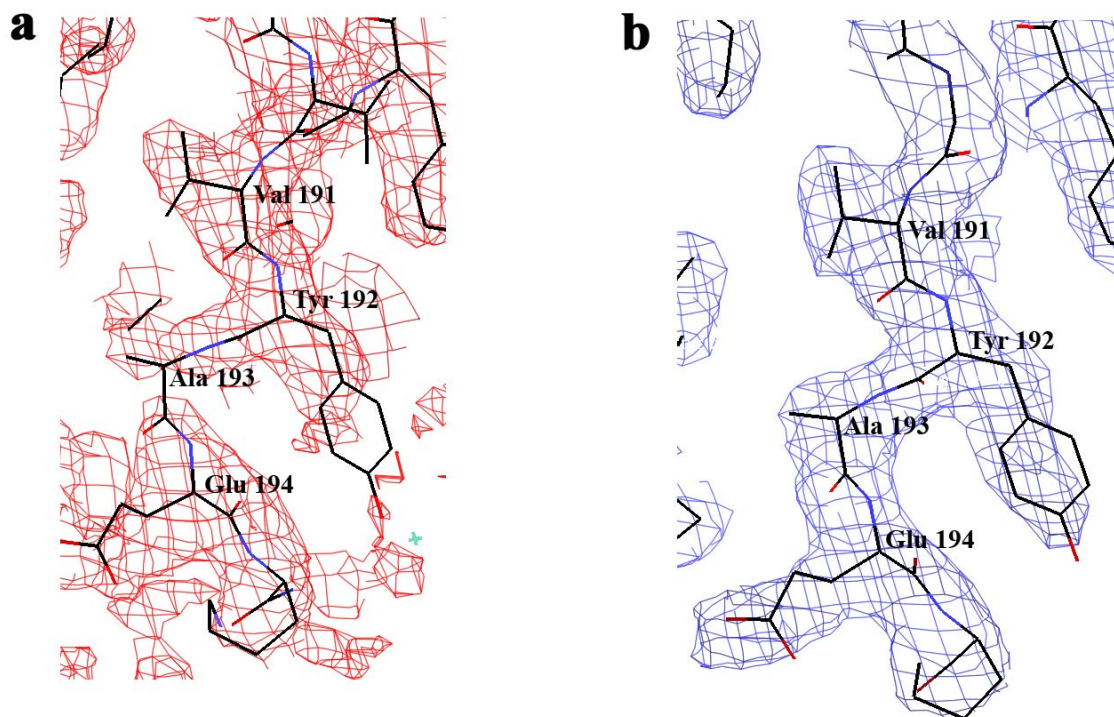


Figure 10 The Electron Density Maps before and after Density Modification.

a. The Electron Density Map (Contoured at 1σ) Based on Phases Generated by SOLVE

b. The Electron Density Map (Contoured at 1σ) Based on SOLVE Phases Improved by RESOLVE

Table 5 Refinement Statistics of MVK Apoprotein Crystal Structure.

No. of Residues	
Protein	317
Water	92
Dioxane	2
R_{work}	0.197
R_{free}	0.282
Average B factor (Å²)	32.5
r.m.s.d from Ideal Geometry	
Bond Distance (Å)	0.012
Bond Angles (°)	1.622
Ramachandran Plot (%)	
Favored	88.5
Allowed	10.8
Generously Allowed	0.7
Disallowed	0.0

Overall structure of the apoprotein

Similar to other GHMP kinases, the MVK monomer is a dumbbell-shaped molecule and contains two structural domains. Its size is about 60×30×25 Å (102). At the domain interface there is a cleft with the dimension about 30×15 Å, where the active site is located (Figure 11a). The secondary structural elements in this protein are listed in Table 6.

A DALI search (103) has been performed to investigate structural homologs of MVK. The top hits include the rat MVK (1KVK Z-score 32.5), galactokinase (1S4E Z-score 30.7), phosphomevalonate kinase (1K47 Z-score 28.4), HSK (1FWK Z-score 26.1),

CDP-ME kinase (1UEK Z-score 21.1), *C. elegans* sexual determinant Xol-1 (1MG7 Z-score 16.9), mevalonate 5-diphosphate decarboxylase (1FI4 Z-score 15.9). They all belong to the GHMP kinase family. With the structural superposition program TOPP

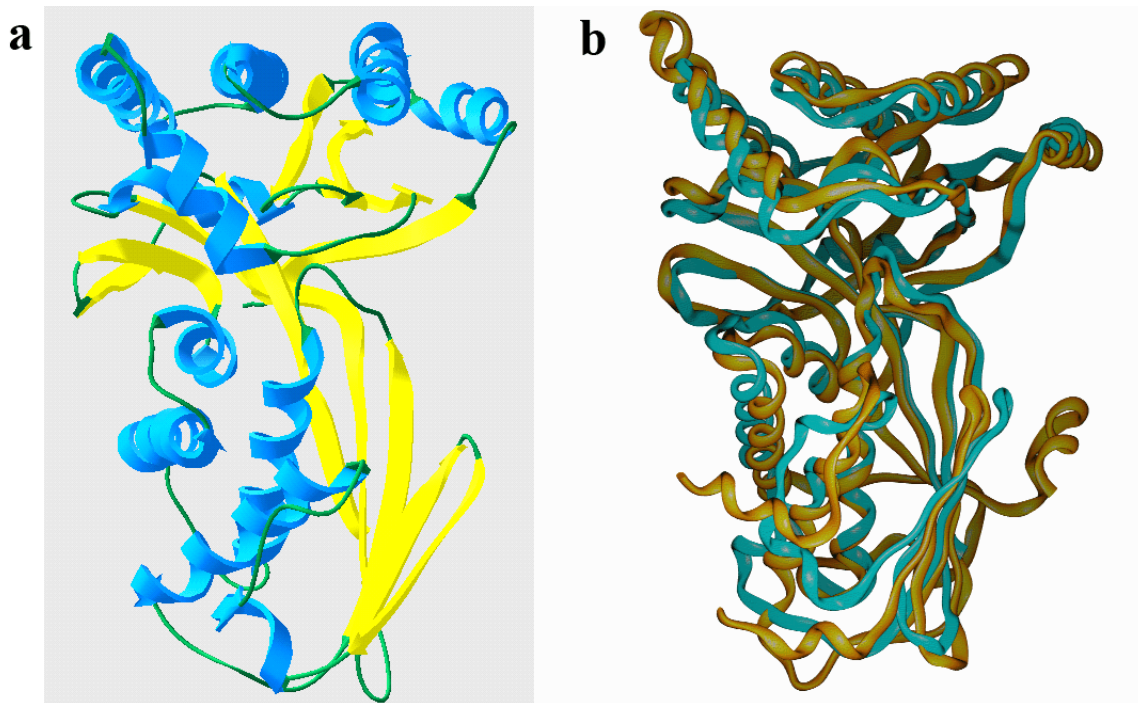


Figure 11 Rat and *M. jannaschii* MVK.

a. Structure of the *M. jannaschii* MVK

b. Structural Alignment of the Rat (golden) and *M. jannaschii* (cyan) MVK

(104), the structure of MVK has been aligned with CDP-ME kinase (173 residues match and r.m.s.d of 1.77 Å), homoserine kinase (191 residues match and r.m.s.d of 1.96 Å), phosphomevalonate kinase (174 residues match and r.m.s.d of 1.93 Å) and galactokinase

(204 residues match and r.m.s.d 1.63 Å). The structural similarity indicates the conservation among GHMP kinases.

The quality of the structure was checked with the program PROCHECK. From the Ramachandran plot, the geometry of the structure is good, with 88% non-glycine residues in the most favorable region (Figure 12). Notably, for a typical protein structure,

Table 6 List of MVK Secondary Structural Elements.

α-Helices	Residues
A	70 to 87
B	110 to 127
C	132 to 148
D	154 to 162
E	176 to 185
F	200 to 208
G	213 to 230
H	232 to 250
I	254 to 268
J	287 to 300
β-Strands	Residues
a	1 to 12
b	23 to 38
c	42 to 49
d	52 to 58
e	95 to 101
f	163 to 168
g	170 to 175
h	188 to 195
i	269 to 274
j	280 to 285
k	303 to 309

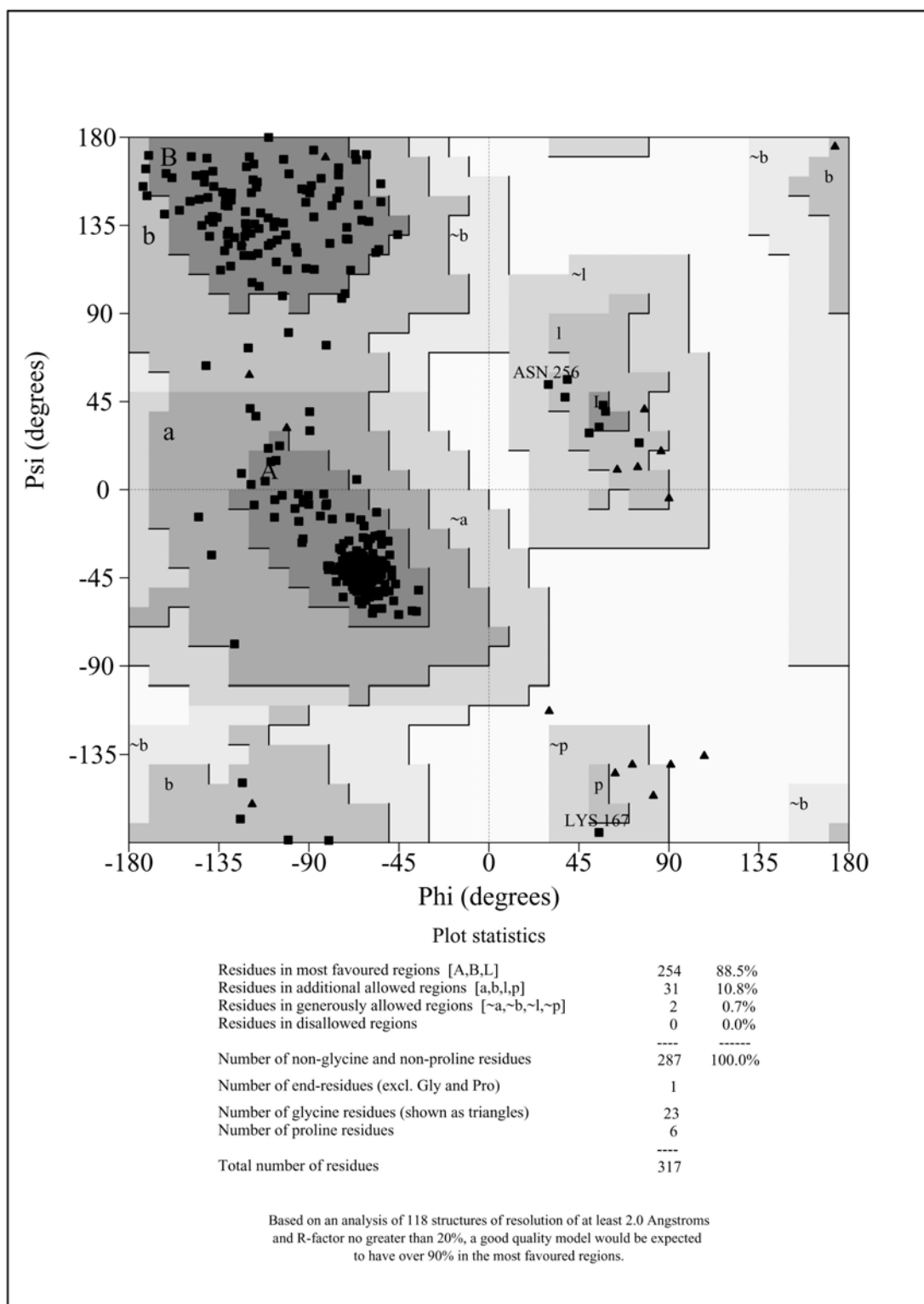


Figure 12 Ramachandran Plot of MVK Crystal Structure.

about 78.2% of the residues are in the most favorable region. The other parameters are also comparable or better than the typical crystal structure at this resolution level.

Fold analysis

The N-domain can be identified as a “ribosomal protein S5 domain 2–like” fold and probably originated from a nucleic acid binding protein. One feature of this domain is a highly conserved phosphate-binding loop, composed of a long loop and the N-terminus of an α -helix (the helix **B**), as the site for the nucleotide binding, which is shared among all GHMP kinases. The other feature of this domain is a large mixed five-stranded β -sheet (formed by strands **a**, **b**, **c**, **d**, **e**) and a smaller anti-parallel four-stranded β -sheet (formed by strands **a**, **b**, **f**, **g**) at one side packed with four long α helices (helices **A**, **B**, **C** and **D**) at the other side. (Figure 13a)

The structure of this N-domain was compared with that domain in other GHMP kinases (Figure 13a, **b**, **c**, **d**, **e**, **f**). Most part of the N-domain seems to be conserved. However, there are substantial differences around the region equivalent to the β strand **d** in MVK. HSK and CDP-ME kinase do not have this strand. Instead, in these two proteins, the element corresponding to the strand **c** of MVK immediately connects to the next helix (equivalent to the helix **A** in MVK) by a short loop (Figure 13b, c). Mevalonate 5-diphosphate decarboxylase lacks the equivalent the strand **d** and its overall structure shows more deviation from the rest of the GHMP kinases (Figure 13f).

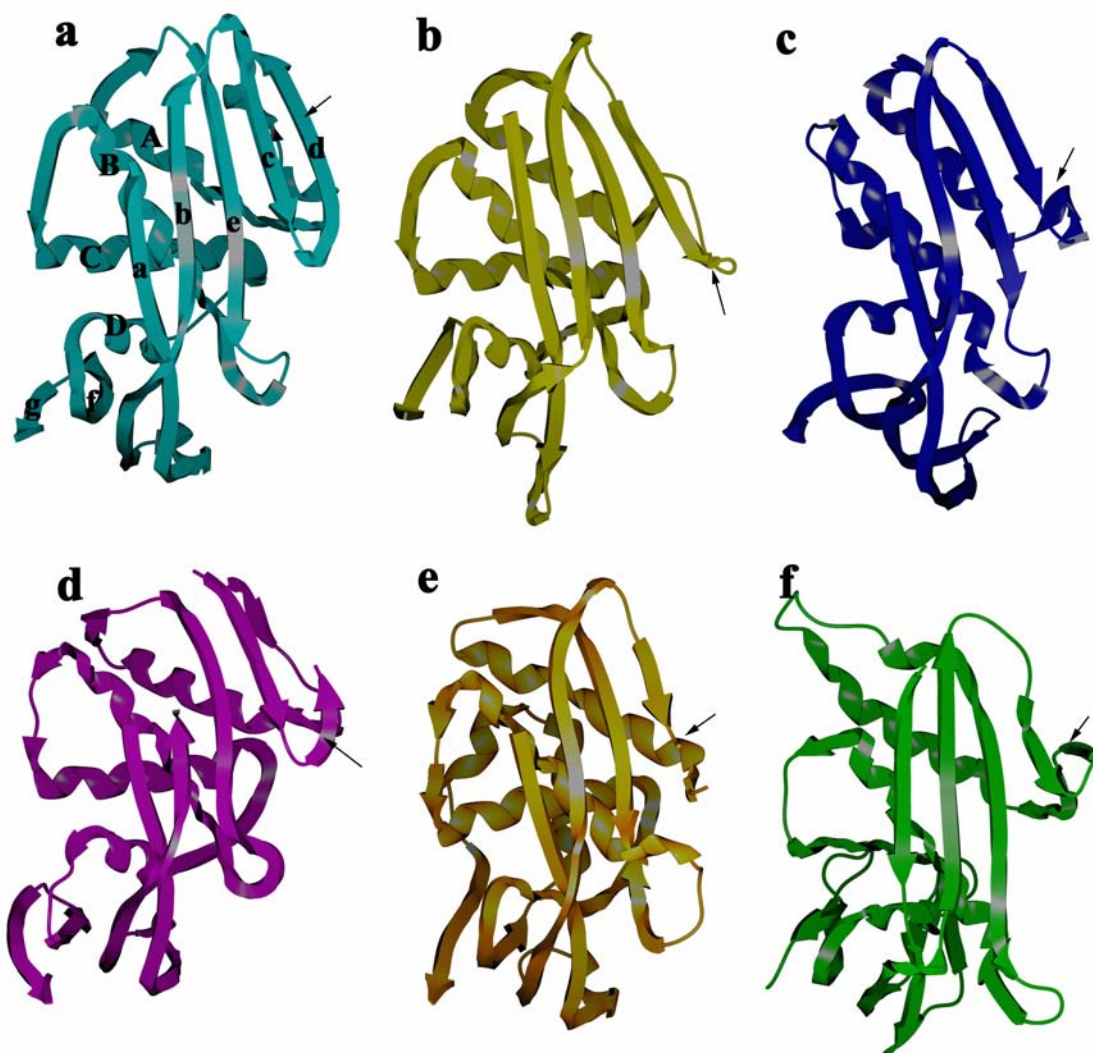


Figure 13 Fold Comparison of N-domains of GHMP Kinases.

N-domains from MVK (a), HSK (b), CDP-ME kinase (c), galactokinase (d), phosphomevalonate kinase (e) and mevalonate 5-diphosphate decarboxylase (f) are shown. The secondary structural elements in MVK are labeled. Regions where major structural deviation takes place are marked by arrows.

The C-domain is composed of a conserved structural core that adopts a $\beta\alpha\beta\beta\alpha\beta$ fold, formed with a four-strand anti-parallel β -sheet (strands **h**, **i**, **j**, **k**) packed with two α helices (helices **I** and **J**). There are three additional helices (helices **F**, **G** and **H**) that are considered to be an insertion between the strand **h** and the helix **I** (Figure 14a). The structure of the insertion is more variable in each member of the GHMP kinases family

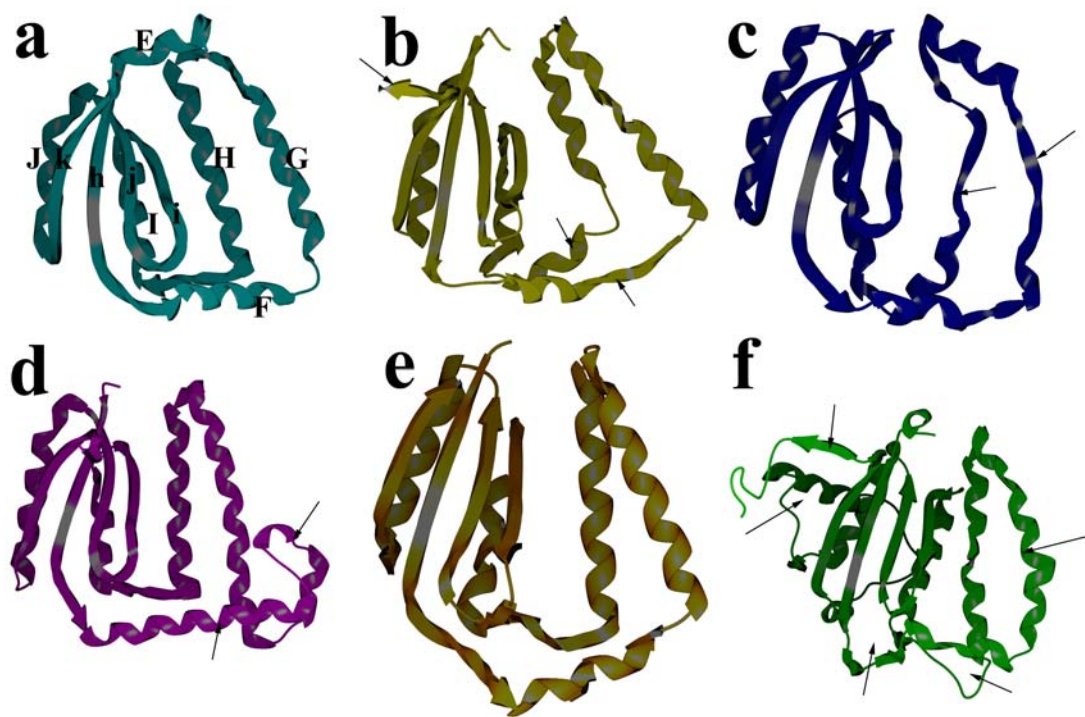


Figure 14 Fold Comparison of C-domains of GHMP Kinases.

C-domains from MVK (a), HSK (b), CDP-ME kinase (c), galactokinase (d), phosphomevalonate kinase (e) and mevalonate 5-diphosphate decarboxylase (f) are shown. The secondary structural elements are labeled in MVK. Regions where major structural deviation takes place (corresponding to the insertion part of the structure) are marked by arrows.

(Figure 14a, b, c, d, e, f). In CDP-ME kinase, instead of two long helices, the region corresponding to MVK helices **G** and **H** is composed of two long β -strands as well as a short α -helix (Figure 14c). Galactokinase has an additional helix and a short loop insertion between two helices equivalent to helices **F** and **G** in MVK. In addition, compared with the helix **F** in MVK, its equivalent in the galactokinase is much longer (Figure 14d). HSK has a short strand in the location corresponding to the MVK helix **F**. It also has an additional helix insertion between the equivalents to MVK helices **H** and **I** (Figure 14b). Mevalonate 5- diphosphate decarboxylase has one more helix after the equivalent to the helix **J** in MVK. Furthermore, its structural element corresponding to the helix **G** in MVK is bent. (Figure 14f)

The comparison between the rat and *M. jannaschii* enzymes

The structure of the rat MVK was solved shortly after the *M. jannaschii* MVK structure. Overall, the structures are similar (Figure 11b). Superimposition of two structures generates an r.m.s.d at about 1.47 Å. Several differences in the structural elements have been identified. First, the long loop connecting the strand **d** and the helix **A** has a different conformation and in the rat MVK, this loop is apparently longer. Second, the helices **A** and **C** are not superimposed well. These two regions are involved in the nucleoside base binding and Mg^{2+} coordination, respectively. As will be discussed in later sections, difference in these regions may affect the mode of ATP binding.

Furthermore, between helices **B** and **C** in the rat MVK, there is a flap in contact with a long C-terminal extension that folds into the N-domain and joins the large β -sheet. Both the flap and the long C-terminal extension are missing in the *M. jannaschii* MVK.

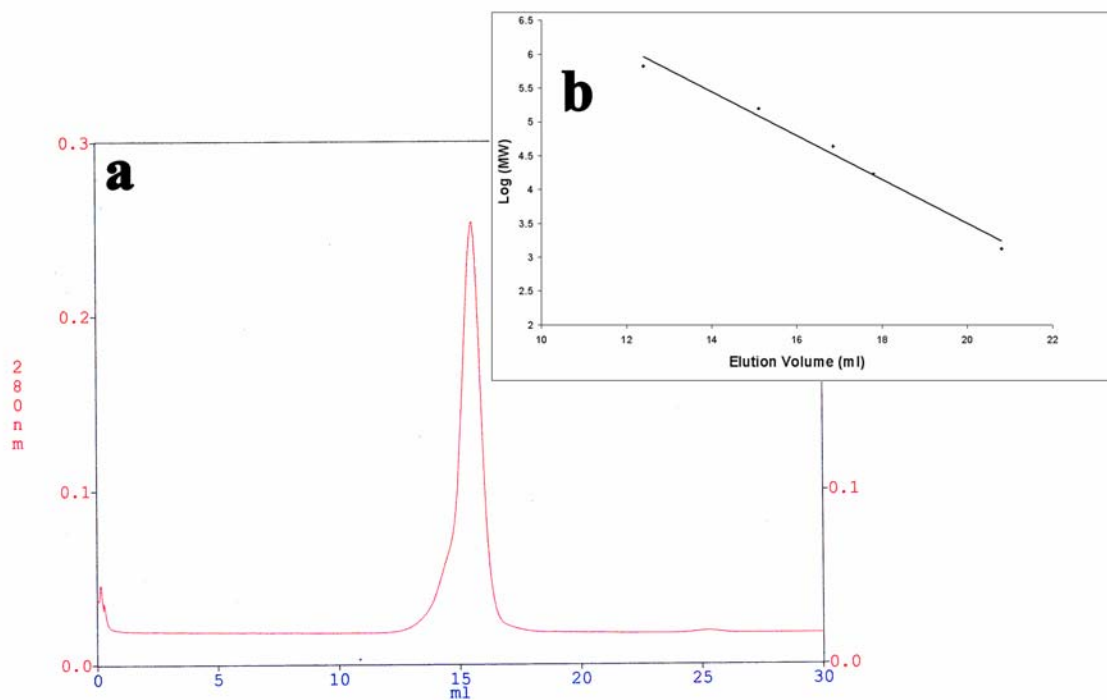


Figure 15 Superose 6 Size-exclusion Chromatography of MVK.

a. The Elution Profile of MVK

b. Calibration of the Superose-6 Column

Thyroglobulin, bovine gamma-globulin, chicken ovalbumin, equine myoglobin and vitamin 12 (from left to right) were used to calibrate the size-exclusion column. The logarithm of the molecular weight (MW) and the elution volume were fit with linear regression. The resulting equation was used to estimate the molecular weight of MVK from its elution volume.

