STRUCTURAL STUDIES OF BACTERIOPHAGE LYSINS AND THEIR IMPLICATION IN HUMAN DISEASES

A Dissertation

by

QINGAN SUN

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

May 2011

Major Subject: Biochemistry

STRUCTURAL STUDIES OF BACTERIOPHAGE LYSINS

AND THEIR IMPLICATION IN HUMAN DISEASES

A Dissertation

by

QINGAN SUN

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,	Ryland Young
	Andreas Holzenburg
	Siegfried Musser
Head of Department,	Gregory Reinhart

May 2011

Major Subject: Biochemistry

ABSTRACT

Structural Studies of Bacteriophage Lysins and Their Implication in Human Diseases.

(May 2011)

Qingan Sun, B.S., Fudan University; M.S., Fudan University

Chair of Advisory Committee: Dr. James C. Sacchettini

Structural biology lays the molecular foundation for the modern field of life sciences. In this thesis, X-ray crystallography is the primary resource for atomic detail structural information and is the major technology employed in our research. Three examples show how structural biology addresses the basic processes of life. Firstly, two crystal structures of R^{21}, corresponding to two biological states, reveal a new activation mechanism of SAR(Signal-Anchor-Release)-endolysin, which not only complements the previous model, but is also more generally applicable to the endolysin family. It turned out that the SAR domain in R^{21} not only modulates the topology of the protein but also becomes an integral part of the active enzyme. This role is brand-new to the current knowledge on the SAR domain. The structural information was further corroborated by NMR(Nuclear Magnetic Resonance) data in solution. The second example is the crystal structure of mycobacteriophage lysin B. Combined with further biochemistry and genetics study, it identified the function of the Lysin B in phage D29 as a novel mycolylarabinogalactan esterase. This finding tackles the long-standing question of how mycobacteriophage circumvent the mycolic acid-rich outer membrane of mycobacterium

to release the phage progeny, and provides potential treatment to mycobactaria-related diseases, like tuberculosis. The last example is the homology modeling of the *Plasmodium* ribosomal L4 protein. In the modeled structure, azithromycin binds to the L4 protein, thus will block protein systhesis of the apicoplast ribosome. However, the model of the mutant L4 from azithromycin-reisistant strain is in conflict with the bound azithromycin. The action mode for the drug in *Plasmodium* was proposed based on that, which accounts for the anti-malaria effect of azithromycin.

DEDICATION

All my thoughts are dedicated to the last *Fish* men*. At the cross of the Pacific Ocean and the Yangtze River, it is the country of the *Fish* men. They were the Carthaginian in the Oriental; they were the warriors on the lands; they were the mighty over the sea. When the glory is fading in the history, they are, still, the poets of life.

* *Fish* is pronounced as *Ng* in the native language.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Sacchettini, and my committee members, Dr. Young, Dr. Holzenburg, and Dr. Musser, for their guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues in the Sacchettini lab for making my time at Texas A&M University a great experience. I also want to extend my gratitude to all the collaborators who shared their expertise throughout my research. Besides the collaborations with Young lab, Hatfull lab, and Fidock lab which were published in papers and pointed out in Chapters II, III, and IV, I would thank to Dr. Savva from the Holzenburg lab for the Electron Microscopy work on *M. tuberculosis* FAS I; Meng Chen from the Musser lab for previous collaboration on SufI; also Dr. Dangott for protein analysis, and Dr. Pai for mass spectrometry work. I would like to thank Tracey Musa for her extensive review of the manuscript.

Finally, thanks to my mother and father for their encouragement and to my wife for her patience and love.

NOMENCLATURE

AZ	Azithromycin
CQ	Chloroquine
DV	Digestive Vacuole
EM	Electron Microscopy
HSQC	Heteronuclear Single Quantum Coherence
LSU	Large Subunit (of Ribosome)
NMR	Nuclear Magnetic Resonance
SAR	Signal-Anchor-Release Domain
SP	Signal Peptide
XFEL	X-ray Free Electron Laser

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION: STRUCTURAL BIOLOGY LAYS THE FOUNDATION FOR LIFE SCIENCES

What is life? This question is often asked by humanity in both a philosophical and scientific context. We are living in a biosphere, in balance with various animals, fungi, and plant life, baffled and sometimes killed by pathogens. It is not only scientific curiosity but also mundane imperative to understand the nature of life and how it interacts and evolves.