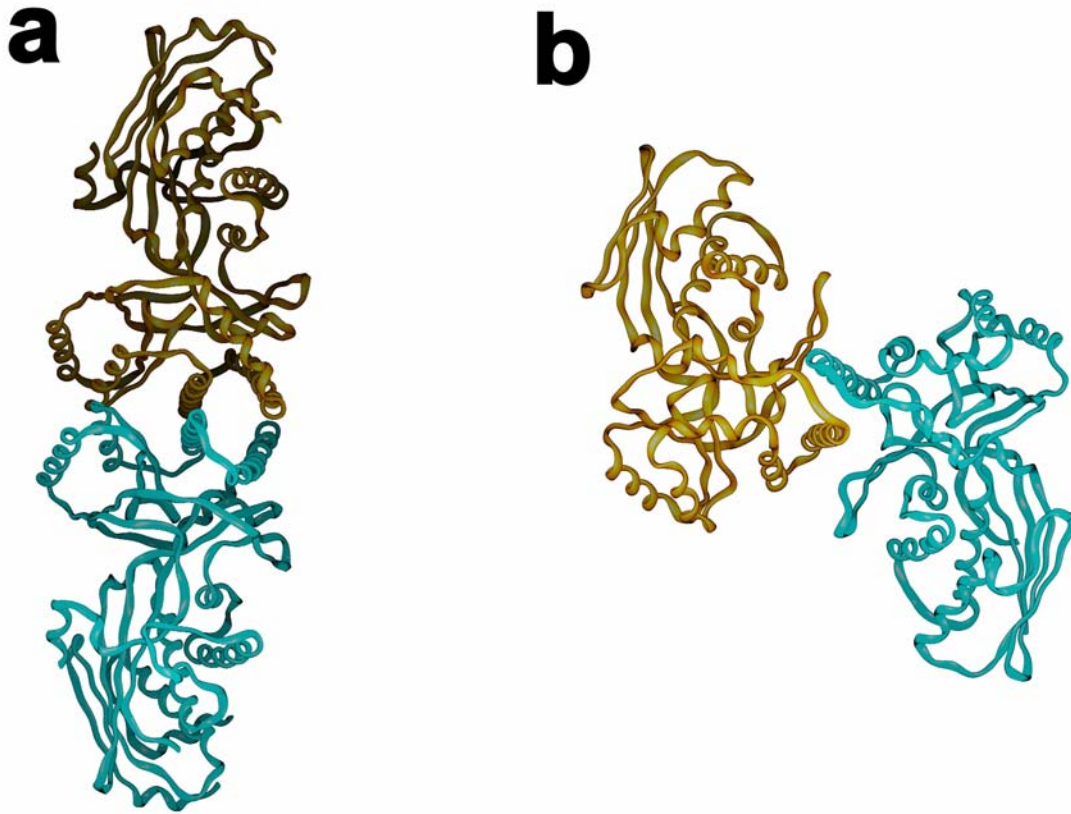


Figure 16 Oligomer State of MVK and HSK Crystal Structures.

a. The Dimer of the Rat MVK

b. The Dimer of the *M. jannaschii* HSK

Interestingly, Val 377 is located in the long C-terminal extension and the human equivalent to this residue is related to HIDS. Therefore, this part of the structure might be involved in protein stability. The *M. jannaschii* protein, as a thermophilic protein, may have a different stability profile from that of the rat protein, which probably explains the difference in this region. Finally, in the C-terminal domain, the helix G and

the following loop show large structural deviation between the rat and *M. jannaschii* proteins, which, as will be discussed in the next section, may have implications in the dimerization of MVK.

Oligomeric state of MVK

The size-exclusion chromatography studies using superose 6 column (analytical grade with separation range between 5 to 5000 KD) suggested that the molecular weight of MVK is 90646 Da (Figure 15). As the molecular weight for a monomer is 37340 Da (based on its amino acid sequence), it is estimated that MVK exists as a dimer in solution. In the crystal structure, the *M. jannaschii* MVK has only one molecule in the asymmetric unit. In order to reveal any possible dimers whose dimeric symmetry axis is coincident with the crystal symmetry, symmetry operation was applied and all molecules in a unit cell were inspected. However, no dimer was found.

Dimeric forms were reported from the crystal structures of both the rat MVK and the *M. jannaschii* HSK. In both proteins, the dimeric interface is at the C-domain. Interestingly, in the rat MVK, helices **H**, **G** and **I** are involved in the protein-protein interaction, which generates an elongated dimer (Figure 16a). In HSK, however, only the equivalent to the helix **G** is involved in the dimer formation. Consequently, the HSK dimer has a different shape and the dimeric interface from those of the rat MVK dimer (Figure 16b), which indicates that the dimer formation is not through a conserved mechanism within the GHMP kinase family. Sequence comparison also shows that the hydrophobic residues in the dimeric interface in the rat MVK are not conserved in the *M. jannaschii* protein. As stated earlier, the helix **G** also shows structural differences in the

rat and *M. jannaschii* MVK. Therefore, the dimeric interactions in the *M. jannaschii* MVK may be different and weaker than that from the rat enzyme, which makes it readily disrupted by crystal packing forces.

Interaction between structural motifs

Cys 281 in the motif III is approximately 2.0 Å from Cys 107 in the motif II. The relative orientation of these two residues, as well as a continuous electron density connecting their side chains, suggest a disulfide bond (Figure 17). In MVKs from other sources, only the residue corresponding to Cys 281 is conserved. Neither of these two

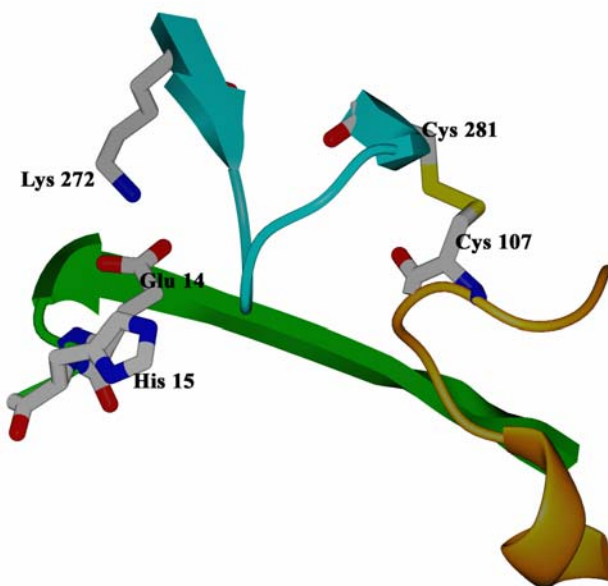


Figure 17 Interactions between Structural Motifs.

Motif I (green), II (golden) and III (cyan) are shown in this figure. The disulfide bond between Cys 107 and Cys 281 are shown. The residues Glu 14, His 15 and Lys 272 are also shown. The residues are colored according to the atom types: carbon (white); oxygen (red); nitrogen (blue) and sulfur (yellow). This coloring scheme is applied to residues and ligands throughout unless described elsewhere.

cysteines is conserved among other GHMP kinases, except the bacteria phosphomevalonate kinase, in which the residue equivalent to Cys 281 is conserved. This residue probably plays some roles in binding the mevalonate moiety of the substrate. The disulfide bond might have its unique function in the *M. jannaschii* MVK, probably by “locking” the relative orientation of motif II and III and contributing to its thermostability.

Lys 272 in motif III is 3.1 Å away from Glu 14, a motif I residue (Figure 17). Their relative orientation indicates the formation of a salt bridge that is strictly conserved among MVKs. The mutation of His 20 in the rat enzyme (equivalent to His 15 in the *M. jannaschii* enzyme) to lysine completely abolished the activity of this protein, while its mutation to leucine or tyrosine only moderately decreased the enzyme activity (105). The destructive consequence of H20K is probably due to its disruptive effects on this salt bridge. This salt bridge also exists in phosphomevalonate kinase and galactose kinase (where the lysine is replaced with an arginine) crystal structures.

The active site

The active site of MVK is located in a cleft between the N- and C-domains. The phosphate-binding loop, composed of a long loop and the N-terminus of the helix **B**, forms a “ridge” dividing the active site cavity into two pockets. The smaller pocket accommodates the purine and sugar moieties of ATP, while the larger pocket has a negatively charged patch around the residue Glu 144 that provides the site for Mg²⁺ ion coordination. There is also a positively charged region extended from the phosphate-binding loop towards the residue His 15 in the larger pocket, probably as the binding site

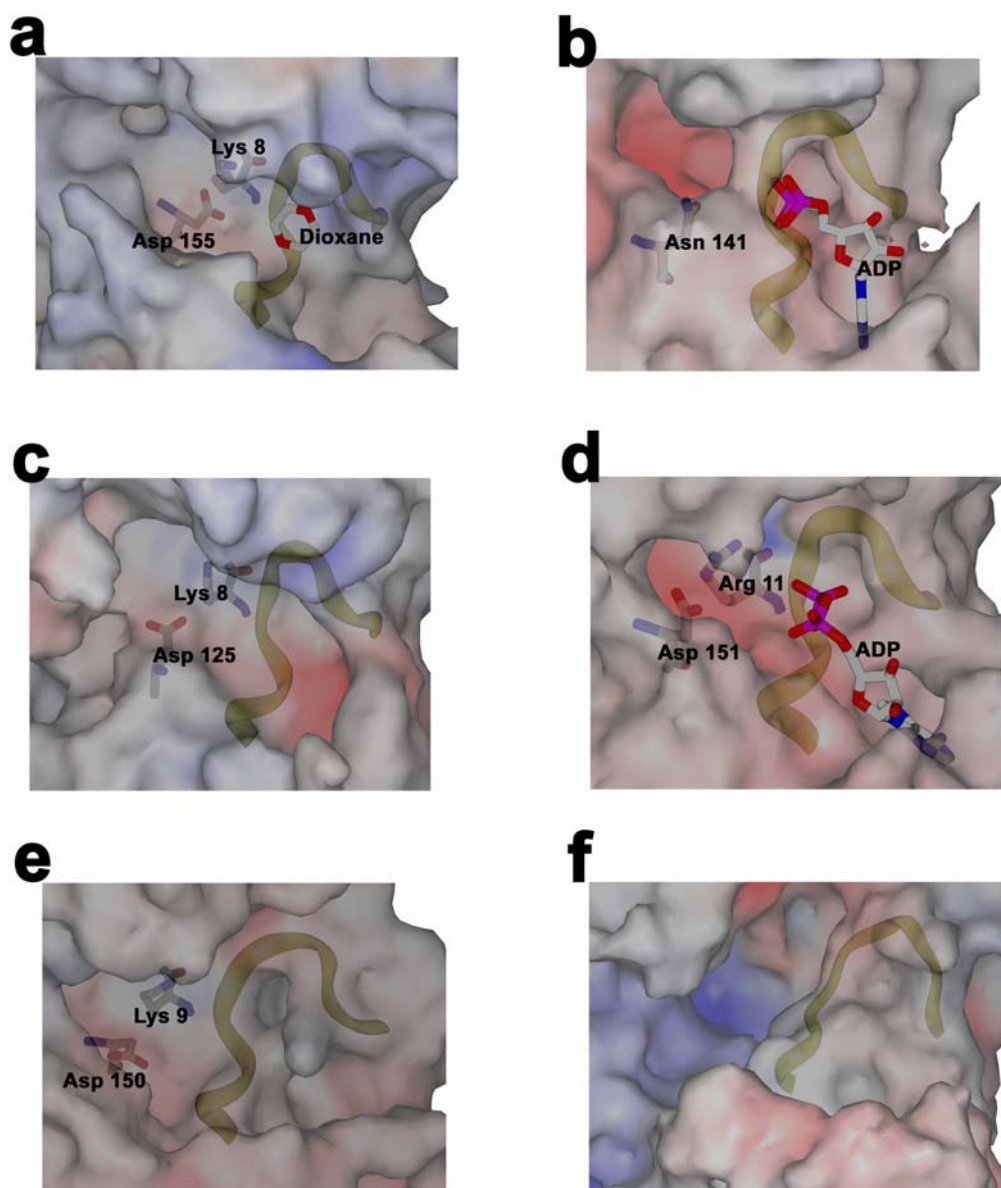


Figure 18 Active Site Pocket of GHMP Kinases.

Active site cavities of MVK (a), HSK (b), CDP-ME kinase (c), galactokinase (d), phosphomevalonate kinase (e) and mevalonate 5-diphosphate decarboxylase (f) are shown. The yellow loop is the phosphate-binding loop. The catalytic salt bridge as well as the ADP and dioxane ligands is displayed. The surface is colored according to the electrostatic potential: negatively charged (red), uncharged (white) and positively charge (blue). This coloring scheme is applied to other surfaces in figures throughout this dissertation unless described elsewhere.

for mevalonate, since its positive charges are compatible with the carboxyl group as well as the 3' and 5'-hydroxyl groups of mevalonate.

In the apoprotein crystal structure, a dioxane molecule is bound about 3.0 Å from the main chain amides of residues Ser 111 and Ser 112, probably forming hydrogen bonds. In the HSK crystal structure in complex with ADP, the β-phosphate group occupies this position and forms hydrogen bonds with main chain amides of Ser 97 and Ser 98 (equivalents to Ser 111 and Ser 112). Therefore, it is likely that the dioxane molecule occupies the phosphate position, which has been verified by the crystal structures in complex with nucleotide ligands, as will be described later in this chapter.

Conserved residues and the three structural motifs are aligned around the active site cavity. Residues from motif II form the phosphate-binding loop that is involved in ATP binding. Although distant in sequence, motif I and III are in proximity in the crystal structure. Together, they form the mevalonate binding site. Asp 155 forms a salt bridge with the residue Lys 8, with distance 2.72 Å. As will be presented later, this salt bridge, in particular the residue Asp 155, plays the central role for the enzyme catalysis.

The active site pockets of MVK, HSK, CDP-ME kinase, galactokinase, phosphomevalonate kinase and mevalonate 5-diphosphate decarboxylase are compared in Figure 18. One common feature is the phosphate-binding loop that has a similar conformation in all six proteins. The salt bridge formed by an aspartate and a lysine residues exists in MVK, CDP-ME kinase, phosphomevalonate kinase and galactokinase (where an arginine is in place of the lysine) crystal structures, suggesting a conserved catalytic mechanism based on a catalytic general base. HSK does not have the lysine

residue and has an asparagine residue in place of the aspartate. HSK is known to lack a catalytic base and use a transition state stabilization mechanism for catalysis, which is consistent with its loss of this salt bridge (106). Mevalonate 5-diphosphate decarboxylase does not have any residues equivalent to this pair and its active site pocket is relatively different from other GHMP kinases (Figure 18), which is consistent with the fact that the reaction it catalyzes is different from a simple phosphoryl transfer.

The nucleotide usage of MVK

Enzyme activity assays showed, in addition to ATP, the *M. jannaschii* MVK can also use GTP, UTP and CTP as the phosphoryl donor. Our assay method based on the PK/LDH coupling system does not permit to compare the nucleotide specificity, because PK prefers ADP as its substrate much more than CDP, GDP and UDP. Therefore, the kinetic property of MVK was characterized by using ATP as the phosphoryl donor. The mean K_m values for ATP and (RS)-mevalonate were determined to be $184.91 \pm 20.94 \mu\text{M}$ and $20.09 \pm 2.68 \mu\text{M}$, respectively. This is comparable with the data from the human enzyme (Table 7). We also investigated the product inhibition to MVK. Because ADP is the substrate of the PK/LDH coupling system, the non-hydrolysable analog ADP β S was used to analyze the product inhibition by ADP and its K_i was determined to be $90.96 \pm 11.90 \mu\text{M}$. AMP did not inhibit MVK at 2 mM concentration, which suggests that a minimal di-phosphate moiety is necessary for nucleotide binding. GDP β S and pyrophosphate were not inhibitory at 2 mM, either, indicating that an adenine base is required for inhibition by the nucleotide. (Table 7)

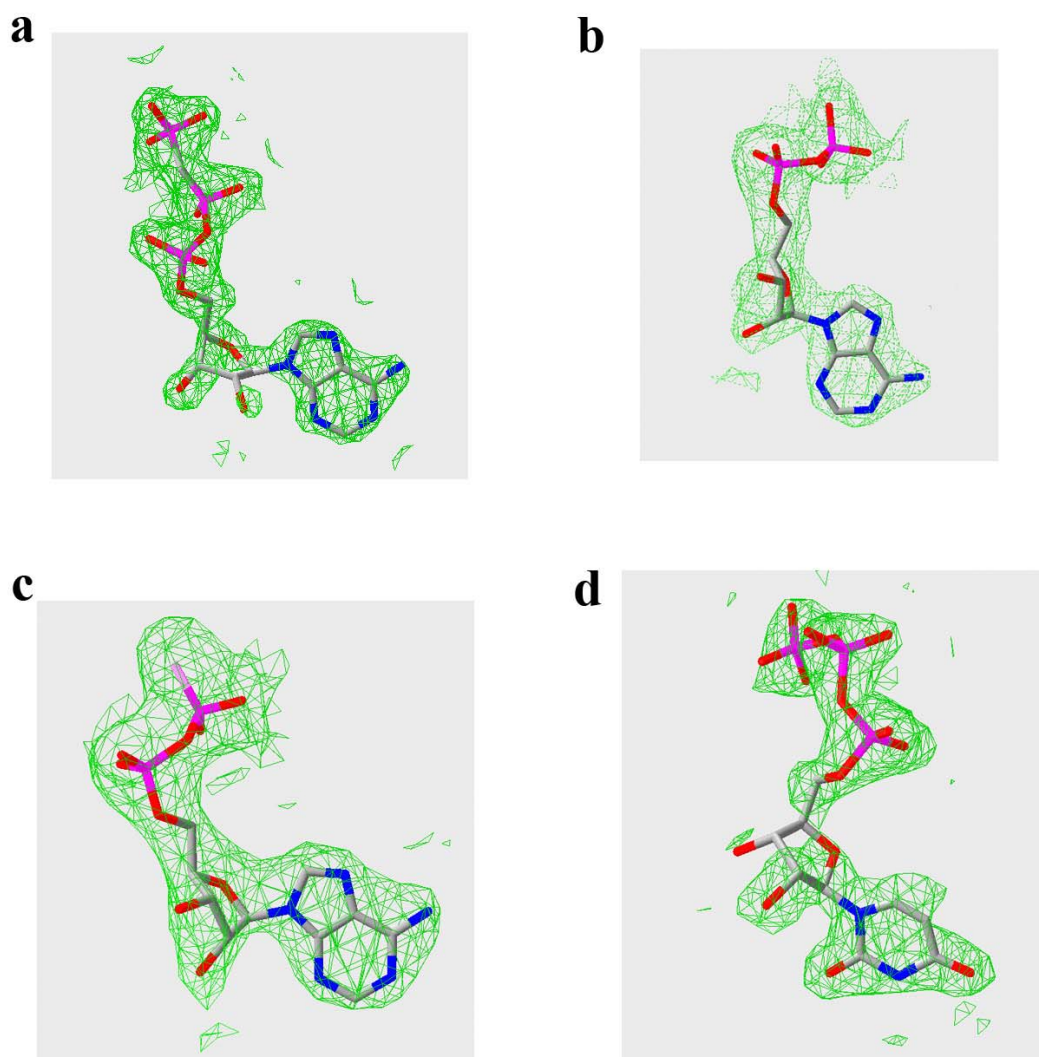


Figure 19 The Electron Density of Nucleotide Ligands.

Electron densities at 1σ in $2F_o - F_c$ maps for nucleotide ligands AMPPCP (a), ADP (b), ADP β S (c) and UTP (d) are shown in green. The phosphorus atoms are colored in purple and the rest atoms are colored as described earlier.

Determining the ligand bound crystal structures

Using the apoprotein structure as the model, the initial $2F_o-F_c$ and F_o-F_c maps showed an additional density that resembled a putative ligand. In general, phosphate groups had the best density. The density for nucleoside base was good in AMPPCP, ADP and ADP β S but weak in UTP. The sugar moiety had poor density in both AMPPCP and UTP, leading to the discontinuity in this region. The model phases were then improved using Shake 'n' wARP, which significantly improved the ligand electron density. As a result, AMPPCP was totally connected and the pyrimidine ring of UTP was clearly revealed. These results further validated the existence of the ligand in the crystal. The ligand was fit into the electron density and after several rounds of refinement, the group occupancy of AMPPCP, ADP, ADP β S and UTP were refined to 0.72, 0.84, 0.83 and 0.80 respectively. The final $2F_o-F_c$ map for nucleotide ligands are shown in Figure 19. The final refinement statistics are summarized in table 3.

For GPP, FPP and GGPP bound MVK crystals, after initial refinement and map calculation in REFMAC, a density contoured at 5σ resembling a pyrophosphate moiety was found near the phosphate-binding loop. Shake 'n' wARP and composite omit map produced some more densities extending from the pyrophosphate. However, the nature of that part of the density is unclear, since it could be either from the disordered carbon chain of terpenoids or a solvent associated with the pyrophosphate moiety. Therefore, only a pyrophosphate has been built.

Table 7 Kinetic Studies on MVK.**NI: No inhibition*****: Taken from Reference 83****** : Taken from Reference 89**

Compound	Constant	<i>M.jannachii</i> MVK	Human MVK
Pyrophosphate	K _i	NI (to 2 mM)	
Geranyl Pyrophosphate	K _i	208.41 μM	116 nM*
Farnesol	K _i	NI (to 2 mM)	72 μM*
Farnesyl Pyrophosphate	K _i	13.02 μM	104 nM*
Farnesyl Monophosphate	K _i	6.54 μM	
Farnesyl Triphosphate	K _i	25.03 μM	
Geranylgeraniol	K _i	NI (to 2 mM)	
Geranylgeranyl Pyrophosphate	K _i	1.02 μM	59 nM*
Geranylgeranyl Monophosphate	K _i	0.66 μM	
Geranylgeranyl Triphosphate	K _i	1.65 μM	
ATP	K _m	184.91±20.94 μM	440 μM* or 74 μM**
(RS)-mevalonate	K _m	20.09±2.68 μM	150 μM* or 24 μM**
ADPβS	K _i	90.96±11.90 μM	
GDPβS	K _i	NI (to 2 mM)	
AMP	K _i	NI (to 2 mM)	

MVK shows little flexibility under different liganded states

The crystal structures of MVK in complex with AMPPCP, ADP, ADPβS and UTP were compared and no major conformation differences were identified (Figure 20a). The only exception is the disulfide bond between residues Cys 107 and Cys 281. In the UTP bound structure, the distance between the two sulfur atoms is about 3.35 Å, indicating that the disulfide bond is broken. In all other structures, this distance is less than 2.20 Å, consistent with the existence of the disulfide bond.

Positions of the phosphate-binding loop in MVK structures under different liganded states are nearly identical, despite the fact that this structural motif is mainly composed of small residues and might be relatively flexible. Interestingly, the main chain amides of Ser 106, Cys 107, Gly 108 and Leu109 form an extensive hydrogen bond network via an ordered water molecule (Figure 20b). This feature also exists in several other GHMP kinase crystal structures, including HSK, phosphomevalonate kinase, the rat MVK and CDP-ME kinase and may contribute to the rigid conformation of the phosphate-binding loop.

The crystal structure of the MVK-AMPPCP-Mg²⁺ reveals the mode of ATP binding and Mg²⁺ coordination

Consistent with the published rat MVK-ATP-Mg²⁺ structure (96), AMPPCP is in an *anti* conformation in the MVK-AMPPCP-Mg²⁺ crystal structure. The phosphate groups are bound to the phosphate-binding loop via several hydrogen bonds with main chain amides. Interaction with the phosphate-binding loop is mainly made by the β - and γ -phosphate groups (Figure 21a).

The Mg²⁺ ion is octahedral coordinated, by three water molecules as well as the side chain of Ser 112 and oxygen atoms from the β - and γ -phosphate groups of AMPPCP (Figure 22a). The three water molecules and the β -phosphate oxygen form a plane with the Mg²⁺ ion in the center, while Ser 112 and the γ -phosphate oxygen coordinate from the above and below, respectively. Water molecules involved in Mg²⁺ coordination form hydrogen bonds with side chains of several conserved residues, including Ser 112, Asp 155 and Glu 144. Interestingly, the conserved residue Glu144

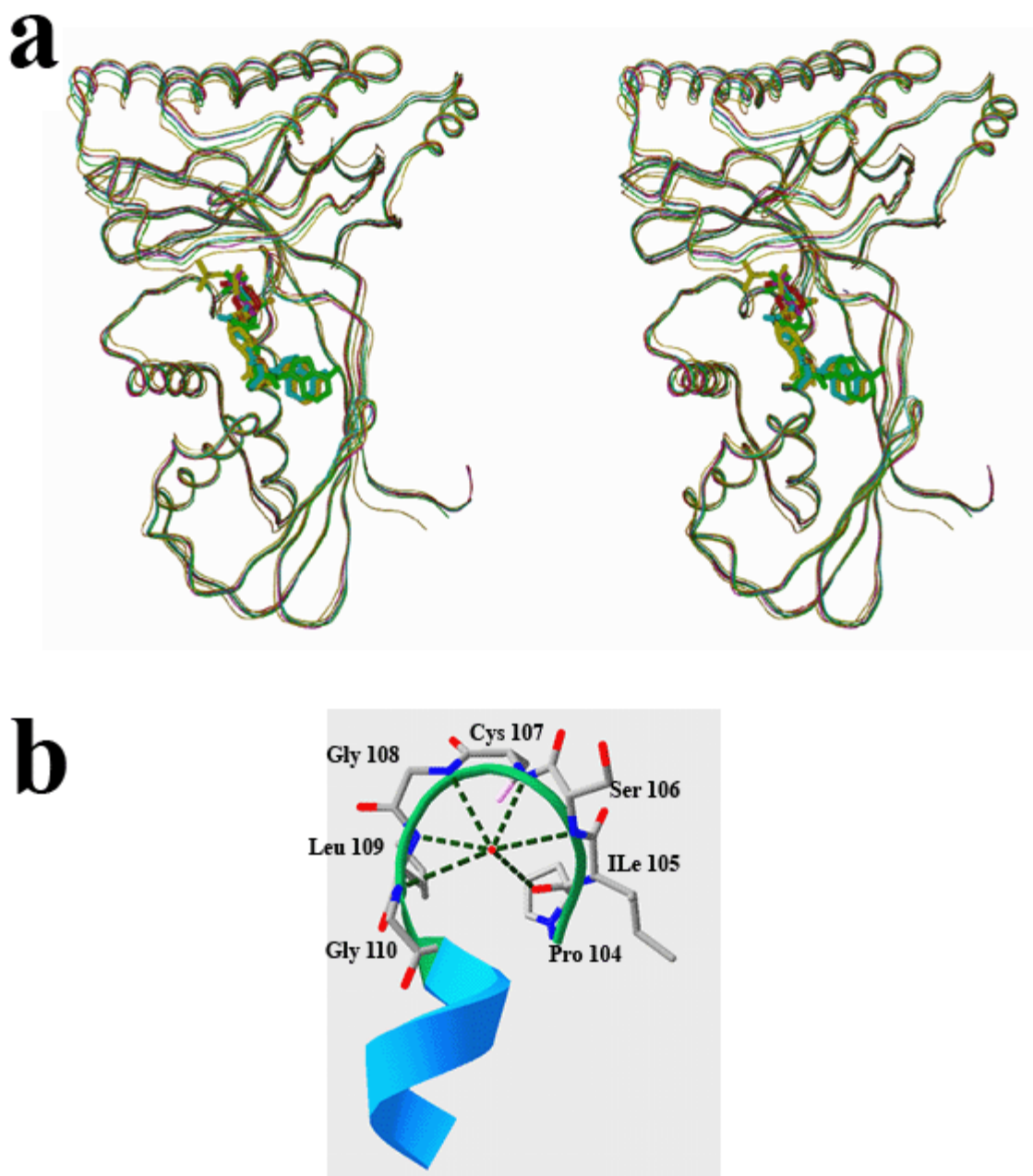


Figure 20 Overall Structures.

a. The Alignment of MVK Crystal Structures in Complex With AMPPCP (green), ADP (cyan), ADP β S (golden) UTP (yellow), FPP(purple) and Dioxane (red)

b. An Conserved Water Molecule Coordinated by the Phosphate-binding Loop

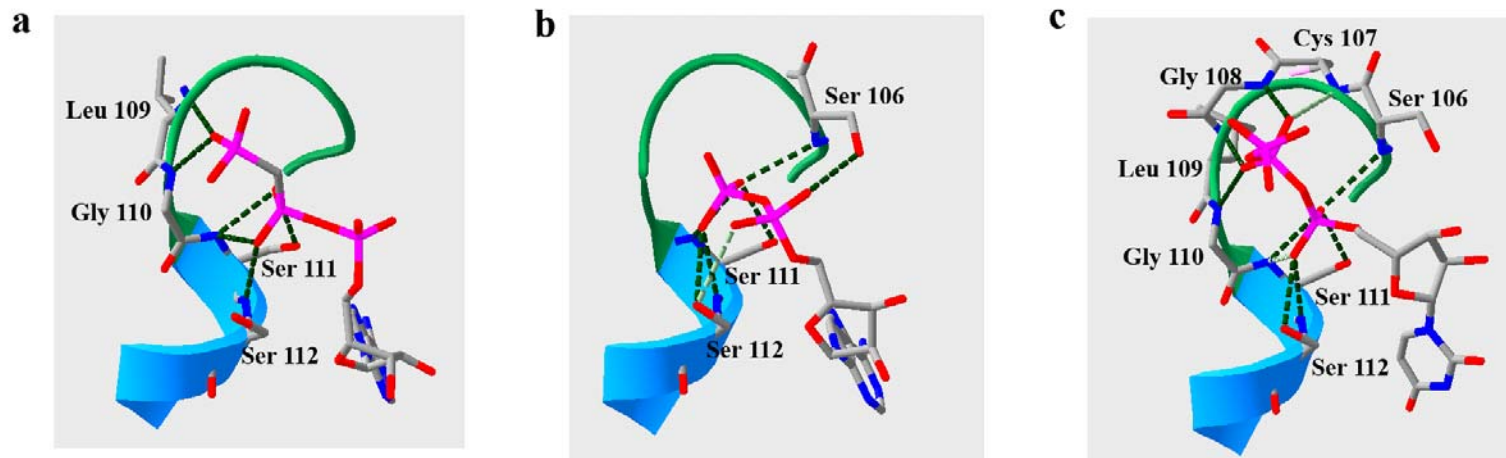


Figure 21 The Binding Mode of the Phosphate Groups.

a. The Interaction of the Phosphate Groups of AMPPCP with the Phosphate-binding Loop

b. The Interaction of the Phosphate Groups of ADP with the Phosphate-binding Loop

c. The Interaction of the Phosphate Groups of UTP with the Phosphate-binding Loop

does not directly coordinate the Mg^{2+} ion, as was reported in the rat MVK crystal structure (96). In the rat MVK, the distance between the equivalent glutamate residue (Glu 193) and the Mg^{2+} ion is closer (2.84 Å in the rat protein versus 3.95 Å in the *M. jannaschii* protein). After superimposing two MVK molecules, it is apparent that, the AMPPCP- Mg^{2+} in the *M. jannaschii* MVK is not aligned well with the ATP- Mg^{2+} in the rat MVK (Figure 22b). As mentioned previously, several structural differences are observed between the rat and *M. jannaschii* MVK, near the nucleoside base binding pocket, which probably affects the binding of the nucleotide ligand and Mg^{2+} coordination.

The two hydroxyl groups on the ribose moiety form hydrogen bonds with the carbonyl of Asp 72 and also form indirect hydrogen bonds (bridged by a water molecule) with the main chain amides of Lys 74, Tyr 75 and Cys 76. The purine ring is involved in both hydrophobic and hydrogen bonding interactions with MVK. The pocket that accommodates the nucleoside base is relatively hydrophobic and the adenosine ring is located in a site surrounded by residues Leu 46, Leu 48, Leu 51, Ile 99, Ile 105 and Ile 115. Hydrogen bonding is formed between N6 of the adenosine ring and the side chain of Asp 50. In addition, the side chain and main chain amide of Ser 101 form indirect hydrogen bonds with N7 and N6 of the adenosine ring, respectively. Main chain amides of Asn 49 and Asp 50 also form indirect hydrogen bonds with the N1 of the purine ring (Figure 23a). The hydrogen bonding pattern may provide the basis for the selectivity on the nucleotide usage.

The potential site for the phosphoryl acceptor

Although mevalonate was included in the crystallization condition, no electron density could be identified as this ligand. Instead, a dioxane molecule was found to partially overlap with the potential mevalonate binding site, which may compete with mevalonate and preclude its binding.

In the MVK-AMPPCP-Mg²⁺ structure, the side chains of Asp 155 and Lys 8 form a hydrogen bond with an ordered water molecule. This same water molecule is also close to the γ -phosphorus of AMPPCP (3.46 Å) (Figure 24a). In addition, after superimposing the HSK-homoserine structure with the MVK-AMPPCP-Mg²⁺ structure, we aligned the 4'-hydroxyl group of homoserine (its phosphoryl acceptor) to almost the same location of this water molecule. Furthermore, this water molecule, the γ -phosphorus and the bridging carbon atom between the β - and γ -phosphorus of AMPPCP are approximately in the “in-line” geometry. All these clues suggest that the location of this ordered water molecule marks the binding site for the phosphoryl acceptor, the 5'-hydroxyl group of mevalonate. Based on the position of this ordered water molecule and the HSK-homoserine structure superimposed to MVK, we modeled a mevalonate molecule into the MVK active site (Figure 24b). As expected, the mevalonate was modeled to the positively charged area in this pocket. Its carboxyl group is facing the side chain of Ala 276, whose equivalent residue in the human enzyme (Ala 334) is indicated in HIDS. Probably, the introduction of a bulkier side chain of the threonine in the A334T mutant from the HIDS patients interferes with the carboxyl group and decreases the binding of mevalonate.

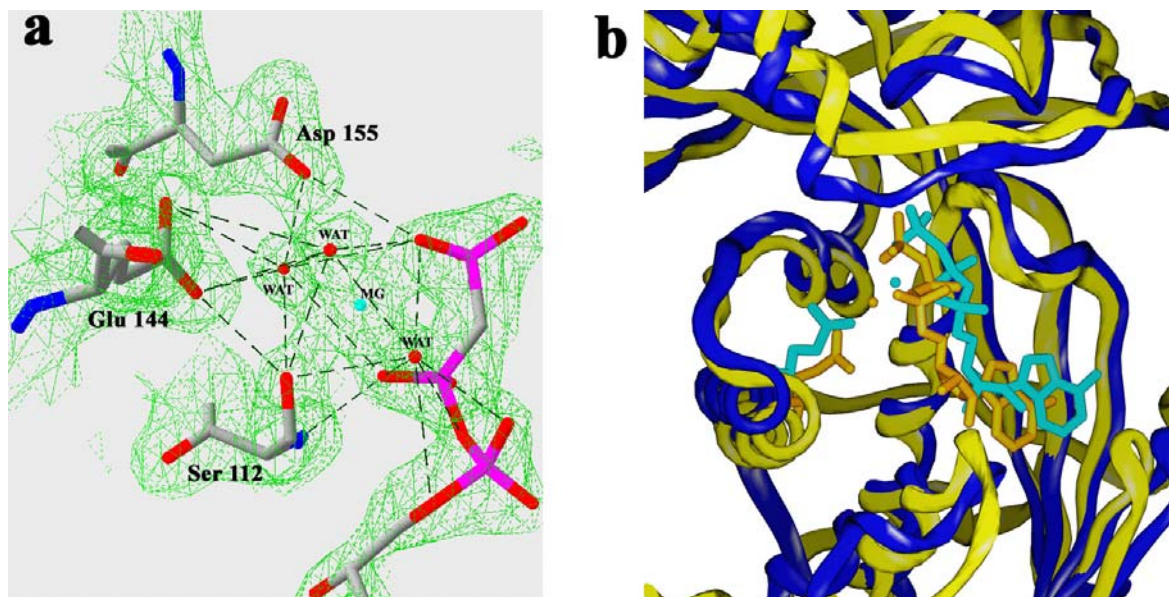


Figure 22 The Mode of Mg^{2+} Coordination and the Comparison of the Rat and *M. jannaschii* Enzymes.

a. The Mode of Mg^{2+} Coordination

The electron density (contoured at 1σ) around the Mg^{2+} ion (cyan) is shown in green. The residues important for Mg^{2+} coordination are displayed. The coordination water molecules are shown as red spheres. The hydrogen bonds are displayed in black dashed line.

b. The Structural Alignment of the Rat and *M. jannaschii* MVK

The *M. jannaschii* (blue) and rat (yellow) MVKs are superimposed. The Mg^{2+} , Glu 144 and AMPPCP from *M. jannaschii* protein is shown in cyan, while the Mg^{2+} , Glu 193 and ATP from rat protein is shown in golden.

Is this ordered water molecule an actual phosphoryl acceptor? Though it was reported that HSK has an ATPase activity in addition to its kinase activity (107), we did not find the ATPase activity from MVK based on our enzyme activity assay. Co-crystallization of MVK with ATP showed only an ADP density in the active site, which might suggest that in the special situation within a crystal, where the protein concentration is very high, MVK could hydrolyze ATP, as has been demonstrated in other ATP binding proteins (101). However, no direct evidence exists for this hypothesis.

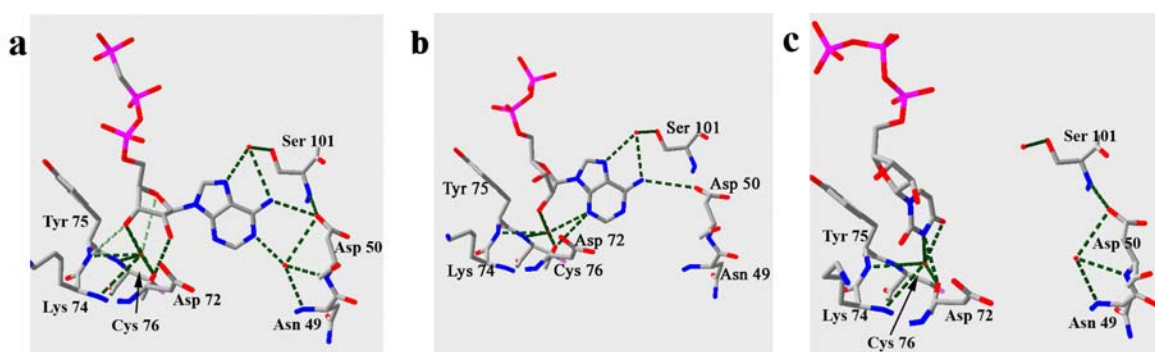


Figure 23 The Binding Mode of the Nucleoside Base.

a. The Binding Mode of the Nucleoside base of AMPPCP

b. The Binding Mode of the Nucleoside base of ADP

c. The Binding Mode of the Nucleoside base of UTP

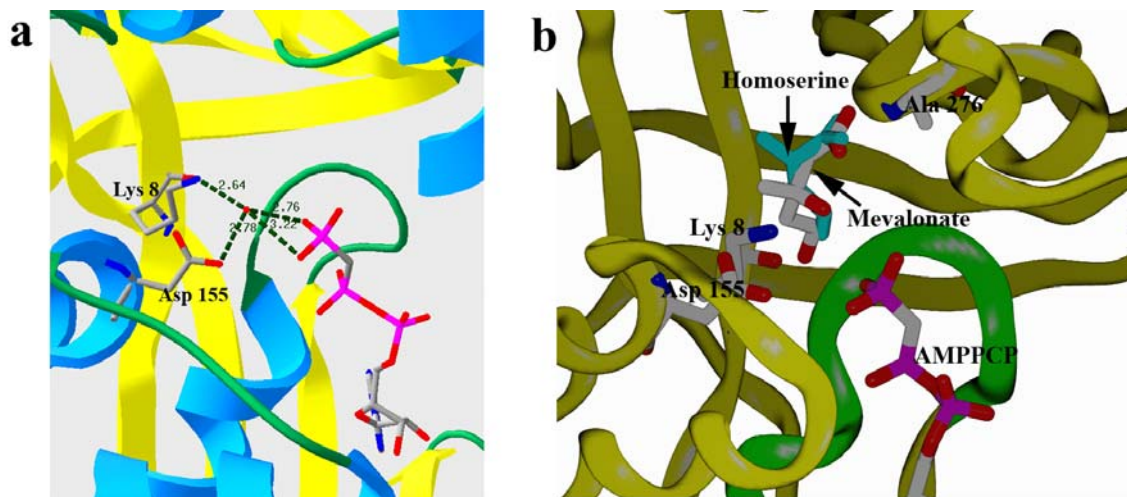


Figure 24 The Phosphoryl Acceptor Binding Site.

a. The Potential Phosphoryl Acceptor Site

The ordered water molecule is shown as a red sphere. AMPPCP and catalytic residues Asp 155 and Lys 8 are displayed. The hydrogen bonds are displayed as green dashed line and the bonding distances are displayed in Å.

b. The Modeling of a Mevalonate Molecule

The phosphate-binding loop is shown in green. Homoserine from the superimposed HSK is in cyan and the modeled mevalonate is colored according to the atom types. AMPPCP as well as key residues Lys 8, Asp 155 and Ala 276 are displayed and colored according to the atom types. The phosphorus of AMPPCP are colored in purple.

The crystal structure of the MVK-ADP-Mg²⁺ and the basis for product inhibition by ADP

The binding mode of ADP was revealed in the crystal structures of the MVK-ADP-Mg²⁺ and MVK-ADPβS-Mg²⁺, solved to 2.4 Å and 2.3 Å, respectively. In both structures, the ligands have almost exactly the same conformation. Therefore, we will focus discussion on the MVK-ADP-Mg²⁺ structure. In ADP, both the α- and β-phosphate

groups are in close contact with the phosphate-binding loop and participate in Mg^{2+} coordination (Figure 21b). As was also revealed in the MVK-AMPPCP- Mg^{2+} structure, a di-phosphate moiety is involved in making hydrogen bond with the phosphate-binding loop. This will again be demonstrated in the UTP and FPP (as from its partial density) bound structures in later sections. AMP certainly does not have this di-phosphate moiety, which probably prevents it from making optimal binding to MVK, explaining its lack of inhibition.

After superimposing the MVK crystal structure in complex with AMPPCP and ADP, we found that there is a nearly 60° of rotation in the α -phosphate group, by which the α -phosphate group in ADP is brought into contact with the phosphate-binding loop and replaces one water molecule at the Mg^{2+} coordination position. As suggested from the crystal structures of HSK (101, 106) and the rat MVK (96), the phosphate groups of the nucleotide are flexible. In the rat MVK (96), the ATP phosphate tail can be fit with two conformations. The major conformation is similar to the AMPPCP structure reported in this dissertation, with the β - and γ -phosphate involved in Mg^{2+} coordination. In the minor conformation, the α -phosphate group also coordinates Mg^{2+} , which is similar to the ADP structure (minus the γ -phosphate). The flexibility of the phosphate groups ensures the phosphate groups of ADP to make sufficient interactions with the phosphate-binding loop and forms a stable complex with MVK, which might cause the product inhibition.

Compared with AMPPCP, the ribose moiety in ADP is twisted for about 45° and, with the purine ring as a rigid body, has a shift for about 1.3 Å. As in AMPPCP, the

purine ring nitrogen atoms N6 and N7 form the same hydrogen bonding interactions with the protein. However, due to the shift of the purine ring, N1 does not form hydrogen bonding interactions with the protein. Instead, N3 of the adenosine ring forms a direct hydrogen bond with the carbonyl of Asp 72 and indirect hydrogen bond with main chain amides of Tyr 75 and Cys 76 (Figure 23b).

The binding mode of an alternative phosphoryl donor

In order to reveal the binding mode of alternative nucleotide ligands, we soaked or co-crystallized GTP, UTP and CTP with MVK. Only UTP showed an unambiguous density, in which the α - and β -phosphate groups are associated with the phosphate-binding loop and form several hydrogen bonds (Figure 21c). Interestingly, the γ -phosphate group of UTP is rotated approximately 180° away from the phosphate-binding loop and does not form hydrogen bonding interactions with its residues. Instead, it forms hydrogen bonds with the side chains of Ser 200 and Arg 196.

The Mg^{2+} is located at approximately the same place as that in the MVK-AMPPCP- Mg^{2+} and MVK-ADP- Mg^{2+} structures, coordinated by the side chain of Ser 112, three water molecules and one α -phosphate oxygen. From the mechanistic point of view, the γ -phosphate oxygen atom should be involved in Mg^{2+} coordination in order for the nucleotide to be activated for the phosphoryl transfer. In the UTP bound structure presented here, the γ -phosphoate group is not involved in Mg^{2+} coordination and it is relatively far away from the catalytic Asp 155-Lys 8 pair. Therefore, it is not likely to be in the proper in-line geometry required for the enzyme activity and our structure may represent a non-productive binding mode that is energetically most stable. It is certainly

possible that conformational changes after mevalonate binding might lead to a proper UTP conformation, considering the flexibility of the phosphate groups.

There is no interaction between the ribose moiety and the protein, which is consistent with the weaker electron density in this region (Figure 19d). Since UTP has a pyrimidine ring that has different hydrogen bonding and geometry properties from the purine ring of AMPPCP, it is not surprising to find that the nucleoside base of UTP adopts a different binding mode. In the crystal structure, the pyrimidine ring of UTP is bound to a site close to the aromatic ring of Tyr 75 and the side chain of Ile 115, approximately the ribose moiety binding site of AMPPCP and ADP. N3 of the uridine moiety forms one hydrogen bond with the main chain carbonyl of residue Asp 72. Both N3 and the carbonyl oxygen O4 of the uridine ring also form indirect hydrogen bonding interactions with main chain amides of Lys 74 and Tyr 75 via a bridging water molecule (Figure 23c). Interestingly, most of these hydrogen bonding sites are used by the purine atom N3 in ADP.

Compared with the adenosine moiety in AMPPCP and ADP, fewer hydrogen bonds are formed between the uridine moiety and MVK. Normally, both uridine carbonyl oxygens O2 and O4, as well as N3, make hydrogen bonding interactions with protein residues, as was demonstrated in the dUTPase crystal structures (108). Therefore, the binding of UTP in MVK is not optimized. The hydrogen bonding pattern between the nucleoside base and the protein suggests the nucleoside base binding pocket is probably selective for the adenine ring, as adenine forms more hydrogen bonds with the protein. However, even for adenine, its binding modes are slightly different in the

AMPPCP and ADP bound crystal structures. The uridine, on the other hand, binds quite differently. These data suggest that the nucleoside base binding pocket has a degenerate nature, probably based on its relatively large size and the abundance of potential hydrogen bonding partners which allows the base ring to be fit in different orientations. This probably explains why alternative nucleotides can be used as phosphoryl donors.

The possible mechanism for MVK

Based on the available data, a probable mechanism for MVK is proposed (Figure 25). First, the substrate binding is mediated by the three conserved motifs. Motif I and III are involved in mevalonate binding. Motif II provides the interactions sites for a diphosphate group and is critical for the nucleotide binding. The nucleoside base binding pocket accommodates the adenine ring of ATP.

Second, the catalysis is performed by the conserved aspartate residue (Asp 155) that functions as a general base by abstracting a proton from the nucleophilic 5'-hydroxyl group of mevalonate. A nearby lysine residue (Lys 8) lowers the pK_a of that hydroxyl group and may also neutralize the negative charges accumulated on the γ -phosphorus during the transition state. The transition state is further stabilized by the positively charged N-terminus of the helix **B** pointing towards the β - and γ -phosphate groups. The reaction is promoted by an Mg^{2+} ion coordinated by Ser 112 and Glu 144 (via bridging water molecules) (Figure 25).

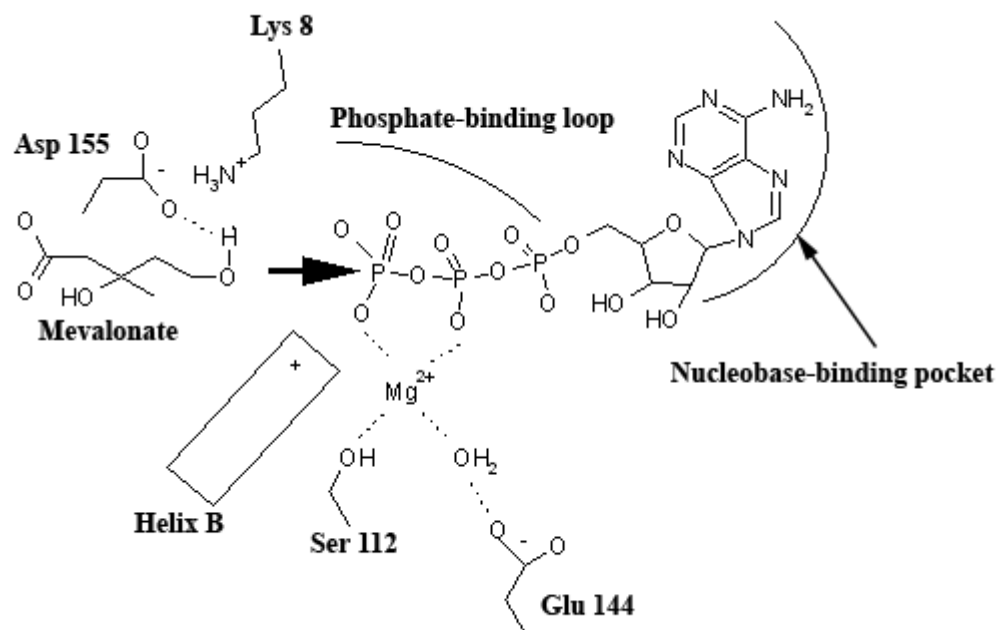


Figure 25 The Proposed Mechanism for the *M. jannaschii* MVK.

Analysis of inhibition by terpenoids and their analogs

We then investigated the effect on the terpenoid inhibition by the carbon chain and phosphate groups. We found that the length of the carbon chain is a positive determinant for inhibition, with K_i of GGPP < FPP < GPP, which is consistent with studies in human enzymes (Table 7) (83). The effect of the phosphate groups is less clear. Interestingly, the monophosphate group was sufficient for inhibition, which is different from the nucleotide as mentioned in section 2. 10. Farnesol and geranylgeraniol, however, did not show any inhibitory effects at a concentration of 2 mM. Therefore, at least one phosphate group is required for the terpenoid inhibition (Table 7).

Why is the di-phosphate moiety of terpenoids not required for inhibition? It is probable that as hydrophilic ligands, nucleotides bind to MVK via electrostatic interactions, which is mainly through the interaction with the phosphate-binding loop. As a monophosphate group is not sufficient to maintain this interaction, AMP cannot inhibit MVK. As hydrophobic ligands, terpenoids might largely rely on the contribution by its carbon chain to bind the protein, as the effect of the carbon chain length has already demonstrated. Although the interaction with the phosphate-binding loop is still necessary, its contribution to the overall binding is not as prominent as in nucleotides. Therefore, the weakening of the interaction with the phosphate-binding loop by losing one phosphate group might be readily compensated by the hydrophobic interactions.

The K_i values for GPP, FPP and GGPP of the *M. jannaschii* enzyme are much higher than those of the human enzyme (Table 7). However, the K_i for FPP in our enzyme is similar to that of the *S. aureus* MVK (~45 μ M) (79). As the phylogenetic analysis suggested that the archaeal MVK is more closely related to the prokaryotic enzyme (52), our observation probably revealed the difference in the response to the terpenoid feed-back inhibition by the eukaryotic and prokaryotic/archaea enzymes. In section 2.6, we compared the crystal structure of the *M. jannaschii* MVK with rat MVK. Differences in the ATP site were identified, which might be the underlining cause for the difference in the terpenoid sensitivity.

Possible mode for terpenoid inhibition

Crystallization trials with terpenoids (GPP, FPP and GGPP) and MVK reveal only a pyrophosphate moiety that occupies the same location of the β - and γ -phosphate groups



Figure 26 The Terpenoids Binding Site.

The pyrophosphate moiety (purple: phosphorus; red: oxygen) of a FPP molecule is in complex with MVK. The grey molecule is an AMPPCP superimposed onto this structure.

of AMPPCP. This is consistent with the competitive nature of terpenoid inhibitors with ATP (81, 83) (Figure 26). Earlier results also show that terpenoids are uncompetitive with mevalonate (81, 83), which may indicate that the binding of mevalonate is required

for the proper binding of terpenoids. As mevalonate does not bind in our crystal structures, this probably explains why the carbon chain is disordered in these structures.

But where could the carbon chain go? Docking experiment suggest that the carbon chain can be placed into the nucleoside base binding pocket. This pocket is relatively hydrophobic and may readily accommodate the hydrophobic carbon chain. We certainly cannot rule out other possibilities, as no experimental evidence is available. The other question is about the mechanism on how mevalonate affects the terpenoid binding? In HSK, the binding of the substrate homoserine leads to the closure of the active site pocket (106). Maybe a similar event is triggered by the presence of mevalonate in MVK, which could form a more hydrophobic binding environment that is preferred by the carbon chain of terpenoids.