However, it is impractical, if not impossible, to provide a short, final, comprehensive definition of life although we inherently perceive the distinct features that qualify living beings. One key quality of life is its evasion of the second law of thermodynamics, which dictates that the entropy of a closed system constantly maximizes (Landau and Lifshits, 1969) . In contrast, the biological world seems capable of the opposite, increasing its organization over generations. In his book *What is Life*, Schrödinger ascribed this to the living organisms's consumption of so-called "negative entropy" (Schrödinger, 1992). By using information as a form of negative-entropy, life is able to transform, accumulate, propagate, and eventually perpetuate itself by managing the flow of matter and energy through living forms.

This dissertation follows the style of Cell.

The key to the mystery of the life lies in the biomolecules, or "aperiodic crystals" as Schrödinger dubbed them (Schrödinger, 1992), which are made largely of the common elements: carbon, nitrogen, oxygen, phosphorous, and hydrogen. The unique features of biomolecules do not arise from a uniqueness in their atomic composition, but from their distinct 3-D structures, i.e., the spatial distribution of the atoms and the inter-atomic connectivity and interactions. This falls in the field of structural biology, which lays the foundation for understanding the science of life.

X-ray crystallography is the primary resource of structural biology

There are three common techniques in structure biology used to visualize macromolecules at the atomic level, these are X-ray crystallography, Nuclear Magnetic Resonance (NMR), and Electron Microscopy (EM). X-ray crystallography is by far the most common of the three methods. The latter two methods are limited by the macromolecule's size, lower resolution, or both. X-ray crystallography has been used to solve the eukaryotic ribosome complex which is more than 3 MDa, and can reach a resolution of better than 1 Å. As of January 2011, there were 61266 searchable structures in the RCSB protein data bank (PDB) solved by X-ray, while NMR and EM contributed 8739 and 340 structures, respectively (http://www.rcsb.org). EM and NMR have their unique power of course, e.g., NMR can reveal the valuable information about structure and dynamics in solution as a supplementary to x-ray crystallography, which will be briefly demonstrated in Chapter II. Below, I will focus on how X-ray crystallography works.

X-ray diffraction is the magnifier of molecular electron structure

X-rays cannot be magnified through a lens system as in optical microscopy or EM. However, atomic spatial information can still be detected at the macroscopic level through the interference of the scattered coherent X-rays by the electrons in the molecule. This phenomenon is called diffraction, as demonstrated in **Figure 1-1** (Feynman et al., 2010). In X-ray diffraction experiments, most X-rays pass through the empty space within the atom, but when a single X-ray photon encounters an electron, the waves are scattered. X-ray waves scattered by two identical electrons have the same amplitude while their phases are determined by the positions of the electrons, they form an interference pattern on the detector screen. Since the distances to the light source and the detector (L) are infinitively large compared to the spread of the scatters (d) in the molecule, the distance between the peaks of the diffraction pattern (D) can be approximately given by

$$\frac{D}{L} = \frac{\lambda}{d} \tag{1.1}$$

As illustrated in **Fig. 1-1B**, the X-ray wavelength (λ) is comparable to d, and the distance between the diffraction peaks (D) is a on the same scale of the detector-sample distance (L). The diffraction pattern is a macroscopic observation, from which the electron distribution in the molecule can be deduced at the atomic level.