Overview of virtual screening

To further explore the active site of MVK, virtual screening was performed. More than 40,000 NCI compounds and 10,000 Maybridge compounds were docked into the MVK active site with FLEXX. In addition, about 78,000 compounds from the LUDI/ACD library were docked using the program LUDI (Table 8). After visual inspection, 29 compounds were tested, and the results suggested that two compounds showed weak inhibition of MVK. (Table 9) These two compounds were further characterized, which included structure activity relationship study based on one of the compounds. Further virtual screening was performed using the “anchoring” moieties identified from the two hits; however, there was no significant improvement in the inhibitory effect.

Table 8 Virtual Screening Statistics.

Number of Molecules	NCI	Maybridge	LUDI/ACD
Initial	40,000	12,000	78,000
Extracted (Cscore, FLEXX)	400	120	N/A
Visual Inspection	400	120	300
Ordered for Testing	9	9	11
Inhibiting MVK	1	1	0
Inhibiting PK/LDH system	2	5	1

Visual inspection

Visual inspection suggested that the top docking hits tend to cluster around the phosphate-binding loop. A majority of the compounds have an aromatic ring structure as the “anchoring” group which was docked near the phosphate-binding loop. Hydrogen bonding with main chain amides was predicted (Figure 27a, b), which partially mimics the binding of phosphate groups to this structural motif. The other part of the compounds extends away from the phosphate-binding loop, mostly into the potential mevalonate binding site. The apparent central role of the phosphate-binding loop for docking is probably due to its main chain amides that provide most of the interaction sites and ensure that the electrostatic interactions play a large role in the total binding energy. However, this pattern can potentially create two problems: i) the docking and scoring may be less accurate because docking programs are insufficient to estimate electrostatic interactions (42) and ii) the compounds may show nonspecific inhibition since they have the same binding pattern as the phosphate (see below). In addition, the phosphate-binding loop may cause low inhibition because of the compounds’ inability to compete

Table 9 FLEXX Hits Chosen for Initial Validation.

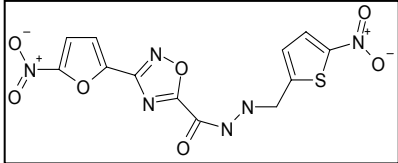
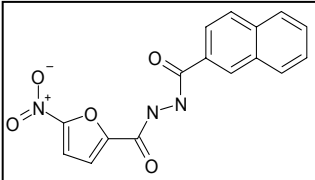
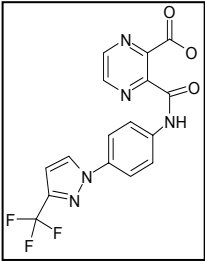
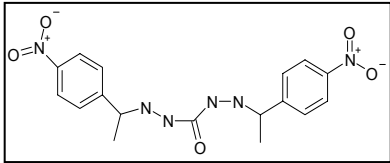
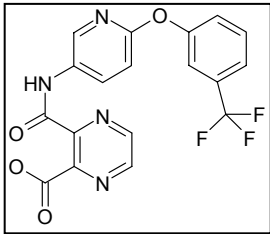
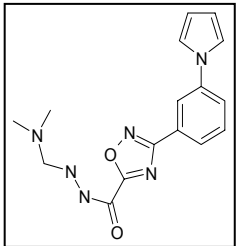
Compound	Structure	F score	CScore	MW (Da)	LogP
SPB00929		-38.74	5	378.28	1.72
RH01369		-39.07	5	325.28	2.26
HTS04725		-38.35	4	377.28	2.45
ML00294		-38.24	5	384.35	2.96
HTS08391		-36.42	4	404.30	3.31
SPB01501		-39.44	3	324.34	3.33

Table 9 Continued.

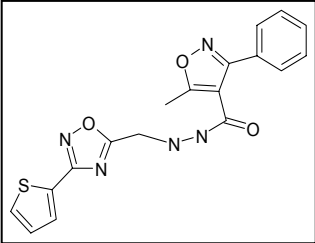
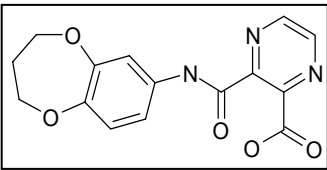
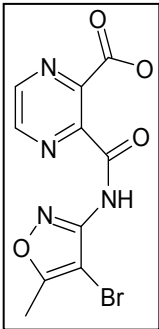
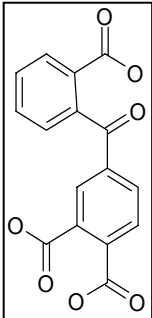
Compound	Structure	F score	CScore	MW (Da)	LogP
HTS08322		-38.75	4	395.40	2.57
HTS03748		-40.98	4	315.28	0.98
HTS07725		-43.99	5	327.09	0.69
NSC28620		-39.40	4	311.23	0.98

Table 9 Continued.

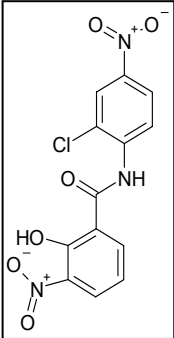
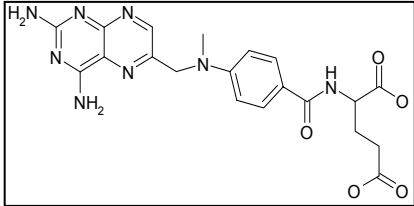
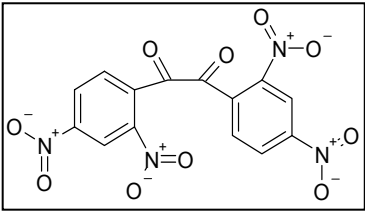
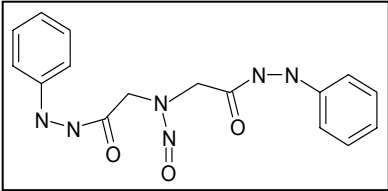
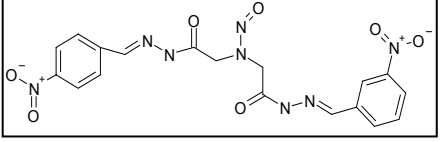
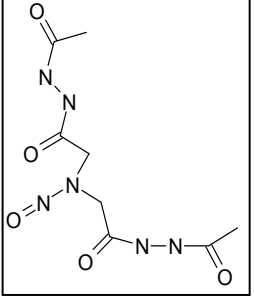
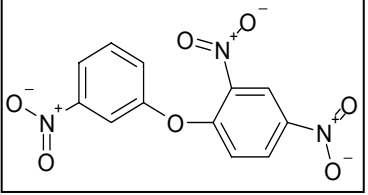
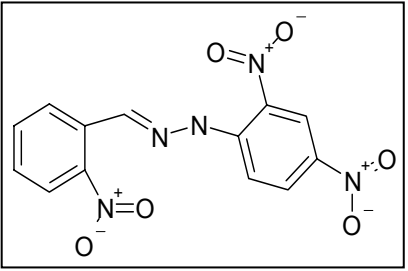
Compound	Structure	F score	CScore	MW (Da)	LogP
NSC82020		-34.20	N/A	464.80	N/A
NSC740		-40.80	N/A	454.44	-0.98
NSC1060		-38.10	N/A	604.60	2.83
NSC21356		-39.80	N/A	384.40	0.37

Table 9 Continued.

Compound	Structure	F score	CScore	MW (Da)	LogP
NSC21357		-41.20	N/A	676.40	0.29
NSC20799		-36.20	N/A	345.80	-4.27
NSC6883		-34.60	N/A	428.00	2.66
NSC12049		-38.30	N/A	486.60	2.65

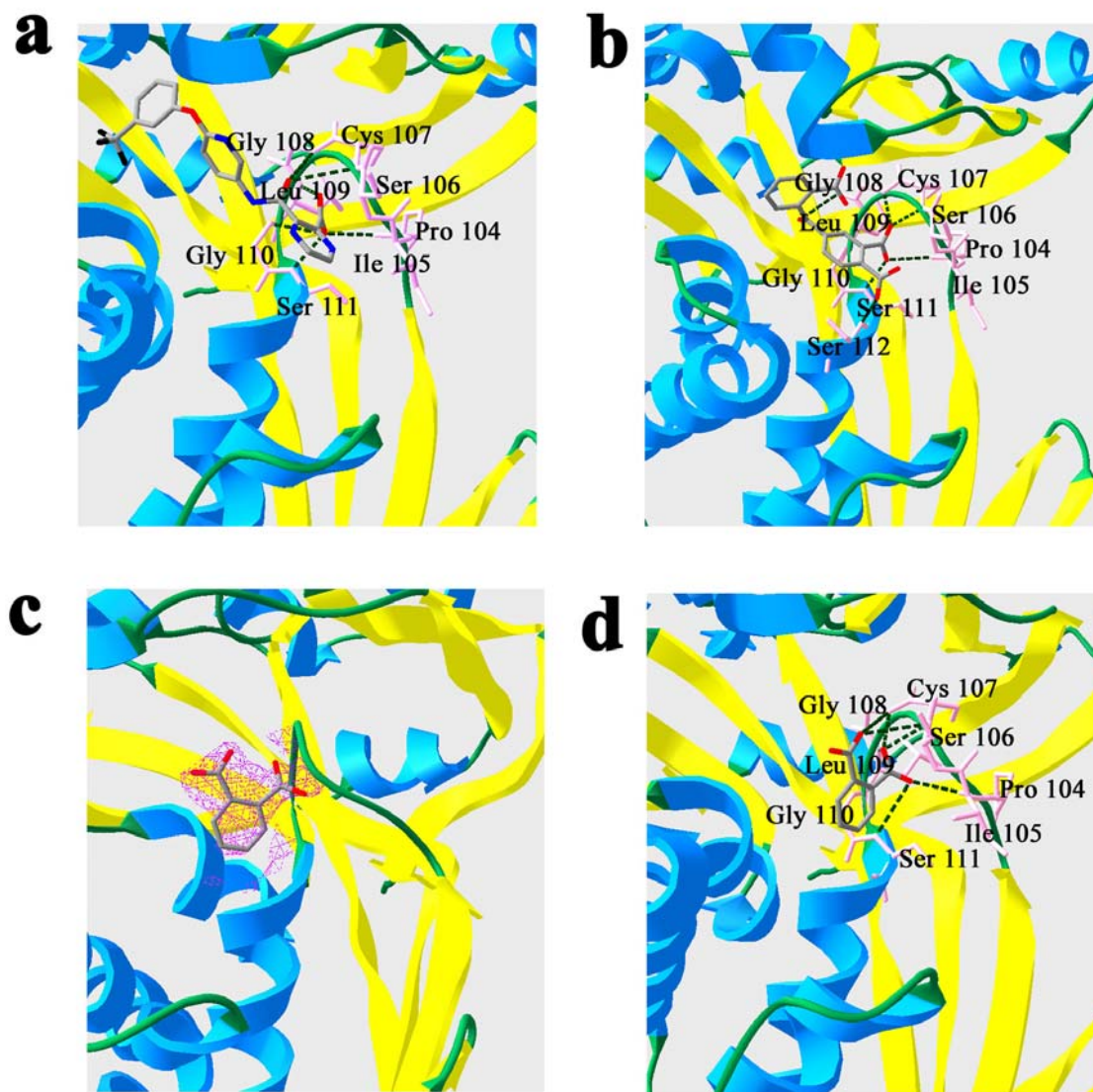


Figure 27 Two Virtual Screening Hits.

a. The Predicted Binding Mode of HTS 08391

b. The Predicted Binding Mode of NSC 28620

c. The Partial Density for NSC28620 (contoured in pink at 1σ)

d. The Binding Mode of the Phthalic Acid Determined from the Crystal Structure

with the phosphate groups from ATP that are optimized for binding to this loop through years of evolution.

To overcome this problem, we inspected more hits from the docking results and identified some that only interact with the nucleoside base binding pocket. Most of these compounds have been predicted to form hydrogen bonds with the side chains of Asp 50 and Asp 72, as well as main chain amides or carbonyls from residues Ile 99, Ser 101, Asn 49 to Asp 50, Asp 72 and Lys 74 to Cys 76, which are similar to the interactions made by the purine or pyrimidine rings from the nucleotide ligands. However, these compounds have poor docking scores (Table 10), predicting that they are not good ligands.

Initial inhibitor validation

In all, 10 Maybridge compounds and 10 NCI compounds from FLEXX docking and 11 ACD compounds from LUDI docking were selected for validation (Table 8, 9). Selected compounds were evaluated, and those with more than 10% inhibition at 10 μ M concentration were subjected for further tests. Many compounds (ML00294, RH01369, SPB00929, HTS08322, HTS08391, NSC82020, NSC740 and 1-hydroxy-2-naphthoic acid) showed 10 to 30% inhibition against coupling enzymes (PK or LDH). This is probably due to the fact that most hits were selected for their affinity to the phosphate-binding loop by FLEXX and might mimic the properties of the phosphate. As both PK and LDH have a ligand that contains phosphate moieties (ADP and NADH, respectively), these compounds might also bind and inhibit them. As mentioned before, compounds that only bind to the nucleoside base binding pocket were visualized, and they are

Table 10 Virtual Screening Hits Exploring the Nucleoside Base Binding Pocket.

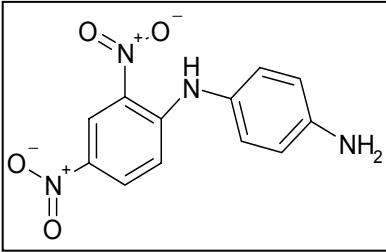
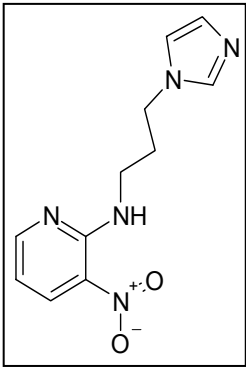
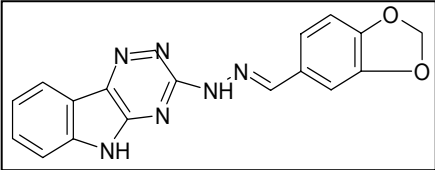
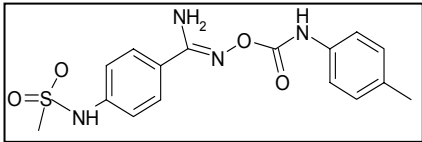
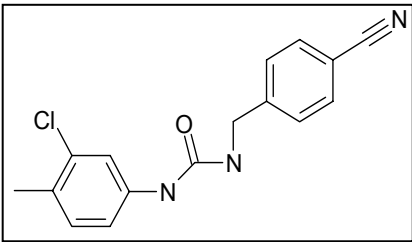
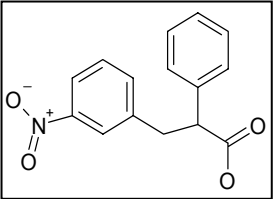
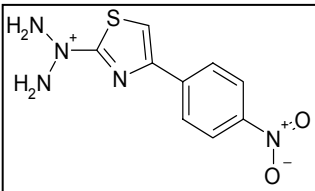
Compound	Structure	F score	CScore	MW (Da)	LogP
JFD03224		-27.05	5	274.24	2.28
CD07440		-24.85	5	247.26	1.62
HTS12557		-25.92	5	332.32	3.05
SPB06772		-27.88	5	362.40	2.38

Table 10 Continued.

Compound	Structure	F score	CScore	MW (Da)	LogP
HTS10933		-24.40	5	299.76	3.20
XBX00106		-26.52	5	268.25	3.05
CD01617		-28.47	5	264.28	1.13

associated with poor docking scores. Therefore, not surprisingly, *in vitro* tests on some of these compounds did not identify any inhibitory effects.

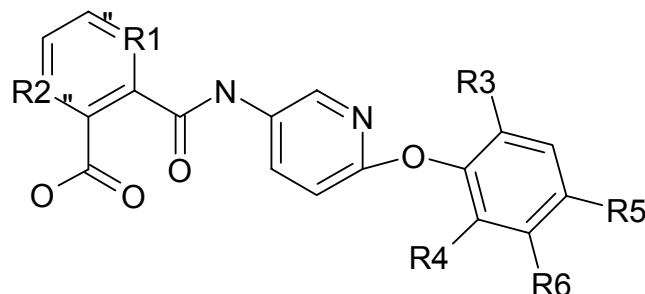
Two compounds NSC28620 (4-(2-carboxybenzoyl)phthalic acid) and HTS08391 (3-[(6-[3-(trifluoromethyl)phenoxy]-3-pyridinyl)amino]carbonyl]-2-pyrazinecarboxylic acid) were identified from initial assays. They showed weak inhibition (about 15%) at 10 μM and did not inhibit the coupling enzymes at this concentration. Visual inspection showed these two compounds were predicted to be anchored to the phosphate-binding loop, via a phthalic acid (NSC28620) and pyrazinecarboxylic acid (HTS08391) moiety (Figure 27a, b).

Determine IC₅₀ and mode of inhibition

Because of its limited solubility and its inhibition on the PK/LDH system when the concentration is higher than 100 μM , compound HTS08391 could not be further tested to determine its IC₅₀. However, compound NSC28620 could be characterized, and its IC₅₀ was determined to be 255.9 μM . To determine the mode of its inhibition, reactions under 20 μM to 3 mM ATP and 0, 50 and 200 μM inhibitor were performed. The initial velocity was plotted with the ATP concentration, and the Lineweaver-Burke plot analysis indicated that compound NSC28620 competes with ATP. Therefore, its K_i value was directly estimated from the IC₅₀ as 69.2 μM .

Crystallographic studies

Compound NSC28620 was then co-crystallized with MVK and a complete dataset was collected from this crystal. No intact drug density could be identified after the electron density map was calculated, even if 2 mM of compound was included

Table 11 Structural Activity Studies Based on HTS08391.

Compound	R1	R2	R3	R4	R5	R6	F score	Inhibition (10 μ M)
HTS08391	N	N	H	H	H	CF ₃	-36.42	14%
HTS11134	N	N	Cl	Cl	H	H	-27.28	No inhibition
HTS08512	C	C	H	H	C(CH ₃) ₃	H	-36.38	No inhibition
HTS08390	C	C	H	H	H	CF ₃	-36.11	No inhibition
HTS03728	N	N	H	H	C(CH ₃) ₃	H	-39.55	14%

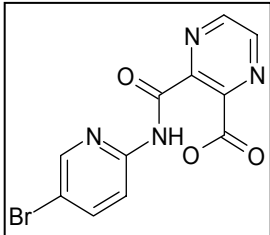
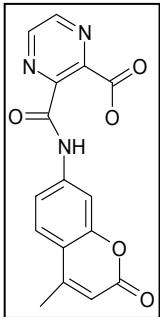
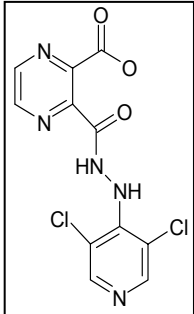
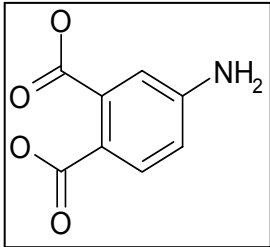
during the crystallization. However, a density near the phosphate-binding loop was identified and was verified by running the program Shake 'n' wARP. It was fit with a phthalic acid moiety (Figure 27c). Prediction from the docking study suggested that the carboxyl groups of the phthalic moiety make hydrogen bonds with main chains amides of Pro 104, Cys 107, Gly 108 and Ser 111 (Figure 27b). The same amides also form hydrogen bonds with the carboxyl groups of the phthalic acid moiety in the crystal

structure (Figure 27d). Interestingly, the orientation of the phthalic acid moiety in the crystal is different from the prediction, with the ring almost completely “flipped” 90°. Since we do not have the complete density, the exact binding mode of this compound remains unclear.

Structural activity studies

Structural activity studies were performed based on compound HTS08391 (Table 11), because derivatives of this compound were commercially available. The only analog that still retained the inhibition is HTS03728 (about 14% at 10 μ M). It has modifications on the distal benzene ring by removing the CF₃ group (position R6) and adding a bulkier butyl group at the neighboring position (position R5). As predicted, these positions are facing away from the active site pocket. Therefore, the functional groups on these positions may not contribute much to the binding energy. In comparison, compound HTS11134 has two chlorine atoms on the distal benzene ring at positions predicted to make more intimate interactions with the protein (position R3 and R4), which may introduce steric clashing. Two analogs, HTS08390 and HTS08512, have modifications on the pyrazinecarboxylic acid moiety by replacing the nitrogen with carbon (position R1 and R2), which converts this moiety to benzoic acid. Neither of these compounds showed inhibition. As HTS08391 and HTS03728 did show inhibition, it seems that the modification of the pyrazinecarboxylic acid moiety is the cause for inactivity, which stresses the importance of this moiety.

Table 12 Additional Virtual Screening Results.

Compound	Structure	F score	CScore	MW (Da)	LogP
ST 201911		-39.70	5	322.09	0.56
HTS07012		-35.80	4	324.37	0.39
HTS03680		-35.94	4	327.11	0.65
524716		-39.52	4	179.13	-0.53

Additional screening based on “anchoring” moieties

As discussed earlier, phthalic acid and pyrazinecarboxylic acid moieties were predicted to interact with the phosphate-binding loop and probably function as “anchors” for inhibitor binding. Based on this, we predicted that more inhibitors would be found by testing compounds that also contain these moieties. We searched the databases on the ZINC website (<http://blaster.docking.org/zinc/>), and all compounds that contain one of these two moieties were collected and docked using FLEXX. However, only four compounds showed good statistics. Visual inspection of the docking results suggested, like the compounds from initial screening, that these compounds were also predicted to be anchored to the phosphate-binding loop. Except for compound HTS07012, all were docked in such a way that the phthalic acid or pyrazinecarboxylic acid moiety interact with the phosphate-binding loop. Three of these compounds (HTS07012, HTS03680 and 524716) were then purchased and tested. Only compound HTS03680 showed about 20% inhibition at 10 μ M concentration. (Table 12)

Conclusion

In this study, we solved the first MVK structure from *M. jannaschii* using selenium MAD phasing technique. Fold analysis shows that MVK shares a core structure with other GHMP kinases. The differences are mostly located in periphery regions. Therefore, the evolutionary conservation of this structural family is apparent, though the overall sequence identity among members are low. Comparison also shows a similar organization of the active pocket within most members of the GHMP family. An aspartate and lysine (arginine) pair is indicated as the catalytic residues in four out of six

GHMP proteins compared, which suggests a conserved mechanism for the phosphoryl transfer.

To reveal the mechanism for substrate binding, catalysis and inhibition of MVK, we solved several ligand bound structures. Based on these crystallographic data, in combination with kinetic assay, the following conclusion were made: i) The phosphate-binding loop is critical for the binding of nucleotides, terpenoids or compounds generated from virtual screening to the MVK active site. Although the loop itself does not change conformation, the phosphate tails of the nucleotides are flexible, which may ensure the minimization of the binding energy. A di-phosphate moiety is the minimal unit for interacting with the phosphate-binding loop. ii) The nucleoside base binding pocket provides several hydrogen bonding interaction sites. ATP seems to be preferred as its adenine ring forms the maximal amount of hydrogen bonding interactions allowed by the pocket. This pocket also has some degree of degeneracy, as is demonstrated in the UTP bound structure, in which a different binding orientation is adopted by the uridine ring. iii) The water in the structure near the catalytic residue Asp 155 indicates the potential location of the phosphoryl acceptor. Based on this, a mechanism consistent with all available data was proposed for MVK catalysis. iv) Studies on terpenoids explain why they are competitive with ATP. Different from nucleotides, the hydrophobic effect contributed by the carbon chains seems to play important roles for their binding. As the carbon chain is disordered in our structure where mevalonate does not bind, it is possible that the binding of mevalonate somehow facilitates the ordering of the carbon

chain, probably via MVK conformational changes, as a plausible mechanism for the uncompetitive nature of terpenoids with mevalonate.

Finally, we performed virtual screening in order to identify inhibitors against MVK. Most compounds with top scores were docked near the phosphate-binding loop. *In vitro* validation was then performed. Although some compounds showed limited inhibition, this process cannot be regarded with much success, since the inhibitory effects were weak and could not be improved through optimization. Therefore, the issue regarding the high false positive rate, as well as the potential problems associated with targeting the phosphate-binding loop, should be taken care of before a more successful experiment can be completed.

CHAPTER IV

BACKGROUND OF CELL DIVISION AND FTSZ

Bacterial Cell Division

Cell division is the fundamental process in any living systems. It is accomplished by two stages: partitioning of the genomic DNA and cytokinesis.

The bacterial genome is partitioned during DNA replication, as the replisome in the mid-cell provides the force for “pushing” DNA towards cell poles (109). Though the origin of the genomic DNA is segregated relatively early, the terminus remains at the mid-cell until the DNA dimer is resolved by enzymes such as XerCD and topoisomerase IV (109, 110, 111, 112). The partitioning of the bacteria genome is also facilitated by DNA condensation, mediated by proteins such as SMC (109, 113). As has been revealed by EM, the homodimer of a SMC has a symmetric structure composed of two long arms linked by a flexible hinge region (114). Individual arms contain the DNA binding site and the movement in the hinge region can shorten the distance between the arms, which may serve as the mechanism for DNA condensation (115).

The cytokinesis is initiated by the formation of the Z-ring, an apparatus mainly composed of the bacterial tubulin homolog FtsZ. The Z-ring formation is regulated by two mechanisms. The first is the nucleoid occlusion, as FtsZ cannot be assembled to form the Z-ring in the vicinity of the nucleoid. As the mid-cell region is normally occupied by the genomic DNA, this mechanism prevents the premature cytokinesis when the genome segregation has not yet completed. Recently, two proteins, Noc in *B.*

subtilis and SlmA in *E. coli*, have been identified as essential players for this process (116, 117). The second regulation mechanism is the Min system. (118, 119) It is composed of three proteins: MinC, MinD and MinE. MinC is the inhibitor of FtsZ polymerization and may also decrease the stability of the preformed Z-ring (118). MinC forms a functional complex with MinD and is anchored to the membrane by MinD. MinE is an anti-MinCD protein that displaces MinCD from the membrane. Importantly, MinCD complex oscillates from the pole to pole in *E. coli* every 30 to 60s, while MinE oscillates between the two $\frac{3}{4}$ positions, which creates a gradient of membrane-bound MinCD with its concentration at the cell poles to be the highest and at the mid-cell the lowest. This allows the Z-ring only to be assembled in the mid-cell area, if this region is cleared of the genomic DNA.