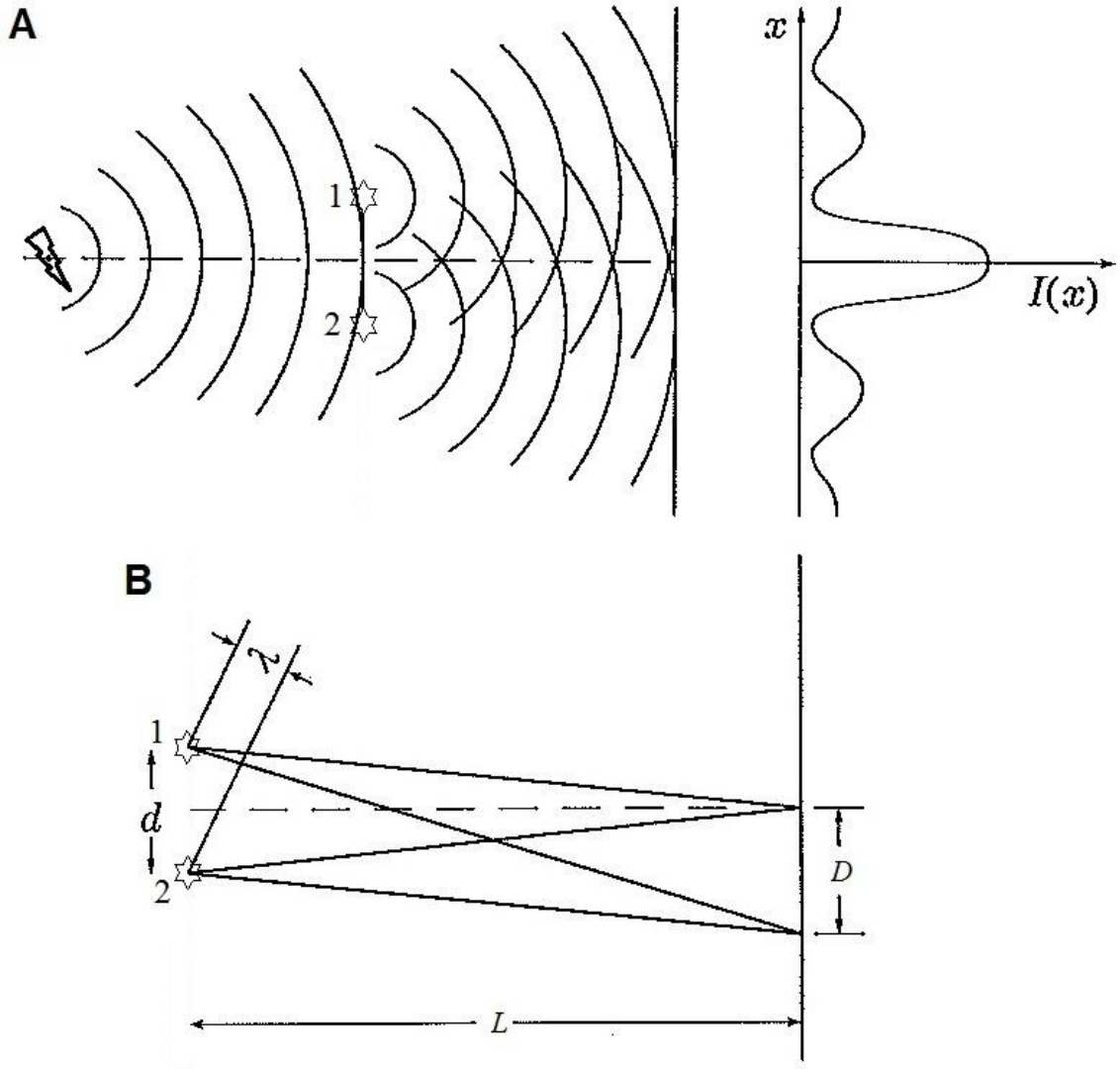


Figure 1-1. Scheme of X-ray diffraction in 2D. A. The analogous experiment in wave interference. The waves arrive at x starting from source (lightening symbol) and scattered by two electrons (represented by the stars). At some points x, the wavelets from 1 and 2 interfere destructively (e.g., a crest from electron 1 arrives at the same time as a trough from electron 2); and others, constructively. This produces the complicated minima and maxima of the intensity curve I(x). **B. Peak position on the distant detector screen**. Two beams of light, starting in phase at electrons 1 and 2, will interfere constructively when they reach the screen B if they take the same time to travel from C to B. This means that a maximum in the interference pattern for light beams passing through two holes will occur at the center of the screen. As we move down the screen, the next maximum will occur at a distance *D*, which is far enough from the center that, in traveling to this point, the beam from electron 1 will have traveled exactly one wavelength λ farther than the beam from electron 2. This figure was modified based upon Feynman's *Quantum Mechanics and Path Integral*. Although in his book Feynman discussed the electron wave going through the holes, the scheme of wave interference holds for x-ray in our case (Feynman et al., 2010).

The qualitative description of the diffraction pattern in a plane by the simple equation 1.1 is based on the approximation that $D \ll L$, which does not always hold true. Generalizing this scenario to 3-dimensions and assuming the scattering power of the electrons are equal, the path-length the X-ray travels through different electrons is what matters. Since the light-source and the detector are both far away from the sample compared to the molecule size, the X-ray beams, both the incident and the scattered, can be viewed as plane-waves (**Fig.1-2**). Electron 2 is located at $\boldsymbol{r}$ relative to electron 1. $\boldsymbol{k}_0$ and $\boldsymbol{k}$ are the unit vectors indicating the propagation directions of the incident and scattered beams, respectively. The angle between $\boldsymbol{k}$ and $\boldsymbol{k}_0$ is defined as 2θ by convention. The difference between the path-length of X-ray scattered by electrons 1 and 2 is:

$$\Delta l = \boldsymbol{k} \cdot \boldsymbol{r} - \boldsymbol{k}_0 \cdot \boldsymbol{r} \tag{1.2}$$

The major elements of biomolecules are carbon, nitrogen, oxygen, phosphorous, and hydrogen, whose energy levels are far below the photon energy (hν, h is Planck's constant, ν is the frequency of the photon) of the X-rays employed in the experiment. Thus the electrons in the molecules can be treated as free charged particles that scatter the X-rays in the Thomson way, i.e., the wavelength (λ) remains unchanged during the scattering process. Here the wave vector, $\boldsymbol{s}$, is introduced, according to its typical usage in crystallography: its magnitude is equal to $1/\lambda$, and its direction is the direction of wave propagation. The phase difference between different electron scattering events will be

$$\varphi = \frac{2\pi \cdot \Delta l}{\lambda} = 2\pi(\boldsymbol{s} \cdot \boldsymbol{r} - \boldsymbol{s}_0 \cdot \boldsymbol{r}) \tag{1.3}$$

Define $\boldsymbol{S} = \boldsymbol{s} - \boldsymbol{s}_0$, then equation 1.3 will be simplified as

$$\varphi = 2\pi \boldsymbol{S} \cdot \boldsymbol{r} \tag{1.4}$$

where $\boldsymbol{S}$ is the change of the wave vector during the scattering, its magnitude is