The exact mechanism of cytokinesis is still unclear, though it has been shown that FtsZ and the Z-ring are indispensable for this process (118). The Z-ring also functions as a platform for many other cell division proteins to bind. Some of them, like ZipA, may stabilize the ring structure and attach it to the membrane (118). Others, like FtsW, FtsN and FtsI, are involved in cell wall biosynthesis (118, 120). Their roles will be discussed in more details later.

FtsZ

FtsZ as a tubulin homolog

FtsZ is the bacterial homolog of tubulin (Figure 28). Although the overall sequence similarity is low, FtsZ and tubulin have the similar structure (121). Both proteins have a GTPase activity and contain a tubulin signature motif GGGTG[S/T]G.

Like tubulin, FtsZ can self-assemble into long filaments promoted by the binding of GTP. Additionally, GTP analog GMPCPP also facilitates FtsZ assembly (122). GMPPNP and GTP γ S, however, do not support polymerization, though the latter can stabilize the preformed FtsZ filaments (122, 123). Furthermore, GDP has been shown to destabilize preformed FtsZ filaments (123).

The morphology of FtsZ filaments have been investigated extensively with EM. Filaments assembled with GTP are mostly straight (122, 124). They can be either paired double protofilaments or sheets composed of multiple protofilaments. Divalent cations, like Ca²⁺, and polycations, such as DEAE-dextran, stabilize the FtsZ filaments and induce extensive bundling (125). This effect has also been observed when FtsZ binding proteins, such as ZipA, are incubated with FtsZ during the polymerization (126). FtsZ can also form helical tubes after being incubated with GDP in the presence of DEAE-dextran (127), GTP in the presence of Ca²⁺ or GMPCPP in the presence of Mg²⁺ (128). It is important to note, however, that these helical tubes are formed by helical protofilament, which is different from microtubules that are formed by the association of 13 to 14 straight tubulin protofilaments (129).

The crystal structures of FtsZ from *M. jannaschii* (121) *M. tuberculosis* (130), *T. maritima* (131) and *P. aeruginosa* (132) have been reported since 1998 (Table 13). Its N-domain is composed of a six-strand parallel β -sheet sandwiched by two and three helices on both sides, adopting the typical Rossmann-fold topology, containing the GTP binding site (121). Its C-domain is composed of a four-strand β -sheet packed with two helices (121). N- and C-domains are connected by a long central helix (121). There are

Table 13 List of FtsZ (a) and Tubulin (b) Crystal Structures.

Unless mentioned, X-ray crystallography was used for structure determination. α - and β -tubulins were used in these studies, unless mentioned otherwise.

a

Species	Resolution (Å)	PDB ID	Comments
<i>M. jannaschii</i>	2.8	1FSZ	monomer with GDP
<i>M. jannaschii</i>	2.7	1W59	straight dimer with apoprotein
<i>M. jannaschii</i>	2.4	1W5A	straight dimer with Mg ²⁺ -GTP
<i>M. jannaschii</i>	2.2	1W5B	straight dimer with GTP
<i>M. jannaschii</i>	2.5	1W58	monomer with GDP
<i>M. jannaschii</i>	3.0	1W5E	W319Y mutant in P1 space group
<i>T. maritima</i>	2.0	1W5F	domain swapped dimer, T7 mutated
<i>P. aeruginosa</i>	2.1	1OFU	complexed with inhibitor SulA
<i>M. tuberculosis</i>	2.1	1RLU	novel dimer with GTP γ S
<i>M. tuberculosis</i>	2.6	1RQ7	novle dimer with GDP
<i>M. tuberculosis</i>	1.9	1RQ2	novel dimer with citrate

b

Species	Resolution (Å)	PDB ID	Comments
<i>H. sapiens</i>	3.0	1Z5W	γ -tubulin
<i>S. scrofa</i>	3.7	1TUB	straight dimer with taxol (electron diffraction)
<i>B. taurus</i>	3.5	1JFF	straight dimer with taxol (electron diffraction)
<i>B. taurus</i>	3.0	1TVK	straight dimer with epothilone (electron diffraction)
<i>B. taurus</i>	4.1	1Z2B	with colchicine, vinblastine and a stathmin-like domain
<i>B. taurus</i>	3.6	1SA0	with colchicine and a stathmin-like domain
<i>B. taurus</i>	4.0	1FFX	with GDP and a stathmin-like domain

several “tublin” loops, designated as T1 to T7. The T1 to T6 loops are around the nucleotide binding pocket of the N-domain and are involved in the binding of GTP. As shown in the *M. tuberculosis* FtsZ, some of these loops are flexible and can adopt different conformations depending on the nucleotide ligand. For example, when the γ -phosphate is not present (apo or GDP state), the T3 loop is disordered (130).

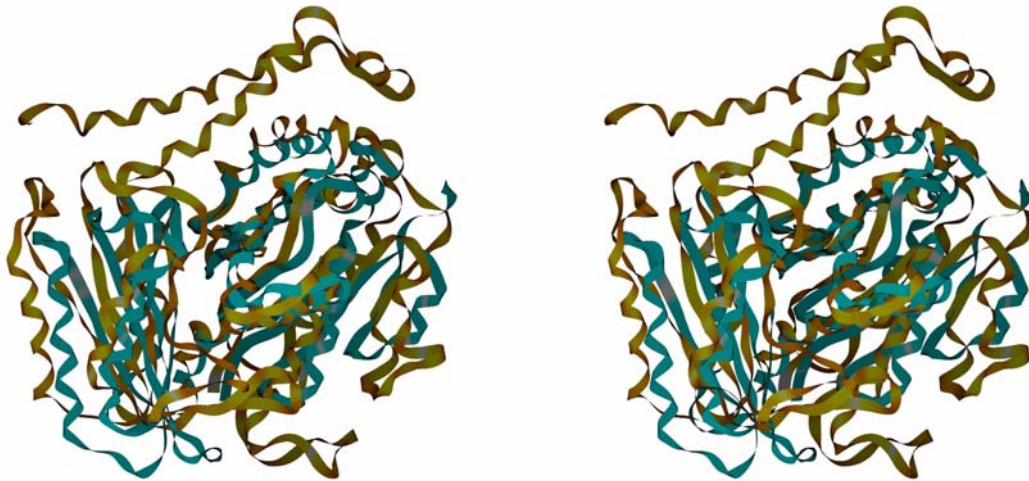


Figure 28 The Structural Alignment of Tubulin and FtsZ.

The Rat Tubulin (golden) and *M. tuberculosis* FtsZ (Cyan) are superimposed and presented in the stereoscopic form.

The T7 loop is located in the C-domain and interacts with the nucleotide binding pocket of the neighboring FtsZ monomer in the FtsZ filament. Experimental data suggests that the GTPase activity is triggered by the polymerization of FtsZ. Mutations

on the T7 loop severely decrease GTP hydrolysis, even in the mixture of mutant and wildtype proteins (133). These data are consistent with the hypothesis that the active site for GTPase is formed by two monomers, as has been suggested from studies on the tubulin.

A crystal structure of the *M. jannaschii* FtsZ protofilament has been reported (131). It shows that two key residues from the T7 loop (Asp 235 and Asp 238) insert into the nucleotide binding pocket in the neighboring monomer, which polarizes a water molecule that seems to be the nucleophile to attack the γ -phosphate. This is similar to the effect of GAPs on G-proteins, in which catalytic residues are provided by the partner across the dimeric interface (134).

Dynamics of the FtsZ filaments

The stability of the FtsZ filament is weakened by the GTP hydrolysis, as inhibiting the GTPase activity (e.g. by depleting Mg^{2+}) allows the fiber to persist (135, 136). Factors modulating the GTPase activity, such as Mg^{2+} , KCl and Ca^{2+} , have profound effect on FtsZ dynamics (135, 136). Some data suggest that FtsZ monomers within a filament are predominantly bound with GDP (probably in the form of GDP: P_i). (123) Therefore, the mere hydrolysis of GTP does not ensure the depolymerization and it has been suggested that one or a few layers of GTP containing subunits at the end of the FtsZ filament might be sufficient to stabilize the whole structure (123), similar to the GTP cap in microtubules (129). However, a more recent study suggests that the FtsZ filament is predominantly GTP bound, which indicates that the hydrolysis of GTP might be the rate-limiting step for filament dissociation instead (137).

Proteins interacting with FtsZ

As the Z-ring is the central apparatus for the cell division, it is not surprising that many proteins make interactions with FtsZ. They may either play some roles in regulating the Z-ring assembly or have their own functions in the cell division (Figure 29a).

There are two types of Z-ring regulators: the inhibitor and the stabilizer. Important Z-ring inhibitors include MinC, as already described. Another important inhibitor is Sula, a protein induced by SOS-response when DNA damage is encountered (132). It binds to FtsZ at the T7 loop containing face and forms an elongated complex Z:A:A:Z (Z for FtsZ and A for Sula). This caps the FtsZ filament, preventing it from further growth (Figure 30a) (132). A third negative regulator is EzrA in *B. subtilis*, a membrane anchor conserved in several Gram-positive bacteria (122). The Z-ring stabilizers include FtsA, ZipA and ZapA. FtsA is an actin homolog that interacts with the C-terminal region of FtsZ. To ensure the proper cell division, a cellular FtsA to FtsZ ratio at about 1:100 is maintained (138). ZipA is a membrane protein that contains motifs common to microtubule-associated proteins. Like FtsA, it interacts with the C-terminal tail of FtsZ. A crystal structure of the complex formed by ZipA and C-terminal 17 residues of FtsZ shows that the FtsZ C-terminal tail forms a β -strand followed by an α -helix and binds in a shallow and hydrophobic cavity in the ZipA C-terminal domain (Figure 30b) (139). ZapA is another Z-ring stabilizer. It is a small protein largely composed of coiled-coil and can dimerize or tetramerize, which probably crosslinks several FtsZ polymers (140).

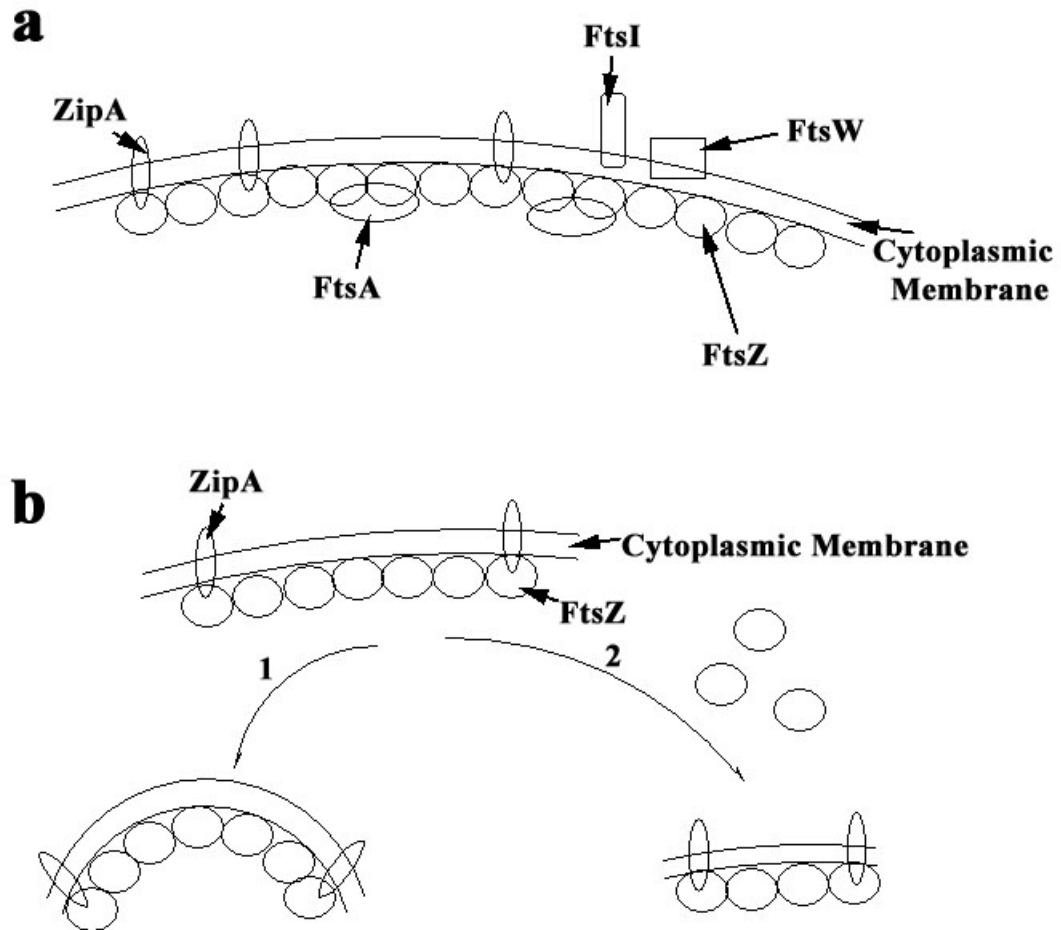


Figure 29 Composition and Constriction of the Z-ring.

a. Proteins Interacting with the Z-ring

b. Mechanism of the Z-ring Constriction

Constriction is triggered by the transition of the FtsZ filament from straight to curved (1) or the depolymerization (2) of the FtsZ filament.

Other division proteins, not directly involved in the Z-ring stability, are also recruited to the Z-ring, either directly by binding to FtsZ or indirectly by interacting with FtsZ binding proteins. Examples include DNA condensing protein MukB (113) and nucleoid occlusion protein SlmA (117). Components of the septum cell wall biosynthesis machinery, such as FtsW, FtsI, FtsN and AmiC, are also recruited (118). It is generally accepted (118) that FtsW is recruited first, followed by FtsI, a transpeptidase. FtsI then recruits FtsN that further recruits AmiC. AmiC cleaves the peptide moiety from N-acetylmuramic acid on the cell wall and allows the cell division to proceed, while FtsN may function as a mechanic linkage connecting the septum ring to the cell wall (118).

FtsW is a membrane bound protein conserved among most bacteria and contains 10 transmembrane helices (118). In addition to the recruitment of FtsI, it also plays a role in translocating lipid-linked precursors for cell wall synthesis (118). It is unclear how FtsW itself is recruited to the Z-ring. In *M. tuberculosis*, FtsW possesses a unique C-terminal extension that interacts with the C-terminal tail of FtsZ (141). It seems that in *M. tuberculosis*, FtsW can directly bind to FtsZ (141). Since neither FtsA nor ZipA is present in *M. tuberculosis*, FtsW might provide an alternative mechanism for the Z-ring stabilization and membrane attachment in this organism.

Questions regarding the *in vivo* function of the Z-ring

The architecture of the Z-ring still remains elusive. It has been suggested that a sheet of FtsZ polymer encircles the cell, which requires each FtsZ fiber to be able to span the whole circumference of the mid-cell and is not supported by available

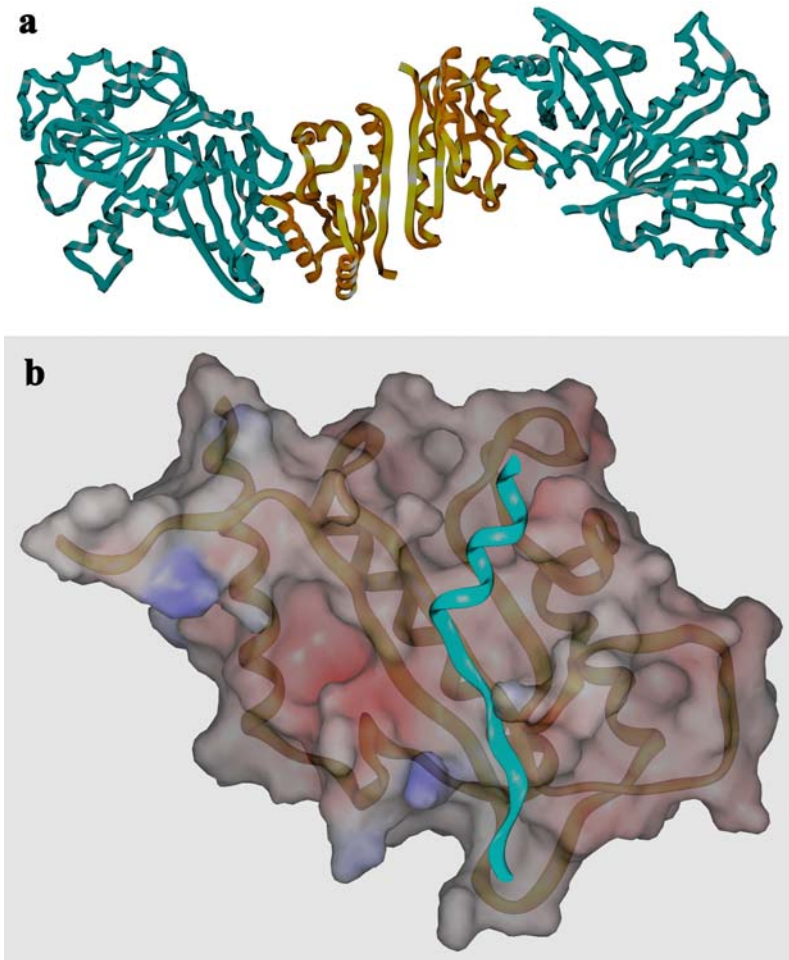


Figure 30 Structures of FtsZ in Complex with Accessory Proteins.

a. The Complex between FtsZ and SulaA

FtsZ is shown in cyan and the SulaA dimer is shown in golden.

b. The Complex between ZipA and the FtsZ C-terminal Domain

ZipA is shown in golden. The FtsZ C-terminal domain forms an α -helix and a β -strand and colored in cyan. The surface is colored according to the electrostatic potential as was described earlier.

experimental data (142). More recent studies suggest that the Z-ring may be constructed by the alignment of many short filaments, either by forming lateral interaction directly between filaments or by FtsZ binding bridging proteins (142, 143). This seems to be more likely, as it does not require a long FtsZ filament as suggested from the first model. *In vitro* observation also reveals that the FtsZ filament have various curvatures, from straight fibers to helical tubes, as described earlier. However, the curvatures of the filaments *in vivo* remain elusive (142, 143).

It is known that at the late stage of cytokinesis, the Z-ring undergoes constriction. What is the mechanism for constriction? Several models have been suggested. One hypothesis involves the transition from the GTP bound, straight FtsZ filament to the GDP bound, curved FtsZ filament (144) (Figure 29b). However, it is unclear whether this type of transition in filament geometry can provide enough force. Another hypothesis suggests that the aligned short FtsZ filaments may slide onto each other, driven by unknown motor proteins, as the myosin and actin do in muscle cells (142). This requires a discovery of unknown motor proteins, which has not been made yet. In a third model, the constriction is suggested to be triggered by the depolymerization of the FtsZ filament, as has been demonstrated in microtubules (142) (Figure 29b). Finally, the FtsZ filament might simply be a passive player, as the inward growth of the peptidoglycan may provide the major force for the Z-ring constriction (118).

Significance of This Work

Although much progress has been made in FtsZ, the structure of the FtsZ filament remains inconclusive. It has been proven that FtsZ can adopt the tubulin-like straight

filament under most *in vitro* experimental conditions. However, the nature of the helical filaments that have also been observed in many circumstances is unclear. In addition, the mechanism of FtsZ polymer disassembly, induced by GTP hydrolysis and the γ -phosphate release, also remains elusive. These questions need to be addressed first before we are able to find answers for questions that are more biologically relevant, such as the Z-ring assembly and dynamics. In the next chapter, we will present a structural biology study that sheds light on the structure of helical filaments and the possible mechanism of FtsZ polymer dissociation.

CHAPTER V

STRUCTURAL STUDY OF AN FTSZ CRYSTAL FILAMENT

Methods

Cloning and purification of FtsZ

The full-length *M. tuberculosis* FtsZ was cloned and purified by our collaborator Lucile White from Southern Research Institute by following the published protocol (135). Selenomethionine substituted FtsZ was produced using a similar protocol as stated in chapter III.

A truncated version of FtsZ (with the removal of the C-terminal region from residue 313 to 376) was also cloned as follows. The *M. tuberculosis* genomic DNA was used as the template. Each reaction (50 μ l) contained 1 μ g genomic DNA, 0.5 μ l (25 pmol) of each primer, 2.5 μ l DMSO, 1 μ l 10 mM MgSO₄, 5 μ l 10 \times thermal buffer (New England Biolabs), 10 μ l 2.5 mM dNTP and 1 μ l Deep Vent DNA polymerase (2 unit, New England Biolabs). The N-terminal primer had the following sequence: 5'-AGA GAA *GCA TAT GAC* CCC CCC GCA CAA C-3', containing an NdeI restriction site (italic and bold). The C-terminal primer had the following sequence: 5'-CCC *AAG CTT TTA* TCA GAA GCC GGC CGC GAT CAC-3', containing a HindIII restriction site (italic and bold) and a stop codon (bold and underlined). 30 cycles of PCR were performed with the temperature profile of 94 °C for 30 seconds, 65 °C for 30 seconds and 72 °C for 90 seconds. There was also a 5 minutes at 94 °C hot start step before the initiation of the PCR cycles and a 7-minute extension at 72 °C after completion of the

PCR cycles. The PCR products were purified on a 0.8% agarose gel containing ethidium bromide and cloned into the pET28b vector. The construct was transformed into BL21(DE3) competent cell and verified by gene sequencing.

The truncated protein (designated as the FtsZ₁₋₃₁₂) was expressed in Rossetta(DE3) *E. coli* strain. Cells were grown in LB medium at 37 °C. When OD₆₀₀ reached 0.5, gene expression was induced by 1 mM IPTG at 16 °C for 20 hours and were then harvested and kept at -20 °C. Cell pellets were thawed and protein purified using a Ni column, following the same protocol as described in Chapter III. The fractions containing the FtsZ₁₋₃₁₂ protein were pooled, concentrated and dialyzed against buffer with 25 mM HEPES (pH 7.2), 100 mM NaCl, 1mM DTT, 0.1 mM EDTA and 10% glycerol. The N-terminal his-tag was removed using thrombin (restriction grade, Novagen) by incubating 45 µl thrombin (45 unit), 4.5 ml FtsZ₁₋₃₁₂ (10 mg/ml), 1 ml 10× thrombin cleavage buffer (Novagen) and 4.5 ml dH₂O at room temperature for 2.5 hours. The reaction mixture was then concentrated and loaded onto a superdex 75 size-exclusion column and eluted with 25 mM HEPES (pH 7.2), 100 mM NaCl, 1mM DTT, 0.1 mM EDTA and 10% glycerol. Fractions containing the FtsZ₁₋₃₁₂ protein were eluted from a wide range, from aggregates to monomers.

Polymerization test

The polymerization of FtsZ (both the truncated and full-length, 10mg/ml to 20 mg/ml) is performed by mixing 20 µl FtsZ, 10 µl 100 mM GTP and 460 µl assembling buffer (100 mM MES pH 6.4, 100 mM NaCl, 12.5 mM MgCl₂). Polymerization was verified by centrifuging the mixture at 100 K in a Beckman TLA100.3 rotor for 30

minutes at 4 °C. 100 µl of the supernatant was removed. The pellet was washed with 100 µl assembling buffer and resuspended in 100 µl assembling buffer. Both the pellet and supernatant were analyzed using SDS-PAGE.

Polymerization results in a 90° rotation of light scattering and reduces the transmission light, which corresponds to an increase in the absorbance and can be directly measured by a conventional spectrometer in real time (142). 20 µl FtsZ was first mixed with 460 µl assembling buffer in a quazi cuvette. The background absorption at 340 nm was monitored for 2 minutes. GTP was then added up to 40 µM and the polymerization of FtsZ was monitored by absorption at 340 nm for 15 to 60 minutes.

EM analysis of FtsZ filaments

The FtsZ protofilament was produced by mixing 4 µl of 10 to 20 mg/ml FtsZ, with 4 µl of 100 mM GTP and 42 µl of assembling buffer. In addition, 0.2 to 1 mg/ml DEAE dextran or 10 mM CaCl₂ was included in some experiments to study the effect of polycationic stabilizer or divalent cations on polymerization. The reaction mixture was left at room temperature for 10 minutes and was then diluted for 2 to 5 fold using the assembling buffer before making grids and being sent for EM analysis (the last two steps were performed by Christos Savva from the Department of Biology).

Crystallization

The full-length FtsZ protein (10 mg/ml) was incubated with 2 mM GDP and 10mM EDTA for 1 hour on ice before setting up the crystallization reaction. Crystals were grown using the hanging drop vapor diffusion method, with 1 µl protein mixed with 1 µl crystallization solution. The crystallization condition is 1.4 to 1.6 M (NH₄)₂SO₄,

0.1 to 0.7 M NaCl, 0.1M HEPES pH 6.8 to 7.3 (145). Crystals appeared in 3 to 5 days to a size of 0.3 x 0.2 x 0.2 mm. Selenomethionine substituted proteins were crystallized under the same condition except with the inclusion of 10 mM DTT.

These crystals diffracted weakly at synchrotron to the resolution ranged from 4 to 8 Å. Optimization trials using additives such as 10 mM ZnCl₂ or 10 mM CaCl₂ did not improve the diffraction power. Because this type of FtsZ crystal has a large unit cell with a high solvent content (about 0.68), a crystal dehydration protocol was tried by sealing the cover slip with a crystal containing drop over a reservoir well containing a higher concentration of (NH₄)₂SO₄ (2.5 to 3.5 M) and allowing to equilibrate for 2 to 3 days. Although there was no visible crystal deterioration, diffraction power was lost. As FtsZ has an unstructured C-terminal region, therefore we tried to crystallize the truncated protein (FtsZ₁₋₃₁₂), under the same crystallization condition. Though it crystallized faster than the full-length protein, it diffracted only to the same resolution and its spots were highly smeared and data processing was not possible. This is most likely due to less order in the crystal.

Heavy atom soaking

Heavy atoms soaking was also attempted to improve crystal diffraction and provide experimental phase information that is unbiased by the model. In order to identify heavy atoms capable of binding FtsZ, compounds were incubated with the FtsZ sample at room temperature for 30 minutes in the presence of the crystallization solution. Samples were then analyzed by a native PAGE (146). Four mercury compounds, including HgCl₂, ethyl mercuric phosphate, p-aminophenyl mercuric acetate and

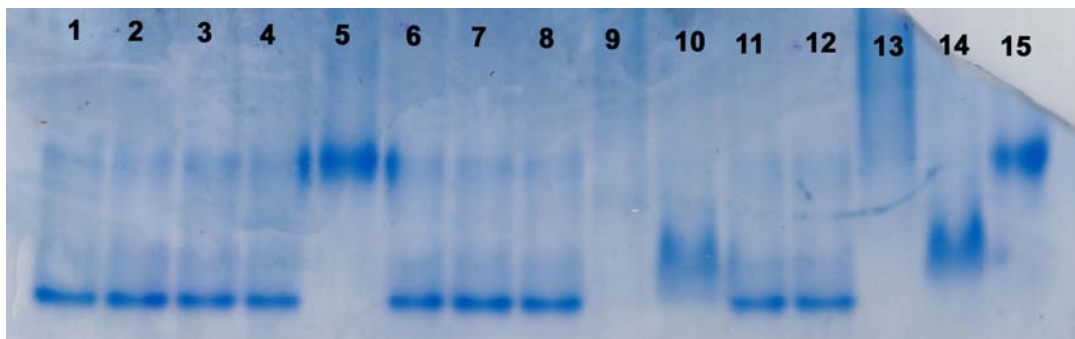


Figure 31 Use of Native PAGE to Screen Heavy Atoms.

- Lane 1: No heavy atom**
- Lane 2: Sodium Tungstate**
- Lane 3: Potassium Tetracyanoplatinate (II)**
- Lane 4: Samarium (III) Chloride**
- Lane 5: Mercury (II) Chloride**
- Lane 6: Lead (II) Acetate**
- Lane 7: Potassium Hexachloroiridate (IV)**
- Lane 8: Ammonium Tetrachloroplatinate (III)**
- Lane 9: Sodium Hexachloroplatinate (IV)**
- Lane 10: Ethyl Mercuric Phosphate**
- Lane 11: Triethyl Lead Acetate**
- Lane 12: Europium (III) Chloride**
- Lane 13: PIP**
- Lane 14: Thimersol**
- Lane 15: p-aminophenyl mercuric acetate**

thimersol, led to a shift of the protein band on the native PAGE, indicating the binding of the heavy atom to the protein (Figure 31). Among them, thimersol was the only one that did not visibly damage the crystal, and thus, crystals were soaked with 2 mM thimersol for about 20 hours and one dataset was collected to 5 Å. But no heavy atom was found by reviewing either the difference Patterson map or the F_0-F_c map. One possible explanation is when the pH is above 6.0, $(\text{NH}_4)_2\text{SO}_4$ can release NH_3 that, as a

good nucleophile, will compete with protein for heavy atoms. Attempts were made to transfer the crystal into new drops that contain PEG instead of ammonium sulfate, but crystals invariably got damaged.

Data collection and crystallographic methods

Crystals were soaked into the mother liquor containing 20% to 25% glycerol before being flash cooled to 100K for data collection. Both a native dataset and a three wavelength MAD dataset (peak, inflection and low energy remote, at 0.97912, 0.97934 and 1.00531 Å, respectively) were collected at 14IDB at APS. For each dataset, the data were collected with a half-degree oscillation and a crystal to detector distance at 300 to 400 mm. A range of 90° was collected for the native dataset. For the MAD data, angle of 180° was collected.

The data were processed with HKL2000 (Table 14). The space group is P6(4)22. Both the native and MAD data showed good statistics up to 4.0 Å. However, the MAD data had low completeness (less than 40%) at resolutions higher than 4.6 Å. The molecular replacement was performed in MOLREP and EPMR, while the MAD phasing was done using SHARP. The model building was performed in XtalView. The model was refined in programs CNS, CORRELS and REFMAC.

Results and Discussion

The polymerization of FtsZ

The ability of FtsZ to form polymers was analyzed. Pelleting assay showed most FtsZ was retained in the supernatant fraction when GTP was omitted. After the addition of 2 mM GTP, most FtsZ was in the pellet fraction, indicating the polymer formation

(Figure 32a). Monitoring absorption at 340 nm suggested that the polymerization of FtsZ clearly reassembled a two phase process. There was a rapid phase, in which the absorbance quickly reached the half height of the maximum and lasted for about 60 seconds. Followed was a slow phase, when the absorbance continued to increase for about one hour. (Figure 32b) This phenomenon is in agreement with the published data collected by directly monitoring the scattering light (135).

The FtsZ filament was observed under EM (Figure 33a). Each filament seems to be composed of two dense regions separated by a gap. The dense region is probably corresponding to a single FtsZ protofilament (Figure 33b). The width across whole filaments is about 90 Å and that for the gap is about 20 to 30 Å. If polycation stabilizers like DEAE-dextran or divalent cation such as Ca^{2+} were present, extensive bundling could be formed, which may suggest that these compounds induce more lateral interactions. These observations are consistent with those made in FtsZ fibers from other sources (124, 125, 147). The polarity of these protofilaments is unclear, though there are reports suggesting that they are in parallel (125, 147). And the putative lateral interaction interface has been (125). However, this has not been supported by mutation studies (148).

E. coli FtsZ was reported to form a helical shaped filament under the condition of GDP and DEAE-dextran (127). From this study, however, we found that the *M. tuberculosis* FtsZ did not polymerize under this condition. Instead, massive aggregation was identified under EM (data not shown). Published studies in the *E. coli* FtsZ also suggested that GDP (without DEAE-dextran) could destabilize pre-formed *E. coli*

Table 14 Data Collection Statistics of FtsZ Crystals.

	Native	Peak	Inflection	Low Energy Remote
Space Group	P6(4)22	P6(4)22	P6(4)22	P6(4)22
Unit Cell Parameters a, b, c (Å) α, β, γ (°)	131.24, 131.24, 328.27 90.0, 90.0, 120.0	131.09, 131.09, 327.78 90.0, 90.0, 120.0	131.56, 131.56, 328.49 90.0, 90.0, 120.0	131.86, 131.86, 328.99 90.0, 90.0, 120.0
Wavelength (Å)	0.90000	0.97912	0.97934	1.00531
Resolution Range (Å)	50.0-3.90	50.0-3.90	50.0-3.90	50.0-3.90
R_{sym}	0.071	0.074	0.065	0.052
Completeness (%)	99.4	58.6	57.2	61.0
I/σ(I)	38.4	23.7	24.2	26.0
MAD Figure of Merit	0.43			

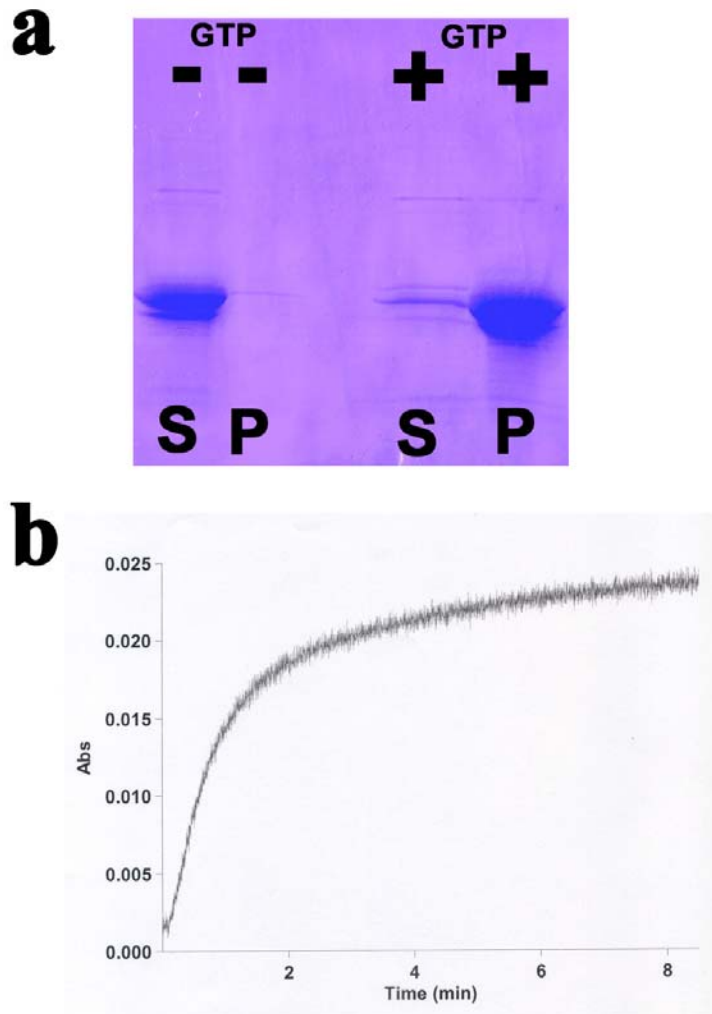


Figure 32 Polymerization Assay of the Full-length FtsZ.

- a. The supernatant (S) and pellet (P) after ultracentrifugation of the FtsZ sample with (+) or without (-) 2 mM GTP
- b. The time course of FtsZ polymerization monitored by absorbance at 340 nm
- c. The time course of slow FtsZ filament disassembly after the addition of GDP monitored by absorbance at 340 nm
- d. The time course of polymerization of FtsZ was monitored by absorbance at 340 nm after being pre-incubated with 0 μ M GDP (1), 100 μ M GDP (2) and 2 mM GDP (3)

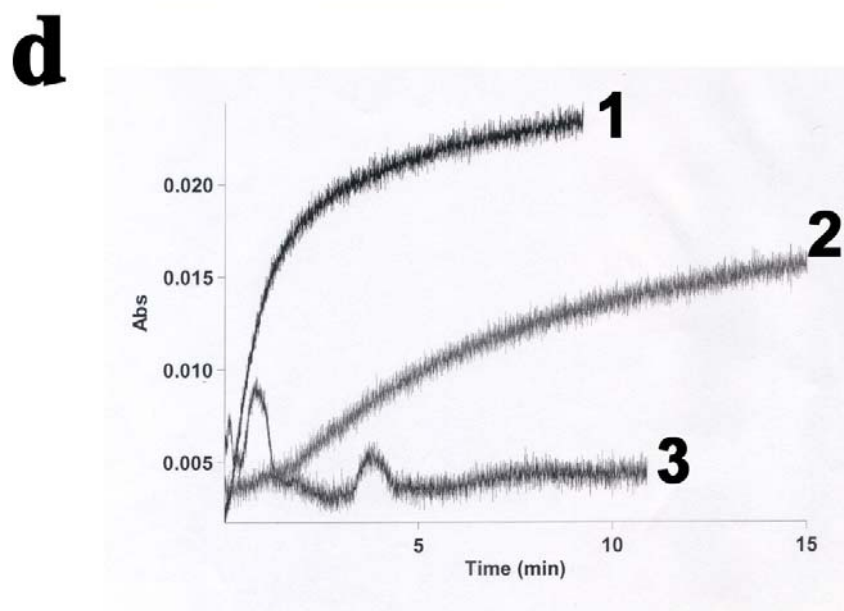
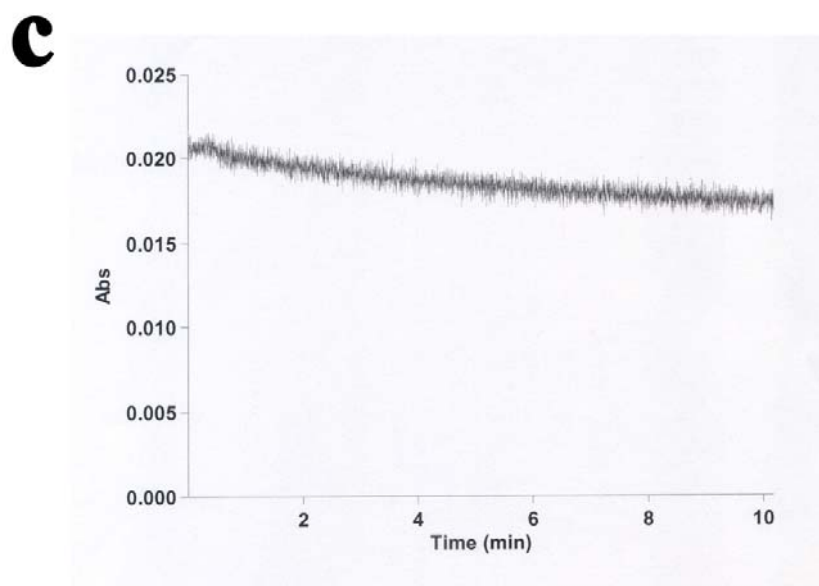


Figure 32 Continued.

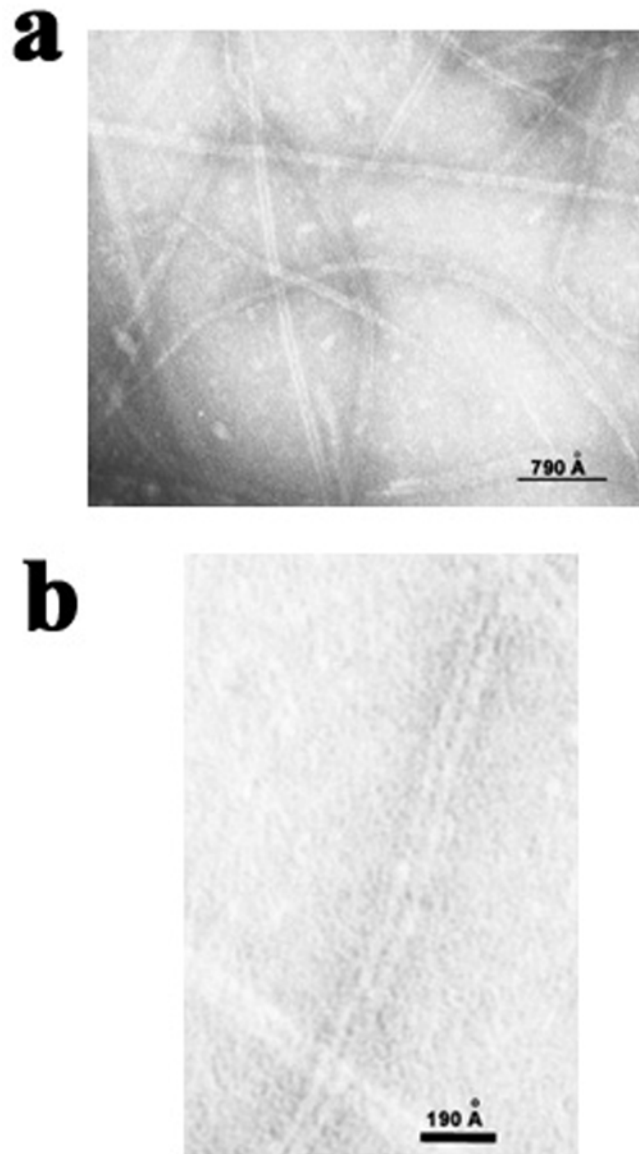


Figure 33 EM Images of FtsZ Filaments.

- a. The filaments formed by the full-length FtsZ**
- b. The enlarged image from (a) showing the morphology of a single full-length FtsZ filament**
- c. The filaments formed by the truncated protein FtsZ₁₋₃₁₂**
- d. The enlarged image from (c) showing the morphology of a single FtsZ₁₋₃₁₂ filament**

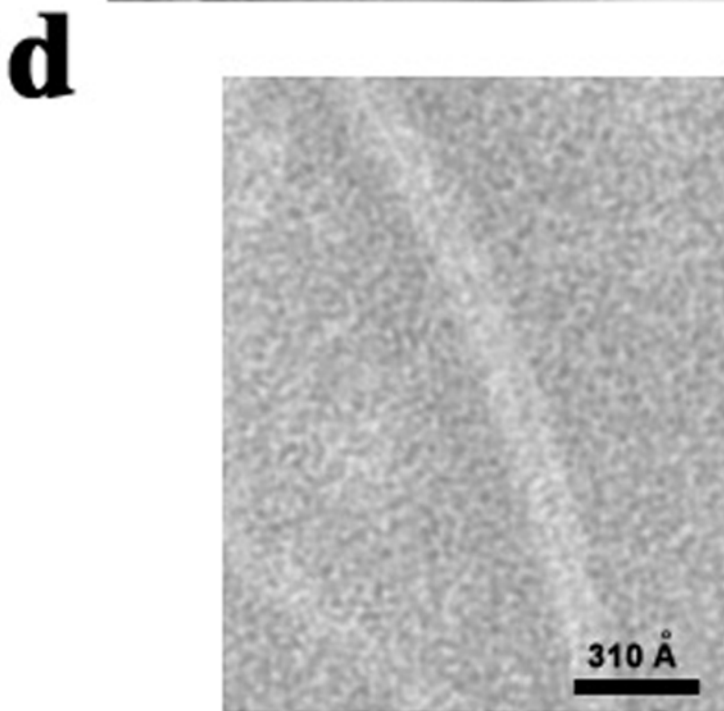
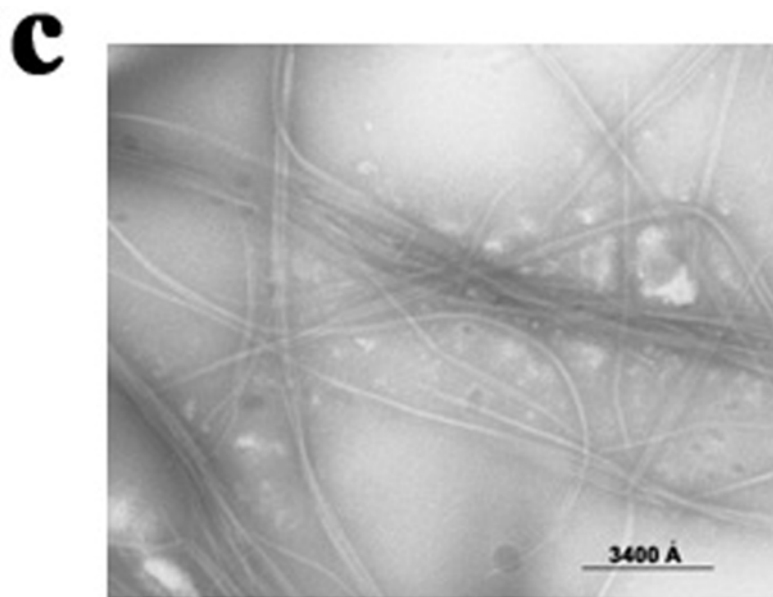


Figure 33 Continued.

FtsZ filaments and pre-incubation of FtsZ with GDP would prevent the formation of polymers by the subsequent addition of GTP (123). Our studies on the *M. tuberculosis* FtsZ suggested that by adding GDP, up to 2 mM, into the preformed FtsZ filament, we observed a slow decrease in the absorption at 340 nm (Figure 32c). Compared with the *E. coli* data, the destabilization (if any) effect is much slower, which may be consistent with the reported slow kinetics of the *M. tuberculosis* protein (135). Pre-incubation of the *M. tuberculosis* FtsZ with GDP, however, indeed inhibited the polymerization in a concentration dependant manner. 2 mM GDP was sufficient to completely block the formation of polymers, judged from monitoring the absorbance change (Figure 32d). When GDP concentration is lowered to 100 μ M, polymerization was allowed at a slow rate (Figure 32d). Based on these data, we conclude that GDP is a destabilization agent for the *M. tuberculosis* FtsZ filament.

Structural solution of the FtsZ crystal filament

Crystal structures from the *M. jannaschii*, *P. aeruginosa* and *M. tuberculosis* FtsZ (Table 13a) were used as models for molecular replacement trials in MOLREP. The structures of the *M. jannaschii* and *P. aeruginosa* FtsZ did not give any solutions. The crystal structure of the *M. tuberculosis* FtsZ shows a novel dimer composed of two molecules adopting different conformations. One molecule is in the “open” form that binds the ligand GTP- γ -S, while the other one is in the “closed” form in which the GTP binding site is empty. Although the dimer and the “closed” molecule did not give molecular replacement solutions, using the “open” monomer as the model, we found one solution composed of two monomers with an R factor of 0.540 and a correlation

coefficient of 0.324. During the process of molecular replacement, both space groups P6(2)22 and P6(4)22 were tried and only the later one led to a correct solution, confirming the proper space group is P6(4)22. A $2F_o-F_c$ electron density map calculated based on this solution clearly showed the secondary structural elements. However, the packing of the model, as well as the high R factor, indicated that there should be at least one more molecule in the asymmetric unit.

By superimposing the “open” and “closed” molecules of the original FtsZ dimer, large structural variations were identified, particularly loops in the N-domain near the nucleotide binding pocket. These regions also correspond to poor electron densities from the $2F_o-F_c$ map calculated using the partial model. We assumed these regions in the model were different from the actual crystal structure. Therefore, they were deleted from the model and the revised model was used for a new round of molecular replacement in EP MR to find the third molecule. The structure was refined using the rigid body refinement protocol in REFMAC and the R factor was lower than 0.50.

In order to obtain more phase information, MAD phasing was tried. However, due to the low resolution, the anomalous signal from the crystal was low. The anomalous signal to noise ratio calculated from XPREP was less than 1.5 as the resolution was higher than 6.0 Å. We were not able to find correct heavy atom positions either using the Patterson search or direct methods. However, an anomalous difference map, calculated based on the phases from the refined molecular replacement solution and the anomalous difference from the peak dataset, showed strong positive densities around the sulfur sites of some methionine residues (Figure 34). In all, 24 potential selenium sites with

densities contoured above 5σ were found per asymmetric unit. Among them, 20 selenium sites were manually built into densities contoured above 8σ and this constellation of anomalous scatterers was used for protein phasing in SHARP. At first, all three wavelengths were used. However, we failed to refine heavy atom sites using this data. Later on, it was found that the low energy remote dataset was problematic, probably because it had little anomalous signal. Therefore, only the peak and inflection datasets were used and the B factors of heavy atoms were constrained at 60.0 during the refinement. The refined selenium atoms were inspected and were seen restrained near their original locations, with satisfactory occupancy (only one atom had very low occupancy, at 0.07). The final figure of merit was 0.43.

The $2F_o-F_c$ map based on the molecular replacement phases was sufficiently well defined so most secondary structures were recognized (Figure 35). In comparison, the MAD phases were relatively poor even after solvent flattening and NCS averaging in DM, although interpretable densities corresponding to secondary structures could be recognized. Maps based on the combination of the MAD and molecular replacement phases were more useful because they showed improved density in some regions. For example, using the combined phases the map showed a density corresponding to the loop from residues 172 and 173 in the molecule A that was missing in the $2F_o-F_c$ map. Combination of the MAD and molecular replacement phases were performed in SIGMAA. Phase extension, density modification and multiple-crystal averaging were then performed in DMMULTI (149). We also tried to further improve the map by using B-factor sharpening, which was suggested as useful for low resolution data (150).



Figure 34 Anomalous Difference Map.

The densities corresponding to anomalous scatterers are shown in golden contoured at 5σ . The methionine residues are also displayed.

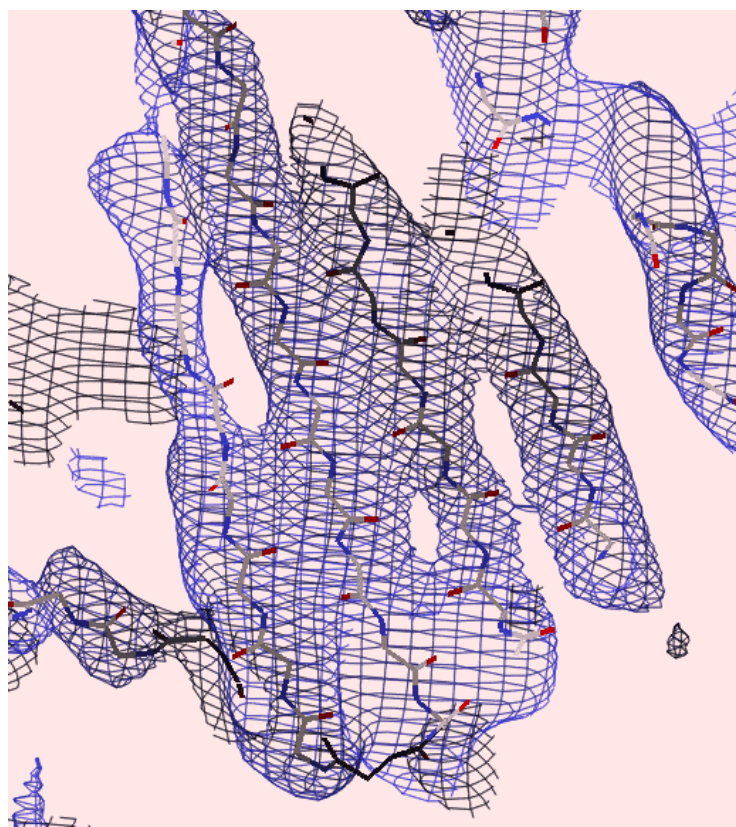


Figure 35 Electron Density Map for FtsZ.

The $2F_o-F_c$ map contoured at 1σ is shown. The main chain of the FtsZ model is also displayed. This region corresponds to the large β -sheet in the C-terminal domain of FtsZ.

Basically, a negative B-factor (-100 to -200) was applied to each amplitude of the MAD data by CAD. The modified amplitudes were subsequently used for map calculation. However, this procedure does not lead to significant improvement on the electron density map of FtsZ.

Rebuilding of the model was performed in regions where the correct electron density for protein structure was unambiguous. The final F_o-F_c map showed a density

above 3σ near the residue Phe 180. As this was the nucleotide binding site and this density resembled a GDP, it was recognized and fit with a GDP molecule (Figure 36a). The major problem for refinement was the low data to parameter ratio, intrinsic to the low resolution diffraction data. Initially, the refinement was mostly restricted to the rigid body refinement, by which we were able to drop R and R_{free} factors to 0.38 and 0.43, respectively. We later found that the model could be restrained properly by applying a very low X-ray term weighting factor (0.001 vs. 0.3 as the default) and tight geometry restraints (restraints for bond distance, bond angle and plane at 20, 5 and 40, respectively vs. 1, 1 and 1 as the default). NCS restraints were also applied. This protocol allows for the restrained refinement being carried out in REFMAC, which decreased the R factor and R_{free} factor to 0.29 and 0.38, respectively. The final model shows an average B factor at 169.4 Å². Such a high B factor was also observed in several other low resolution crystal structures, such as the *E. coli* ABC transporter MsbA (1JSQ), *S. lividans* K⁺ channel KcsA (1BL8) and SIV gp120 envelope glycoprotein (2BF1), and may contribute to the rapid decrease in diffraction intensities as the resolution was increased beyond 4.0 Å in our crystal.