$$|\boldsymbol{S}| = 2|\boldsymbol{s}| \sin\theta \tag{1.5}$$

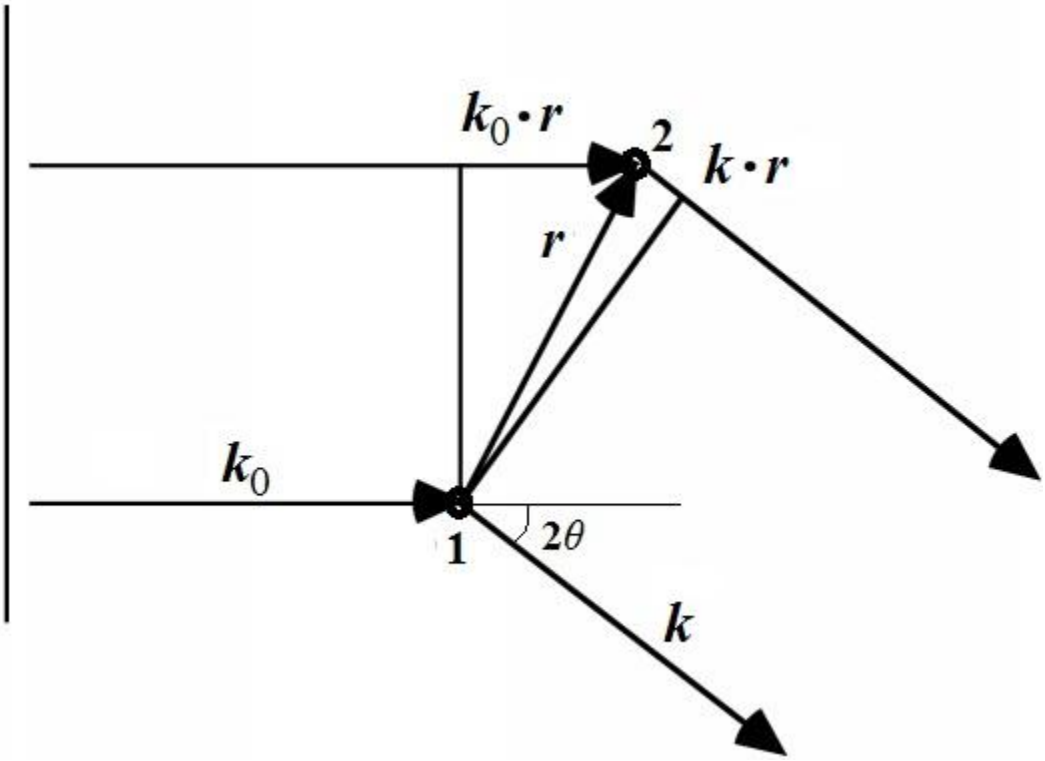


Figure 1-2. The geometry of plane-wave scattered by two electrons. Electron 1 is placed at the origin, while electron 2 is at position $\boldsymbol{r}$; the incident beam's direction is indicated by unit vector $\boldsymbol{k}_0$, which is scattered to direction $\boldsymbol{k}$, the deflected angle is 2 θ.

The intensity of the scattered beam, I, for the scattered beam $\boldsymbol{s}$ will be

$$I(\boldsymbol{s}) = \left|1 + e^{i\varphi}\right|^2$$

I will reach a maximum when $\varphi = 2\pi \boldsymbol{S} \cdot \boldsymbol{r} = 2n\pi$, where n is an integer.

Take d as the projection of $\boldsymbol{r}$ on $\boldsymbol{S}$, and recall that $|\boldsymbol{S}| = 2|\boldsymbol{s}| \sin\theta = 2\sin\theta/\lambda$. Then it can be deduced that

$$2d \sin\theta = n\lambda \tag{1.6}$$

Equation 1.6 shares the same mathematical form and physical insight with the famous Bragg's law, although d in the latter defines the translational symmetry in the crystal lattice.

When there are N electrons in the biomolecule, the intensity will be

$$I(\boldsymbol{S}) = \left|\sum_{j=1}^{N} e^{i\varphi_j}\right|^2 = \left|\sum_{j=1}^{N} e^{i2\pi \boldsymbol{S}\cdot\boldsymbol{r}_j}\right|^2 = |F(\boldsymbol{S})|^2 \tag{1.7}$$

Here the origin is moved away from electron 1, since the origin is arbitrary without any symmetrical restriction. Also the structure factor $F(\boldsymbol{S})$ is introduced, which is the alternative of $I(\boldsymbol{S})$ to describe diffraction. It should be noted that since $F(\boldsymbol{S})$ is the summation of a series of complex numbers, it will generally be complex.

In reality, the electrons are continuously distributed in the molecule, which is described by the electron density function $\rho(\boldsymbol{r})$. The scattering power at a certain position $\boldsymbol{r}$ is therefore proportional to $\rho(\boldsymbol{r})$ at this point. Thus, the summation in equation 1.7 is transformed into the integral:

$$F(\boldsymbol{S}) = \int \rho(\boldsymbol{r}) e^{i2\pi \boldsymbol{S}\cdot\boldsymbol{r}} d\boldsymbol{r} \tag{1.8}$$

This is in the form of a Fourier transformation, the inverse Fourier transform will retrieve the electron density $\rho(\boldsymbol{r})$ from the structure factor $F(\boldsymbol{S})$:

$$\rho(\boldsymbol{r}) = \int F(\boldsymbol{S}) e^{-i2\pi \boldsymbol{S}\cdot\boldsymbol{r}} d\boldsymbol{S} \tag{1.9}$$

So far we have derived the relationship between the electron density function $\rho(\boldsymbol{r})$ and the structure factor $F(\boldsymbol{S})$. Note that the momentum of the X-ray photon $\boldsymbol{p} = h\boldsymbol{s}$, $\rho(\boldsymbol{r})$

and $F(\boldsymbol{S})$ are actually the position and momentum representations of the photon respectively. It is well known in quantum mechanics that the position and momentum operators do not commute and cannot be measured simultaneously. Thus $F(\boldsymbol{S})$ at a certain diffraction direction $\boldsymbol{S}$ cannot be deduced from $\rho(\boldsymbol{r})$ at a certain position $\boldsymbol{r}$, and *vice versa*. We need to measure the full set of $F(\boldsymbol{S})$ in momentum space to get the quantum state implicitly, and then expand it into the position space to get $\rho(\boldsymbol{r})$. In the opposite direction, it also requires integration of $\rho(\boldsymbol{r})$ over all space (in reality, just the molecules involved in scattering) to get $F(\boldsymbol{S})$.