Table 15 Refinement Statistics of FtsZ.

No. of Atoms/Residues	
Protein	841
Ligand	3
Maxium Resolution (Å)	4.0
R_{work}	0.294
R_{free}	0.377
Average B factor (Å²)	169.42
r.m.s.d from Ideal Geometry	
Bond Distance (Å)	0.004
Bond Angles (°)	3.375
Ramachandran Plot (%)	
Favored	34.3
Allowed	42.1
Generously Allowed	16.0
Disallowed	7.6
Model Figure of Merit	0.61

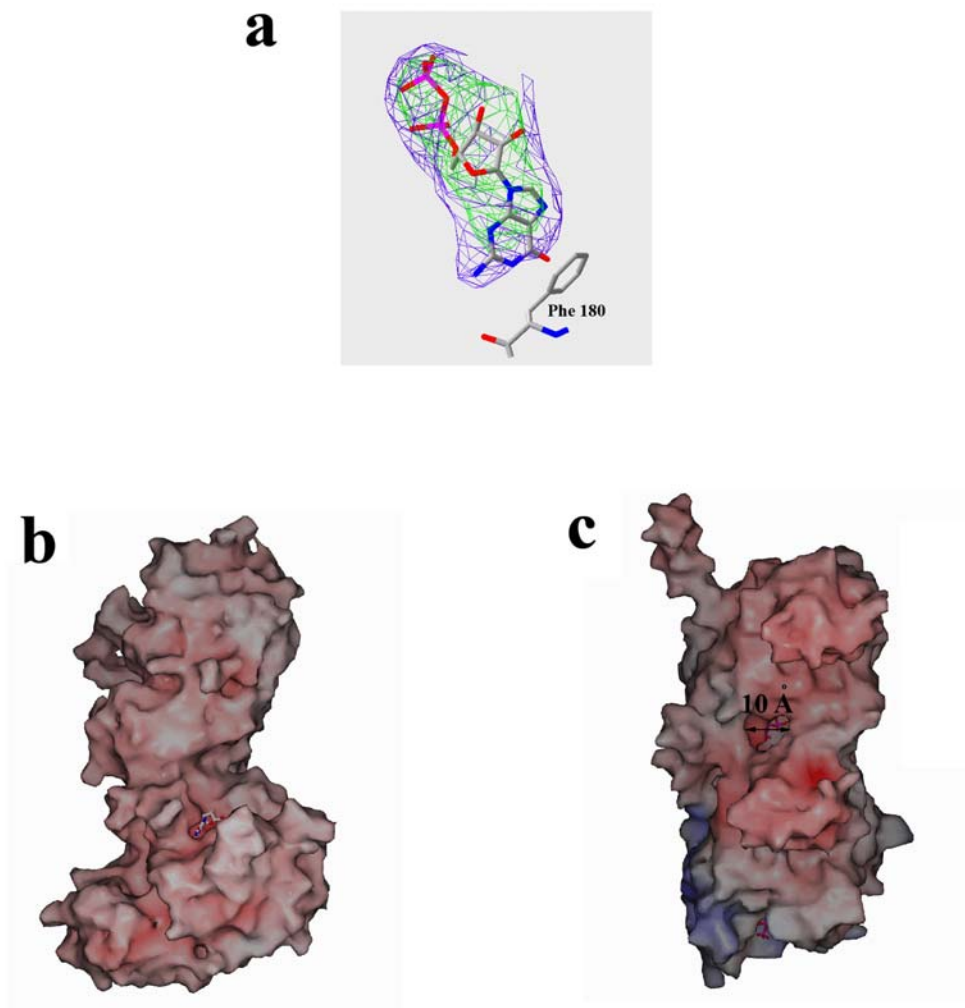


Figure 36 The GDP Binding Site.

a. The F_0-F_c Map (Purple Contoured at 2σ Level and Green at 3σ Level) near Phe 180 Fit with One GDP Molecule

b. The Accessible Surface of the Curved Dimer

c. The Accessible Surface of the Straight Dimer

The overall crystal structure

The final statistics are in Table 15. There are three FtsZ molecules arranged head to tail in each asymmetric unit. Each molecule contains residues up to Gly 311 or Phe 312. Residues after that are disordered in the absence of accessory proteins, as in all reported FtsZ structures. In general, the electron density map for the N-domain (residues 1 to 161) is poorer than that for the C-domain (residues 162 to 312). Consequently, the model is more complete for the C-domain. Particularly, poor densities are identified around residues 34 to 41, 9 to 17, 60 to 70 and 94 to 103. These regions are around the GTP binding site and may be involved in conformational switch (see below).

By applying the crystal symmetry, a left-handed antiparallel continuous double helical crystal filament with the diameter about 131 Å has been constructed. Its diameter is the same as the **a** and **b** axes of the crystal unit cell, as filaments are arranged in parallel to these axes. The surface of the helical tube is negatively charged. Its pitch is around 130 Å, with six molecules for each turn. Between two strands of the double helical filament, there is a gap of about 20 Å and no direct contact between protofilaments has been identified (Figure 37a, b).

In order to get more insights into the crystal filaments, we have looked into the literature. Many proteins can form filaments and several of them have been characterized by crystallography. For example, the DNA recombination protein RecA forms a right-handed helix, with its pitch about 65 to 95 Å. Interestingly, the pitch of the RecA filament depends on the liganded state of this protein, as the ADP bound fiber shows a shorter pitch and the ATP bound fiber shows a longer one (151). Tubulin can also form

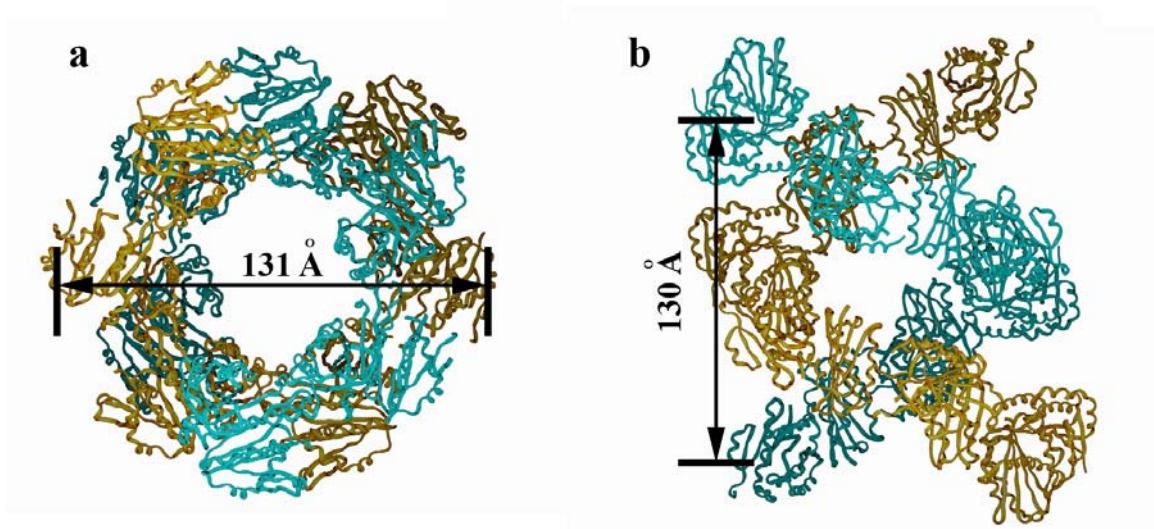


Figure 37 The Double Helical Structure of the FtsZ Crystal Filament.

a. The top-down view

b. The side view

In both views, the two strands are colored in cyan and golden, respectively. The diameter and pitch of the helix are labeled.

the helical filament. The 4 Å crystal structure of the tubulin:stathimin-like domain complex shows that the tubulin forms a curved filament corresponding to part of a shallow helix with a 400 Å pitch (152) (Table 13b). The SAM domain of the transcription repressor TEL is another example. It can form a helical filament with a pitch of 53 Å and a diameter of 83 Å (153). In summary, these results suggest that the helical filaments are regularly observed in crystal structures and the geometry of the

FtsZ helical crystal filament reported here is within the range of other helical filaments characterized so far.

The dimeric interface

The dimeric interface of the *M. tuberculosis* FtsZ is formed with residues 160 to 170, 133 to 137 and 171 to 175 from one molecule and residues 270 to 280 from its neighboring molecule. As it forms part of the helical filament and may correspond with the curved fiber observed from EM, we will refer the dimer observed in our crystal structure as the curved dimer. The recent reported *M. jannaschii* FtsZ dimer is part of a semi-continuous straight filament (Table 13a) and thereby will be referred as the straight dimer. The difference between these two types of dimer can be visualized by superimposing one monomer of each dimer, as a large rotation (about 60°) can be observed between the second monomers (Figure 38a). Importantly, the T7 loop of the curved dimer (around residues 205 to 210) does not make contact with the nucleotide binding site of its neighboring molecule (Figure 38b), while the T7 loop of the straight dimer (around residues 232 to 237) does (Figure 38c). Actually, in the curved dimer, the C_α of the residue Asp 210 is about 20 Å away from the β-phosphorus of the bound GDP, while in the straight dimer, the distance between the C_α of the equivalent residue Asp 235 and the β-phosphorus of the bound GTP is about 11 Å, which demonstrates the exclusion of the T7 loop from the nucleotide binding pocket in the curved dimer. Biochemical data suggest that the T7 loop from one monomer should interact with the nucleotide binding pocket of its neighboring monomer, in order to form the intact GTPase active site (133). Crystallographic data in the straight FtsZ and tubulin dimers

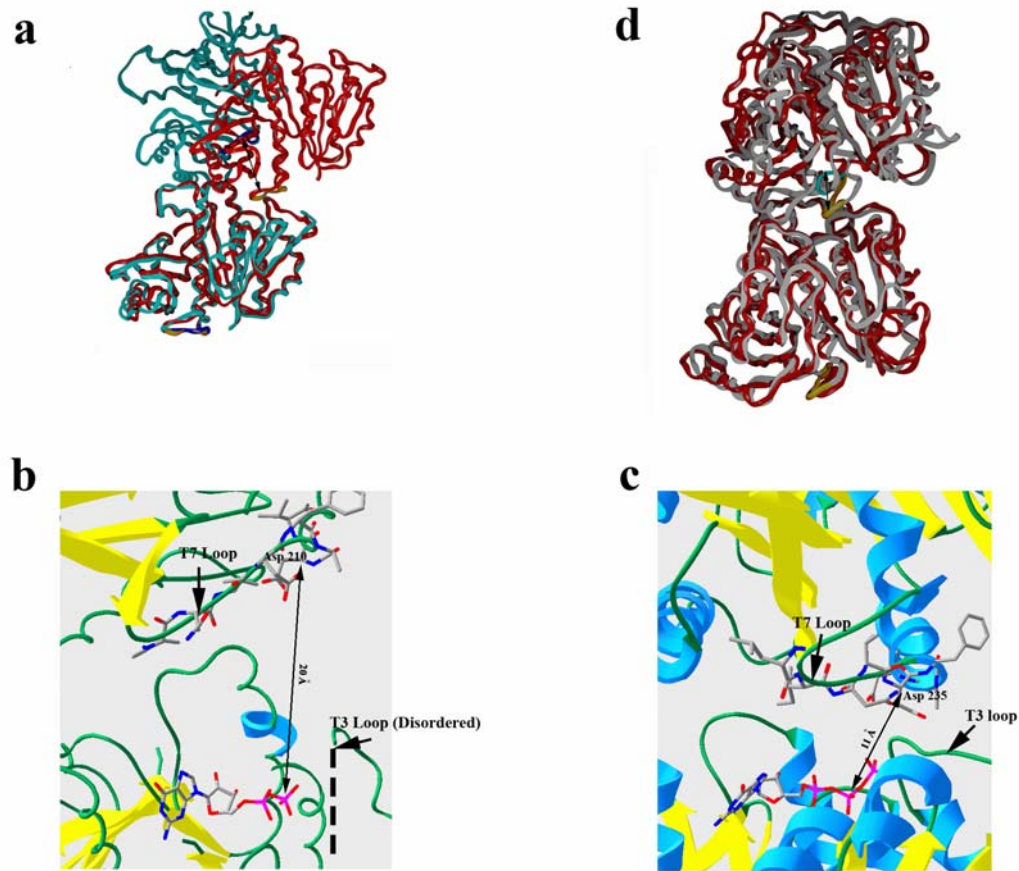


Figure 38 The Difference in the Dimeric Interface between the Curved and Straight Dimer.

a. Comparison of the Curved and Straight FtsZ Dimer

The curved (cyan) and straight (red) FtsZ dimers are aligned at the bottom monomer. The T7 loop is colored in blue and golden in each dimer, respectively. The displacement of the T7 loop is indicated by the arrow.

b. The detailed view of the dimeric interface within the curved FtsZ dimer

The distance between the C_{α} of the residue Asp 210 and the β -phosphorus of GDP is indicated. And the T3 (disordered, represented by black dashed line) and T7 loops are indicated by black arrow. The residues on the T7 loop are also displayed.

c. The detailed view of the dimeric interface within the straight FtsZ dimer

The distance between the C_{α} of the residue Asp 235 and the β -phosphorus of GDP is indicated. And the T3 and T7 loops are indicated by black arrows. The residues on the T7 loop are also displayed.

d. Comparison of the Curved and Straight Tubulin Dimer

The curved (red) and straight (white) tubulin dimer are aligned at the bottom monomer. The T7 loop is colored in cyan and golden in each dimer, respectively. The displacement of the T7 loop is indicated by the arrow.

support this mechanism (131). However, this is not present in our structure. Therefore, we propose that the helical filament we observed is in an “off” state, when the GTPase is inactive.

The GDP binding site

A GDP molecule is bound to the nucleotide binding pocket and its guanosine ring appears to be stacked with the aromatic side chain of Phe 180. The regions around the GDP binding site correspond to the worst electron density. As mentioned in earlier, many loops in this region do not have the well defined density. What is the cause of this problem? As published by Leung, et al. (130), this region is flexible and contains structural elements that are similar to the switch I and II in G-proteins whose conformation are dependent on the liganded state of the nucleotide. Particularly, the T3 loop (residues 60 to 70) is ordered only when the γ -phosphate group is present. Therefore, in the GDP bound protein, as shown from this study, this loop is disordered and cannot be found in the electron density. Other loops around this pocket might take different conformations from those in the model for molecular replacement, as it was originally derived from the GTP γ S bound FtsZ monomer. If we have data to a higher resolution, we might be able to see some of the densities in this region and have more clues on the nature of the conformational changes.

After calculating the solvent accessible surface, the nucleotide binding pocket is found to be completely accessible to the solvent (Figure 36b). In the published straight FtsZ dimer (131), the GTP is mostly buried and accessible to the solvent only via an opening with a diameter of around 10 Å (Figure 36c). Though the 10 Å² opening may

still be enough for nucleotide to diffuse into or out of the pocket (131), it is certainly more restricted. The biochemical studies to compare the accessibility of GTP and GDP bound filaments have not been reported. However, it was shown that that the helical shaped GDP bound filament was readily straightened after being washed with GTP (127), which indicated the GDP bound filament was fairly accessible to the solvent.

The possible mechanism to destabilize the FtsZ and tubulin filament

Tubulin, the eukaryotic homolog of FtsZ, also forms straight and curved dimers (152). The curved form is induced by the binding of GDP as well as a Stathmin-like protein, a microtubule de-stabilizer (in Table 13b, 1Z2B, 1SA0 and 1FFX). By superimposing one monomer of each dimer, we see a similar shift of the second monomer, though with a much less degree than that in FtsZ. Similarly, the T7 loop is “pushed” out in the curved tubulin dimer, though different from the curved FtsZ dimer, in the curved tubulin dimer the T7 loop is never completely out of the nucleotide binding pocket (Figure 38d).

The buried surface area of the curved tubulin dimer is about 1960 to 2200 Å², smaller than that of the straighter tubulin dimer (about 3000 Å²). In FtsZ, the buried area within the curved FtsZ dimer (1100 to 1350 Å²) is also significantly smaller than that of the straight FtsZ dimer (about 2344 Å²). The decrease of the buried surface area probably correlates with the destabilizing effect by Stathmin (in tubulin) and GDP (in tubulin and FtsZ).

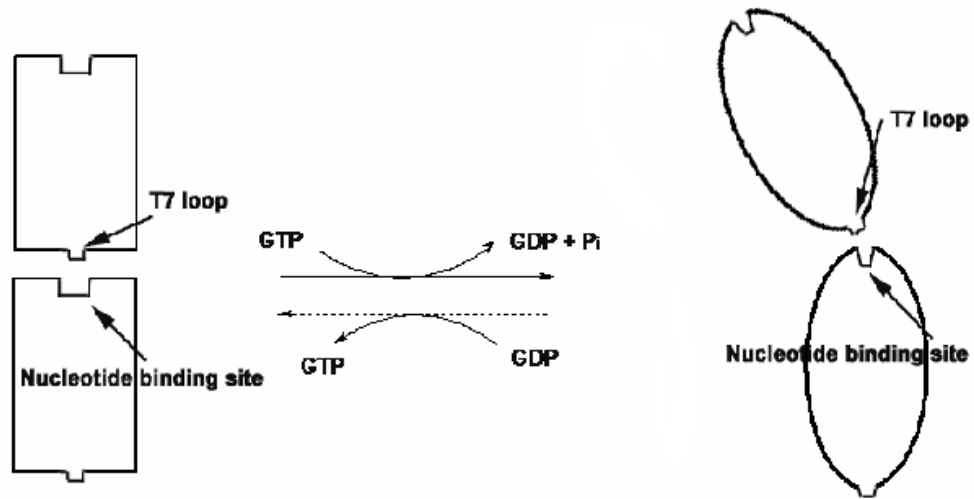


Figure 39 The Putative Mechanism for FtsZ Dynamics.

In the straight dimer of FtsZ, the T3 loop (around residues 93 to 98) is in close contact with the T7 loop (Figure 38c). In the curved dimer, the T3 loop (residues 60 to 70) is disordered (Figure 38b), as was reported earlier (130). Therefore, it is possible that the release of γ -phosphate group may trigger the conformation change of the T3 loop, which may lead to the reorganization of the protein-protein interface. For example, some dimeric interactions, such as those between the T3 and T7 loops, may be abolished. As the result, it might be energetically more stable for the individual FtsZ monomers to be re-oriented to form the helical filament (Figure 39).

As in contrast with the GTPase “on” state of the straight filament, we regard this helical FtsZ filament as GTPase “off”. What is the fate of this helical filament? As it forms under the presence of a destabilizing agent (GDP) and has a decreased buried

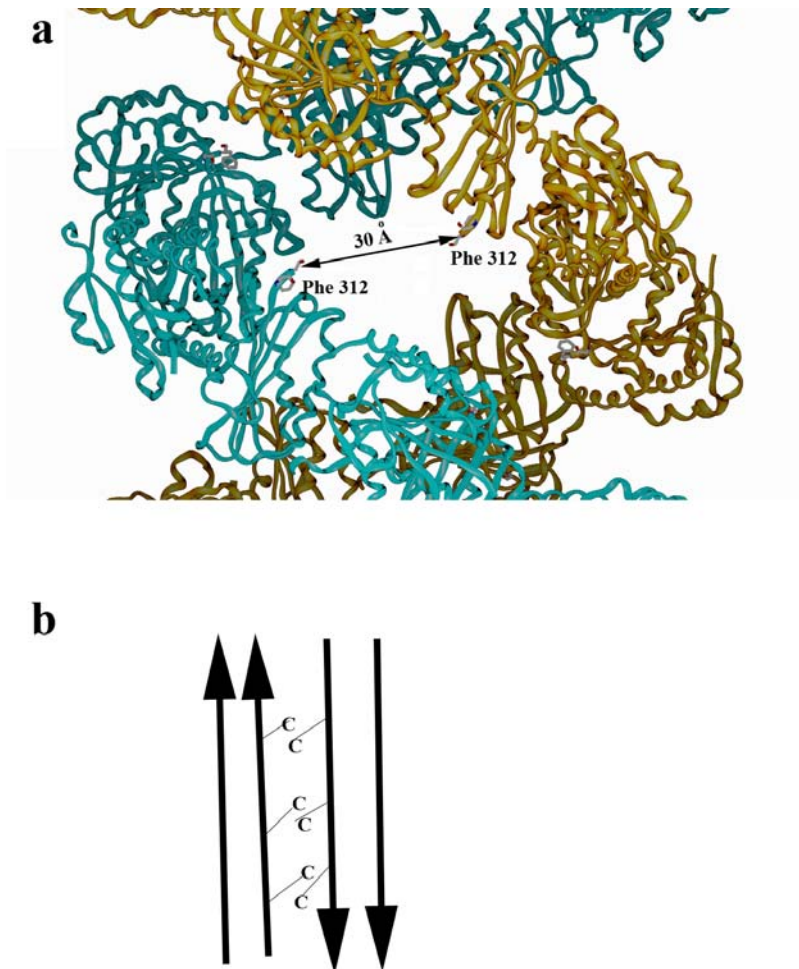


Figure 40 The Potential Role of the C-terminal Tail.

a. The Putative C-terminal Tail Location

The last residues of each monomer in neighboring strands are displayed. Their distance is labeled. The two neighboring strands are colored differently (cyan and golden, respectively).

b. A Model for Lateral Interaction

Proposed by Lowe et al. (125) based on EM observation of an FtsZ sheet, two parallel protofilaments form a thick filament. Thick filaments are arranged in anti-parallel. Though tighter contacts were proposed for interactions within a thick filament, the interactions between thick filaments are much less and were proposed to involve the C-terminal tail region (labeled as “C”).

surface area, the helical shaped FtsZ filament probably represents a less stable state. Therefore, it might eventually undergo disassembly.

The probable role of the C-terminus

Interestingly, the C-terminal residues (Gly 311 or Phe 312) on antiparallel strands of the neighboring protofilaments are pointing to each other with a distance of about 30 Å (Figure 40a). As mentioned earlier, the last 67 residues (residue 313 to 379) of FtsZ are disordered. It is reasonable to speculate that this part of the structure occupies the gap region between the neighboring strands and makes contacts. Actually, in the Ca²⁺ induced FtsZ sheet reported by Lowe, et al. (125), the C-terminal tail between two antiparallel thick filaments did interact (Figure 40b).

To investigate the role of the C-terminal tail, we tested the polymerization of the truncated protein FtsZ₁₋₃₁₂. It polymerized similarly to the full-length FtsZ, which indicated that the C-terminus of FtsZ was not directly involved in the polymerization. Visualized by EM, we found that the FtsZ₁₋₃₁₂ filaments showed more bundling (Figure 33c) than the filaments formed by the full-length protein. Furthermore, the individual FtsZ₁₋₃₁₂ filament did not have the feature of two dense regions separated by a clear gap, as was demonstrated from the full-length FtsZ filaments (Figure 33d).

Based on our data, we hypothesize that the unstructured C-terminal plays a role in modulating the lateral interaction between the FtsZ filaments. Removal of this region seems to promote bundling, probably because the truncated protein forms a more tight lateral interaction. In the cell, the C-terminal tail is involved in the binding of accessory proteins (e. g. ZipA and FtsA). Their binding usually triggers bundling of FtsZ filaments

and also stabilizes the Z-ring (126). It will be interesting to further investigate this part of the structure and their role in the filament interactions.

Conclusion

The polymerization and de-polymerization process is critical for the proper function of FtsZ. In this chapter, we find the *M. tuberculosis* FtsZ can form a left-handed double helical crystal filament. In contrast to the straight filament reported earlier, our findings show the catalytic T7 loop is not in the nucleotide binding loop, which suggests a GTPase inactive form. Furthermore, the helical filament we observed has a smaller buried surface area than the straight filament, which may indicate it is less stable. Therefore, we argue for a mechanism for FtsZ destabilization by GDP, in which the loss of the γ -phosphate group triggers conformational changes in the nucleotide binding pocket and then leads to the re-organization of the dimeric interface. The result is to form a GTPase “off” helical filament that is energetically less stable and may be the first step for the disassembly of the FtsZ polymer. We also describe the tentative role of the flexible C-terminus and its possible contribution in modulating the lateral interaction. Consistent with our crystal structure and EM observations reported by other groups, the C-terminal tails may make contacts between two antiparallel strands. However, the biological relevance of this type of interaction is unclear. To sum up, we think our discoveries will provide a testable hypothesis that may deepen the understanding of the FtsZ polymerization.

CHAPTER VI

GENERAL CONCLUSIONS AND FUTURE DIRECTIONS

Crystallography is a powerful technique to investigate a biological system. It provides a static atomic image of the macromolecule that can be the basis for further analysis. For complete understanding of the scientific problem, however, hypothesis made from the information gained by crystal structures should be verified by perturbing the system. Therefore, the combination of both crystal and solution methods will be the most suitable approach. In this dissertation, we applied both crystallographic and biochemical techniques onto MVK and FtsZ. Our results provided a detailed picture on how these two proteins function and are summarized as follows.

The work on MVK established the fold of MVK within the GHMP kinase family and the structural basis for the substrate binding, catalysis and inhibition of this enzyme. Our data revealed the important role of the phosphate-binding loop, in binding of nucleotide and terpenoid ligands. We also explored the nucleoside base binding pocket and both the selectivity and degeneracy of this pocket was demonstrated. Our studies on terpenoid inhibitors explained how these compounds compete with ATP. The potential mechanism on why these compounds are uncompetitive with mevalonate was proposed as mevalonate might facilitate the ordering of the carbon chain of terpenoids. Though we did not directly solve the location of mevalonate in the crystal structure, our analysis of water structures near the catalytic pair Asp 155-Lys8 revealed the putative location of

the phosphoryl acceptor and helped to model a mevalonate into the active site. Finally, the possible catalytic mechanism was proposed based on our structural data.