The crystalline structure serves as the amplifier and sampler

We have a macroscopic observation - X-ray diffraction - carrying the spatial information of the target molecule at the atomic level. However, the diffraction from a single biomolecule (or complex) is too faint to measure with current X-ray source technology due to the weak interaction between the X-ray photons and atoms. On the other hand, stronger radiation would destroy the structural information well before it could be collected (Neutze et al., 2000).[1] Crystallographers overcome this dilemma by amplifying the signal with X- ray diffraction from a crystal composed of billions of biomolecules.

[1] The development of the X-ray free-electron laser may some day solve atomic structures from the diffraction of a single biomolecule. However, it is still a concept under investigation (Gaffney and Chapman, 2007; Neutze et al., 2000).

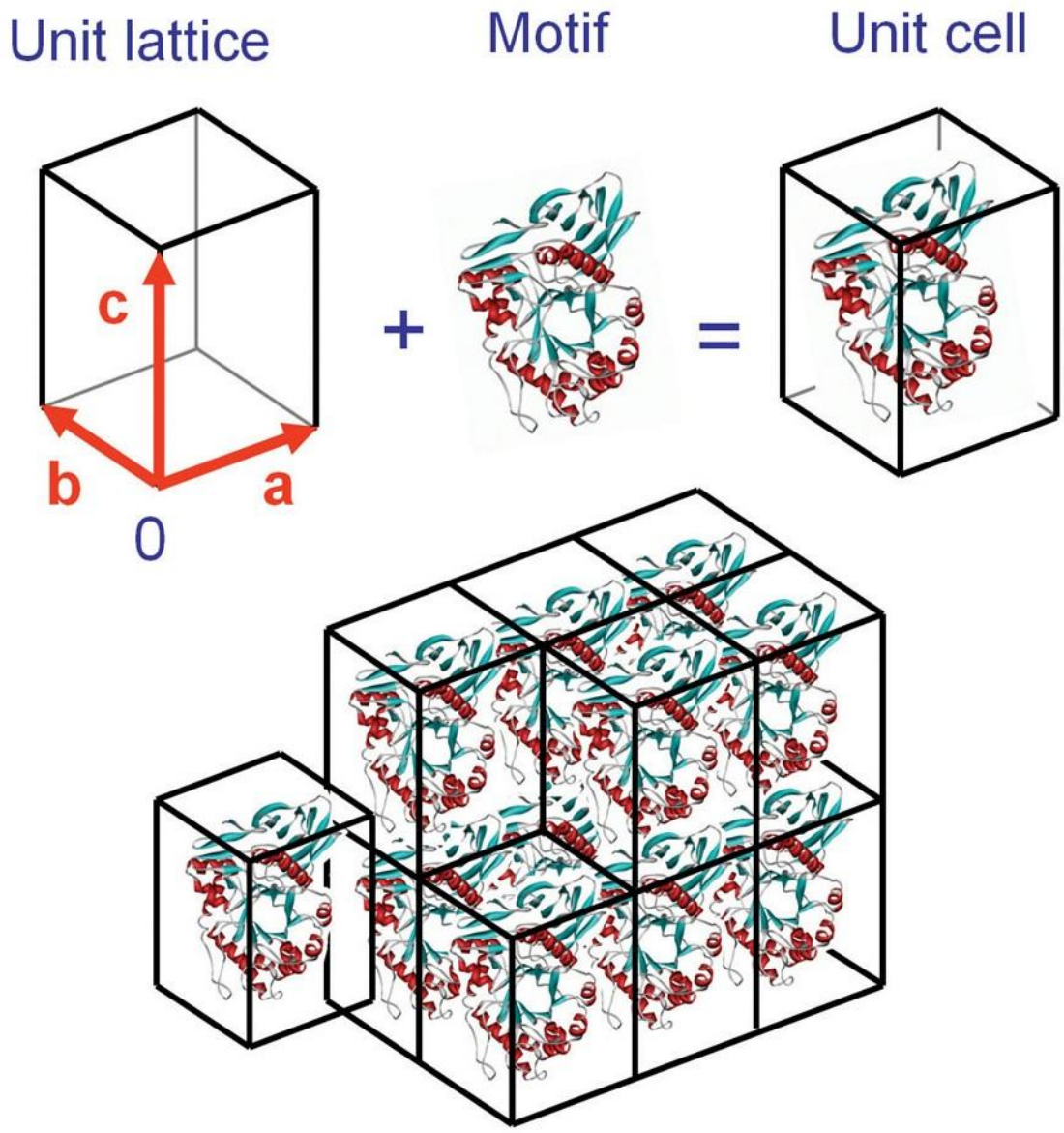


Figure 1-3. Assembly of a primitive triclinic 3-D crystal from unit cells. The unit lattice is filled with a motif, and the crystal is built from translationally stacked unit cells. The basis vectors form a right-handed system [O, **a**, **b**, **c**]. © 2010 From Biomolecular Crystallography by Bernhard Rupp. Reproduced by permission of Garland Science/Taylor & Francis LLC. (Rupp, 2010)

Crystals are a periodic assembly of identical objects such as atoms or molecules as shown in **Figure 1-3** (Rupp, 2010). The basis vectors **a**, **b**, and **c** define the unit cell, as well as the translational symmetry of the crystal, which means in the crystal $\rho(\boldsymbol{r}) = \rho(\boldsymbol{r} + t\mathbf{a} + u\mathbf{b} + v\mathbf{c}) = \rho(\boldsymbol{r}) \otimes \delta(\boldsymbol{r} - t\mathbf{a} - u\mathbf{b} - v\mathbf{c})$, where t, u, and v are integers. A single unit cell usually contains one to several biomolecules, usually with a scale of 10 nm. The size of a protein crystal is typically in the sub-millimeter scale. t, u, and v can be as large as thousands.

If we know the electron density distribution $\rho(\boldsymbol{r'})$ in one unit cell (eq. 1.8), the diffraction pattern from this unit cell will be the Fourier transformation

$$F'(\boldsymbol{S}) = \text{FT}\ (\rho(\boldsymbol{r'}))$$

Assume the dimensions of the crystal are l**a** × m**b** × n**c**, the diffraction from the whole crystal will be

$$\begin{aligned} F(\boldsymbol{S}) &= FT\left(\textstyle\sum_{t,u,v} \rho(\boldsymbol{r'}) \otimes \delta(\boldsymbol{r'} - t\mathbf{a} - u\mathbf{b} - v\mathbf{c})\right) \\ &= F'(\boldsymbol{S}) \times \textstyle\sum_{t=0}^{\mathrm{l}} e^{i2\pi t\boldsymbol{S}\cdot\mathbf{a}} \times \sum_{u=0}^{\mathrm{m}} e^{i2\pi u\boldsymbol{S}\cdot\mathbf{b}} \times \sum_{v=0}^{\mathrm{n}} e^{i2\pi v\boldsymbol{S}\cdot\mathbf{c}} \end{aligned} \tag{1.10}$$

Let's examine the first summation term, $\sum_{t=0}^{\mathrm{l}} e^{i2\pi t\boldsymbol{S}\cdot\mathbf{a}}$. Here $e^{i2\pi t\boldsymbol{S}\cdot\mathbf{a}}$ is a series of unit vectors on the complex plane as shown in **Figure 1-4** (Drenth and Mesters, 2007). Since t is the integer covering a wide range from 0 to more than thousands, $e^{i2\pi t\boldsymbol{S}\cdot\mathbf{a}}$ will distributed evenly around the circle (Fig. 1-4), and the expected value

$$< e^{i2\pi t\boldsymbol{S}\cdot\mathbf{a}} > = 0,$$

unless $\boldsymbol{S} \cdot \mathbf{a} = h$, where h is an integer.

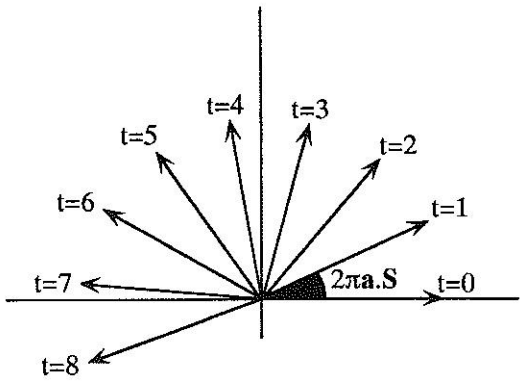


Figure 1-4. Contribution to the diffraction by a series of unit cells in the crystal. Each arrow represents the scattering by one unit cell in the crystal. Because of the huge number of unit cells and because their scattering vectors are pointing in different directions, the scattering by a crystal is, in general, zero. However, in the special case that $\boldsymbol{S} \cdot \mathbf{a}$ is an integer h, all vectors point to the right and the scattering by the crystal can be of appreciable intensity. This figure was reproduced with kind permission of Springer Science and Business Media (Drenth and Mesters, 2007).