In order to further investigate the binding mechanism of MVK, as well as provide lead inhibitors for drugs design, we performed virtual screening on this protein. This study produced two compounds that showed limited inhibition to MVK. However, the whole designing process was far from a success. Their inhibition to MVK was low and no improvement could be achieved by optimization steps. Using our current active site, compounds are routinely docked to the phosphate-binding loop, which may causes weak inhibition, low specificity and poor scoring accuracy.

The work on FtsZ revealed a novel helical filament that was different from the straight filament reported by Lowe's group (131). This filament might correspond with the helical shaped filament of the *E. coli* FtsZ in the presence of GDP and DEAE dextran (127). Our findings also indicated that the catalytic T7 loop is not in the nucleotide binding pocket, in contrast with the straight filament. Therefore, we proposed that our observation might resemble a GTPase "off" state. GDP was demonstrated as a destabilizer of FtsZ polymers and the helical filament in our structure has less buried surface area than the reported straight filament. Therefore, we hypothesized that the conformational changes in the nucleotide binding pocket triggered by GTP hydrolysis might lead to the re-organization of the dimeric interface, which could result in the straight to helical transition in the overall filament and create a less stable helical fiber that may be the intermediate for further disassembly.

We also investigated the role of the FtsZ C-terminal tail. This region is unstructured in all available crystal structures of FtsZ and is involved in the interaction with many FtsZ accessory proteins. From our structure, we inferred that this region may be involved in the interaction between two anti-parallel strands. Removal of this tail still permitted polymerization, though more bundling was observed from filaments formed by the truncated FtsZ. The exact role of the C-terminal tail of FtsZ is still unclear. In this dissertation, we proposed that it might modulate the lateral interactions of neighboring FtsZ filaments during the assembly of the Z-ring.

Our work not only provided insights onto these interesting proteins, but also raised more questions. It is certain that more studies can be performed in the future to further the understanding in this area.

For MVK, we would like to solve the crystal structure with both mevalonate and terpenoid bound to it. As we discussed in Chapter III, with the presence of mevalonate, the carbon chain of the terpenoids might be ordered and any conformational changes, such as the active site pocket closure, might be observed. Currently, only the dioxane condition produces diffracting crystals. As dioxane is likely to compete with mevalonate, a crystallization condition independent of dioxane should be identified. We indeed have several conditions using PEG as the precipitant. However, they only produce thin needles that do not diffract. Additive screening and extensive optimization have been performed without much success. Certainly, more optimization or crystallization screening should be performed. Secondly, more virtual screening can be performed to identify better inhibitors. We currently have a grid based computing system that can be

used to screen one million compounds in a reasonable time frame. In addition to screen for larger database, we should also explore the binding site distinctive from the phosphate-binding loop, because of the problems already being discussed. If the terpenoid carbon chain binding site can be determined by crystallography, that region will provide an attractive target. Compounds designed to bind to this region will be more hydrophobic and do not need to compete with the phosphate groups directly, which may lead to the improvement in both inhibition and specificity.

Our analysis on the FtsZ helical filament raised the question regarding conformational changes at the nucleotide binding pocket during GTP hydrolysis. However, we could not identify many loops around this region, probably due to conformational differences between the model and the actual structure. The low resolution data have further complicated the matter. If the resolution can be improved significantly, either by performing experiments at a more intense beamline or by improving the crystal quality, we may be able to rebuild this region and understand any conformational variations between different nucleotide liganded states.

In Chapter IV, we proposed a model regarding the disassembly of FtsZ filaments based on our observation of the GDP bound helical filament. Our key argument is that the T7 loop is excluded from the nucleotide binding pocket upon the GTP hydrolysis and the γ -phosphate group release. In order to verify this model, it will be critical to seek biochemical evidence on whether the GTPase is active during the GDP bound state. Obviously, this study is difficult to perform directly. However, we may be able to monitor the interaction between the T7 loop with the nucleotide binding pocket of the

neighboring monomer during GTP hydrolysis. Fluorescence might be a useful tool for this investigation. To achieve this, we should first introduce a fluorophore (e.g. a tryptophan residue) into the T7 loop. We could expect to see an increase in both the anisotropy and quenching effect during the time course, as T7 loop is gradually “pushed” out of the nucleotide binding pocket and fully exposed to the solvent.

FtsZ filaments make lateral interactions. Recent studies suggest that the Z-ring is likely to be constructed by alignment of many short FtsZ filaments (142, 143). Therefore, investigations into the mechanism of lateral interactions are instrumental for understanding the architecture of the Z-ring. Though potential lateral interface has been proposed based on EM studies (125, 147), little supporting evidence is available. EM and mutagenesis may be applied to probe the role of the C-terminal domain and to map the interface between protofilaments. Spin label EPR and fluorescence resonance energy transfer may also be useful to provide distance information between residues on neighboring filaments.

Accessory proteins that bind to FtsZ are important for cytokinesis. Many such proteins have been studied, mostly from *E. coli*. Interestingly, in *M. tuberculosis*, only FtsW is found to be FtsZ associated. Future research can be focused on identifying, characterizing and crystallizing new FtsZ binding proteins from *M. tuberculosis*. This will not only advance our knowledge of the cell division machinery in mycobacterium, but also provide potential drug targets for the treatment of tuberculosis.

REFERENCES

1. Perutz, M. F. (1970) *Nature* **228**, 726-734
2. Bourgeois, D., and Royant, A. (2005) *Curr. Opin. Struct. Biol.* **15**, 538-547
3. Vrieland, A., and Sampson, N. (2003) *Curr. Opin. Struct. Biol.* **13**, 709-715
4. Schindler, T., Bornmann, W., Pellicena, P., Miller, T. W., Clarkson, B., and Kuriyan, J. (2000) *Science* **289**, 1938-1942
5. Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B. H., and Wlodawer, A. (1989) *Science* **246**, 1149-1152
6. Rozwarski, D. A., Grant, G. A., Barton, D. H. R., Jacobs Jr., W. R., and Sacchettini, J. C. (1998) *Science* **279**, 98-102
7. Dutta, S., and Berman, H. M. (2005) *Structure* **13**, 381-388
8. Rossmann, M. G., Morais, M. C., Leiman, P. G., and Zhang, W. (2005) *Structure* **13**, 355-362
9. Weber, P. C. (1997) *Methods Enzymol.* **276**, 13-22
10. Riès-kautt, M., and Ducruix, A. (1997) *Methods Enzymol.* **276**, 23-59
11. Luft, J. R., and Detitta, G. T. (1997) *Methods Enzymol.* **276**, 110-131
12. Stura, E. A., and Wilson, I. A. (1990) *Methods* **1**, 38-49
13. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307-326
14. Leslie, A. G. W. (1992) *Joint CCP4+ESF-EAMCB Newsletter on Protein Crystallogr.* **26** (www.mrc-lmb.cam.ac.uk/harry/mosflm/FAQ.html#reference)
15. Evans, P. R. (1993) *Proc. CCP4 Study Weekend.* **R34**, 114-122

16. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. D.* **55**, 849-861
17. Uson, I., and Sheldrick, G. M. (1999) *Curr. Opin. Struct. Biol.* **9**, 643-648
18. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr. D.* **58**, 1772-1779
19. Yao, J.-X. (1983) *Acta Crystallogr. A.* **39**, 35-37
20. de La Fortelle, E., and Bricogne, G. (1997) *Methods Enzymol.* **276**, 472-494
21. Navaza, J. (1994) *Acta Crystallogr. A.* **50**, 157-163
22. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Gross-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta. Crystallogr. D* **54**, 905-921
23. Vagin, A., and Teplyakov, A. (1997) *J. Appl. Crystallogr.* **30**, 1022-1025
24. Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999) *Acta. Crystallogr. D.* **55**, 484-491
25. Wang, B.-C. (1985) *Methods Enzymol.* **115**, 90-112
26. Cowtan, K. (1994) *Proc. CCP4 Study Weekend Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallogr.* **31**, 34-38.
27. Terwilliger, T. C. (1999) *Acta Crystallogr. D.* **55**, 1863-1871
28. Abrahams, J. P., and Leslie, A. G. W. (1996) *Acta Crystallogr. D.* **52**, 30-42
29. Bricogne, G. (1976) *Acta. Crystallogr. A.* **32**, 832-847
30. Reddy, V., Swanson, S., Sacchettini, J. C., Kantardjieff, K. A., Segelke, B., and Rupp, B. (2003) *Acta Crystallogr. D.* **59**, 2200-2210

31. Christopher, J. A. (1998) *SPOCK: The Structural Properties Observation and Calculation Kit. Program Manual.*, The Center for Macromolecular Design, Texas A&M University, College Station, TX
32. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47**, 110--119.
33. McRee, D. E. (1999) *J. Struct. Biol.* **125**, 156-165
34. Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M., and Jones, T. A. (2001) Chapter 17.1, 353-356, 366-367. In Rossmann, M. G. and Arnold, E. (eds.) *International Tables for Crystallography, F. (Crystallography of Biological Macromolecules)*, Kluwer Academic Publishers, Dordrecht, The Netherlands.
35. Kleywegt, G. J., and Jones, T. A. (1996) *Acta Crystallogr. D.* **52**, 826-828
36. Lamzin, V. S., and Wilson, K. S. (1993) *Acta Crystallogr. D.* **49**, 129-147
37. Ioerger, T. R., and Sacchettini, J. C. (2003) *Methods Enzymol.* **374**, 244-270.
38. Sussman, J. L. (1985) *Methods Enzymol.* **115**, 271-303
39. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D.* **53**, 240-255
40. Adams, D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) *Acta Crystallogr. D.* **58**, 1948-1954
41. Brunger, A. T. (1997) *Methods Enzymol.* **277**, 366-396
42. Lyne, P. D. (2002) *Drug Discov. Today* **7**, 1047-1055
43. Anderson, A. C. (2003) *Chem. Biol.* **10**, 787-797

44. Rarey, M., Kramer, B., Lengauer, T., and Klebe, G. (1996) *J. Mol. Biol.* **261**, 470-489
45. Schneider, G., and Fechner, U. (2005) *Nature Rev. Drug Discov.* **4**, 649-663
46. Rees, D. C., Congreve, M., Murray, C. W., and Carr, R. (2004) *Nature Rev. Drug Discov.* **3**, 660-672
47. Congreve, M., Murray, C. W., and Blundell, T. L. (2005) *Drug Discov. Today* **10**, 895-907
48. Sacchettini, J. C., and Poulter, C. D. (1997) *Science* **277**, 1788-1789
49. Magee, T., and Marshall, C. (1999) *Cell* **98**, 9-12
50. Williams, D. C., Carroll, B. J., Jin, Q., Rithner, C. D., Lenger, S. R., Floss, H. G., Coates, R. M., Williams, R. M., and Croteau, R. (2000) *Chem Biol.* **7**, 969-977
51. Avery, M. A., Alvim-Gaston, M., Vroman, J. A., Wu, B., Ager, A., Peters, W., Robinson, B. L., and Charman, W. (2002) *J. Med. Chem.* **45**, 4321-4335
52. Wilding, E. I., Brown, J. R., Bryant, A. P., Chalker, A. F., Holmes, D. J., Ingraham, K. A., Iordanescu, S., So, C. Y., Rosenberg, M., and Gwynn, M. N. (2000) *J. Bacteriol.* **182**, 4319-4327
53. Eisenreich, W., Schwarz, M., Cartayrade, A., Arigoni, D., Zenk, M. H., and Bacher, A. (1998) *Chem Biol.* **5**, R221-R233
54. Istvan, E. S. (2001) *Curr. Opin. Struct. Biol.* **11**, 746-751
55. Istvan, E. S., and Deisenhofer, J. (2001) *Science* **292**, 1160-1164
56. Goldstein, J. L., and Brown, M. S. (1990) *Nature* **343**, 425-430

57. Dhe-Paganon, S., Magrath, J., and Abeles, R. H. (1994) *Biochemistry* **33**, 13355-13362
58. Mueller, C., Schwender, J., Zeidler, J., and Lichtenthaler, H. K. (2000) *Biochem. Soc. Trans.* **28**, 792-793
59. Steinbacher, S., Kaiser, J., Eisenreich, W., Huber, R., Bacher, A., and Rohdich, F. (2003) *J. Biol. Chem.* **278**, 18401-18407
60. Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., Turbachova, I., Eberl, M., Zeidler, J., Lichtenthaler, H. K., Soldati, D., and Beck, E. (1999) *Science* **285**, 1573-1576
61. Wada, T., Kuzuyama, T., Satoh, S., Kuramitsu, S., Yokoyama, S., Unzai, S., Tame, J. R., and Park, S. Y. (2003) *J. Biol. Chem.* **278**, 30022-30027
62. Richard, S. B., Bowman, M. E., Kwiatkowski, W., Kang, I., Chow, C., Lillo, A. M., Cane, D. E., and Noel, J. P. (2001) *Nat. Struct. Biol.* **8**, 641-648
63. Richard, S. B., Ferrer, J-L., Bowman, M. E., Lillo, A. M., Tetzlaff, C. N., Cane, D. E., and Noel, J. P. (2002) *J. Biol. Chem.* **277**, 8667-8672
64. Rohdich, F., Hecht, S., Gartner, K., Adam, P., Krieger, C., Amslinger, S., Arigoni, D., Bacher, A., and Eisenreich, W. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 1158-1163
65. Liang, P.-H., Ko, T.-P., and Wang, A. H.-J. (2002) *Eur. J. Biochem.* **269**, 3339-3354
66. Tarshis, L. C., Proteau, P. J., Kellogg, B. A., Sacchettini, J. C., and Poulter, C. D. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 15018-15023

67. Fujihashi, M., Zhang, Y.-W., Higuchi, Y., Li, X.-Y., Koyama, T., and Miki, K. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 4337-4342
68. Kellogg, B. A., and Poulter, C. D. (1997) *Curr. Opin. Chem. Biol.* **1**, 570-578
69. Hirooka, K., Ohnuma, S.-i., Koike-Takeshita, A., Koyama, T., and Nishino, T. (2000) *Eur. J. Biochem.* **267**, 4520-4528
70. Park, H. W., Boduluri, S. R., Moomaw, J. F., Casey, P. J., and Beese, L. S. (1997) *Science* **275**, 1800-1804
71. Gelb, M. H., Scholten, J. D., and Sebolt-Leopold, J. S. (1998) *Curr. Opin. Chem. Biol.* **2**, 40-48
72. Greenhagen, B., and Chappell, J. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 13479-13481
73. Starks, C. M., Back, K., Chappell, J., and Noel, J. (1997) *Science* **277**, 1815-1820
74. Wendt, K. U., Schulz, G. E., Corey, E. J., and Liu, D. R. (2000) *Angew. Chem. Int. Ed.* **39**, 2812-2833
75. Wendt, K. U., Lenhart, A., and Schulz, G. E. (1999) *J. Mol. Biol.* **286**, 175-187
76. Huang, K.-x., Scott, A. I., and Bennet, G. N. (1999) *Protein Expr. Purif.* **17**, 33-40
77. Tanaka, R. D., Schafer, B. L., Lee, L. Y., Freudenberger, J. S., and Mosley, S. T. (1990) *J. Biol. Chem.* **265**, 2391-2398
78. Hedl, M., and Rodwell, V. W. (2004) *Protein Sci.* **13**, 687-693
79. Voynova, N. E., Rios, S. E., and Mizioroko, H. M. (2004) *J. Bacteriol.* **186**, 61-67
80. Beytia, E., Dorsey, K., Marr, J., Cleland, W. W., and Porter, J. W. (1970) *J. Biol. Chem.* **245**, 5450-5458

81. Dorsey, J. K., and Porter, J. W. (1968) *J. Biol. Chem.* **243**, 4667-4670
82. Levy, H. R., and Popjak, G. (1960) *Biochem. J.* **75**, 417-428
83. Hinson, D. D., Chambliss, K. L., Toth, M. J., Tanaka, R. D., and Gibson, K. M. (1997) *J. Lipid Res.* **38**, 2216-2223
84. Andreassi, II, J. L., Dabovic, K., and Leyh, T. S. (2004) *Biochemistry* **43**, 16461-16466
85. Knowles, J. R. (1980) *Ann. Rev. Biochem.* **49**, 877-919
86. Matte, A., Tari, L. W., and Delbaere, L. T. J. (1998) *Structure* **6**, 413-419
87. Tari, L. W., Matte, A., Pugazhenthii, U., Goldie, H., and Delbaere, L. T. J. (1996) *Nat. Struct. Biol.* **3**, 355-363
88. Cheek, S., Zhang, H., and Grishin, N. V. (2002) *J. Mol. Biol.* **320**, 855-881
89. Potter, D., and Mizioroko, H. M. (1997) *J. Biol. Chem.* **272**, 25449-25454
90. Cho, Y.-K., Ríos, S. E., Kim, J.-J. P., and Mizioroko, H. M. (2001) *J. Biol. Chem.* **276**, 12573 – 12578
91. Houten, S. M., Wander, R. J. A., and Waterham, H. R. (2000) *Biochim. Biophys. Acta.* **1529**, 19-32
92. Biardi, L., and Krisans, S. K. (1996) *J. Biol. Chem.* **271**, 1784-1788
93. Valle, D. (1999) *Nat. Genet.* **22**, 121-122
94. Hinson, D. D., Ross, R. M., Krisans, S., Shaw, J. L., Kozich, V., Rolland, M.-O., Divry, P., Mancini, J., Hoffmann, G. F., and Gibson, K. M. (1999) *Am. J. Hum. Genet.* **65**, 327-335

95. Houten, S. M., Romeijn, G. J., Koster, J., Gray, R. G. F., Darbyshire, P., Smit, G. P. A., de Klerk, J. B. C., Duran, M., Gibson, K. M., Wander, R. J. A., and Waterham, H. R. (1999) *Hum. Mol. Genet.* **8**, 1523-1528
96. Fu, Z., Wang, M., Potter, D., Mizioroko, H. M., and Kim, J.-J. P. (2002) *J. Biol. Chem.* **277**, 18134-18142
97. Matthews, B. W. (1968) *J. Mol. Biol.* **33**, 491-497
98. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283-291
99. Hinson, D. D., Chambliss, K. L., Hoffmann, G. F., Krisans, S., Keller, R. K., and Gibson, K. M. (1997) *J. Biol. Chem.* **272**, 26756-26760
100. Potter, D., Wojnar, J. M., Narasimhan, C., and Mizioroko, H. M. (1996) *J. Biol. Chem.* **272**, 5741-5746
101. Zhou, T., Daugherty, M., Grishin, N. V., Osterman, A. L., and Zhang, H. (2000) *Structure* **8**, 1247-1257
102. Yang, D., Shipman, L. W., Roessner, C. A., Scott, A. I., and Sacchettini, J. C. (2002) *J. Biol. Chem.* **277**, 9462-9467
103. Holm, L., Sander, C. (1995) *Trends Biochem. Sci.* **20**, 478-480
104. Lu, G., (1996) *Protein Data Bank Q. Newslett.* **78**, 10-11
105. Chu, X., and Li, D. (2003) *Protein Expr. Purif.* **32**, 75-82
106. Krishna, S. S., Zhou, T., Daugherty, M., Osterman, A., and Zhang, H. (2001) *Biochemistry* **40**, 10810-10818
107. Huo, X., and Viola, R. E. (1996) *Biochemistry* **35**, 16180-16185

108. Moroz, O. V., Harkiolaki, M., Galperin, M. Y., Vagin, A. A., Gonzalez-Pacanowska, D., and Wilson, K. S. (2004) *J. Mol. Biol.* **342**, 1583-1597
109. Gordon, G. S., and Wright, A. (2000) *Annu. Rev. Microbiol.* **54**, 681-708
110. Yu, X.-C., Tran, A. H., Sun, Q., and Margolin, W. (1998) *J. Bacteriol.* **180**, 1296-1304
111. Aussel, L., Barre, F.-X., Aroyo, M., Stasiak, A., Stasiak, A. Z., and Sherratt, D. (2002) *Cell* **108**, 195-205
112. Donachie, W. D. (2002) *Mol. Cell* **9**, 206-207
113. Lockhart, A., and Kendrick-Jones, J. (1998) *FEBS Lett.* **430**, 278-282
114. Melby, T. E., Ciampaglio, C. N., Briscoe, G., and Erickson, H. P. (1998) *J. Cell Biol.* **142**, 1595-1604
115. Hirano, T. (1999) *Genes Dev.* **13**, 11-19
116. Wu, L. J., and Errington, J. (2004) *Cell* **117**, 915-925
117. Bernhardt, T. G., and de Boer, P. A. J. (2005) *Mol. Cell* **18**, 555-564
118. Errington, J., Daniel, R. A., and Scheffer, D.-J. (2003) *Mol. Biol. Rev.* **67**, 52-65
119. Lutkenhaus, J., and Addinall, S. G. (1997) *Annu. Rev. Biochem.* **66**, 93-116
120. Pastoret, S., Fraipont, C., den Blaauwen, T., Wolf, B., Aarsman, M. E. G., Piette, A., Thomas, A., Brasseur, R., and Nguyen-Disteche, M. (2004) *J. Bacteriol.* **186**, 8370-8379
121. Lowe, J., and Amos, L. A. (1998) *Nature* **391**, 203-206
122. Romberg, L., and Levin, P. A. (2003) *Annu. Rev. Microbiol.* **57**, 125-154

123. Scheffers, D. J., de Wit, J. G., den Blaauwen, T., and Driessen, A. J. M. (2000) *Mol. Microbiol.* **35**, 1211-1219
124. Erickson, H. P., Taylor, D. W., Taylor, K. A., and Bramhill, D. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 519-523
125. Lowe, J., and Amos, L. A. (1999) *EMBO J.* **18**, 2364-2371
126. Hale, C. A., Rhee, A. C., and de Boer, P. A. J. (2000) *J. Bacteriol.* **182**, 5153-5166
127. Lu, C., Reedy, M., and Erickson, H. P. (2000) *J. Bacteriol.* **182**, 164-170
128. Lowe, J., and Amos, L. A. (2000) *Biol. Chem.* **381**, 993-999
129. Desai, A., and Mitchison, T. J. (1997) *Annu. Rev. Cell Dev. Biol.* **13**, 83-117
130. Leung, A. K. W., White, E. L., Ross, L. J., Reynolds, R. C., DeVito, J. A., and Borhani, D. W. (2004) *J. Mol. Biol.* **342**, 953-970
131. Oliva, M. A., Cordell, S. C., and Lowe, J. (2004) *Nat. Struct. Mol. Biol.* **11**, 1243-1250
132. Cordell, S. C., Robinson, E. J. H., and Lowe, J. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 7889-7894
133. Scheffers, D.-J., de Wit, J. G., den Blaauwen, T., and Driessen, A. J. M. (2002) *Biochemistry* **41**, 521-529
134. Scheffzek, K., Ahmadian, M. R., and Wittinghofer, A. (1999) *Trends Biochem. Sci.* **23**, 257-262
135. White, E. L., Ross, L. J., Reynolds, R. C., Seitz, L. E., Moore, G. D., and Borhani, D. W. (2000) *J. Bacteriol.* **182**, 4028-4034

136. Mukherjee, A., and Lutkenhaus, J. (1999) *J. Bacteriol.* **181**, 823-832
137. Romberg, L., and Mitchison, T. J. (2004) *Biochemistry* **43**, 282-288
138. Dai, K., and Lutkenhaus, J. (1992) *J. Bacteriol.* **174**, 6145-6151
139. Mosyak, L., Zhang, Y., Glasfeld, E., Haney, S., Stahl, M., Seehra, J., and Somers, W. S. (2000) *EMBO J.* **19**, 3179-3191
140. Low, H. H., Moncrieffe, M. C., and Lowe, J. (2004) *J. Mol. Biol.* **341**, 839-852
141. Datta, P., Dasgupta, A., Bhakta, S., and Basu, J. (2002) *J. Biol. Chem.* **277**, 24983-24987
142. Scheffers, D. J. (2001) *Polymerization of the Bacterial Cell Division Protein FtsZ*, Ph.D dissertation, University of Groningen, The Netherlands
143. Stricker, J., Maddox, P., Salmon, E. D., and Erickson, H. P. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 3171-3175
144. Erickson, H. P. (1997) *Trends Cell Biol.* **7**, 362-367
145. Leung, A. K. W., White, E. L., Ross, L. J., and Borhani, D. W. (2000) *Acta Crystallogr. D.* **56**, 1634-1637
146. Boggon, T. J., and Shapiro, L. (2000) *Structure* **8**, R143-R149
147. Oliva, M. A., Huecas, S., Palacios, J. M., Martin-Benito, J., Valpuesta, J. M., and Andreu, J. M. (2003) *J. Biol. Chem.* **278**, 33562-33570
148. Lu, C., Stricker, J., and Erickson, H. P. (2001) *BMC Microbiol.* **1**: 7
149. Cowtan, K. D., and Main, P. (1993) *Acta. Crystallogr. D.* **49**, 148-157
150. Chen, B., Vogan, E. M., Gong, H., Skehel, J. J., Wiley, D., and Harrison, S. C. (2005) *Structure* **13**, 197-211

151. Xing, X., and Bell, C. E. (2004) *J. Mol. Biol.* **342**, 1471-1485
152. Gigant, B., Curmi, P. A., Martin-Barbey, C., Charbaut, E., Lachkar, S., Lebeau, L., Siavoshian, S., Sobel, A., and Knossow, M. (2000) *Cell* **102**, 809-816
153. Qiao, F., Song, H., Kim, C. A., Sawaya, M. R., Hunter, J. B., Gingery, M., Rebay, I., Courey, A. J., and Bowie, J. U. (1994) *Cell* **118**, 163-173

VITA

Personal Information

Name: Dong Yang

Address: Department of Biochemistry and Biophysics, 2128 TAMU, College Station,
TX 77843-2128, USA

Email Address: bearofthecity@gmail.com

Education

2006: Doctor of Philosophy, Biochemistry, Texas A&M University,

1998: Bachelor of Science, Biology, Zhe Jiang University, China

Publications

1. Yang, D., Shipman, L. W., Roessner, C. A., Scott, A. I., and Sacchettini, J. C. (2002)
J. Biol. Chem. **277**, 9462 - 9467