The same conclusion can be drawn from the next two summations. Now $F(\boldsymbol{S})$ is no longer continuously distributed over the $\boldsymbol{S}$ space, but has non-zero values at $\boldsymbol{S}$ satisfying what is called the Laue conditions,

$$\begin{aligned} \boldsymbol{S} \cdot \mathbf{a} &= h \\ \boldsymbol{S} \cdot \mathbf{b} &= k \\ \boldsymbol{S} \cdot \mathbf{c} &= l \end{aligned} \tag{1.11}$$

where $F(\boldsymbol{S}) = F'(\boldsymbol{S}) \times \mathrm{l} \times \mathrm{m} \times \mathrm{n} = F'(\boldsymbol{S}) \times \mathrm{N}$, and N is the total number of unit cells in the crystal, which is at the order of 10^8. Recall that the intensity, I, is the square of the structure factor F (Eq. 1.7), the signal is astronomically amplified.

Define a new set of basis vectors, $\mathbf{a}^*$, $\mathbf{b}^*$, and $\mathbf{c}^*$ in the $\boldsymbol{S}$ space (reciprocal space in the terminology of crystallography).

$$\mathbf{a}^* = \frac{\mathbf{b} \times \mathbf{c}}{\mathbf{V}}$$

$$\mathbf{b}^* = \frac{\mathbf{c} \times \mathbf{a}}{\mathbf{V}}$$

$$\mathbf{c}^* = \frac{\mathbf{a} \times \mathbf{b}}{\mathbf{V}}$$

where V is the volume of unit cell in the position space (real space), and can be calculated as $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$.

$F(\boldsymbol{S})$ is a discrete function at evenly distributed points: $\boldsymbol{S}_{h,k,l} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$. Equation 1.9 for the electron density will be reduced to the summation

$$\rho(\boldsymbol{r}) = \sum_{h,k,l} F(\boldsymbol{S}_{h,k,l}) e^{-i2\pi \boldsymbol{S}_{h,k,l} \cdot \boldsymbol{r}} = \sum_{h,k,l} F_{h,k,l} e^{-i2\pi (h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*) \cdot \boldsymbol{r}} \qquad (1.12)$$

Thus, the crystal serves as a sampler of $F(\boldsymbol{S})$ in the reciprocal space without the loss of structural information.

Crystallography is the compiler of structural information

In the previous section we demonstrated that X-ray diffraction carries structural information from the crystal (and the molecules inside) (**Eq. 1.8, 1.9, and 1.12**). Equation 1.8 shows that the structure factor $F(\boldsymbol{S})$ is a complex number in general, which can be written $|F(\boldsymbol{S})| \cdot e^{i\varphi}$ ($|F(\boldsymbol{S})|$ is the amplitude; φ is the phase). Both the amplitude and the phase in the structure factor are required to reconstruct the electron density distribution in equations 1.9 and 1.12. However, in a normal X-ray diffraction experiment, only the intensity of the diffraction can be measured, which is the square of $|F(\boldsymbol{S})|$. The phase information in the entire set of diffraction data has been lost. This

poses the so-called phase problem to the crystallographer, which is the major theoretical difficulty when translating from diffraction to structure.

A little bit of calculation can clarify the severity of the problem. Since φ is completely unknown, it can take any value from 0 to 2π, and the probability density function $p(\varphi)$ will be just $1/2\pi$. Integrating out the unknown phase φ in $F(\boldsymbol{S})$, we get the expected structure factor

$$\langle F(\boldsymbol{S}) \rangle = \frac{1}{2\pi}\int_0^{2\pi} |F(\boldsymbol{S})| \cdot e^{i\varphi}\, \mathrm{d}\varphi = 0 \tag{1.13}$$

And the expected electron density will be

$$\langle \rho(\boldsymbol{r}) \rangle = FT^{-1}(\langle F(\boldsymbol{S}) \rangle) = 0 \tag{1.14}$$

This means that the measurements of diffraction intensity (amplitudes) alone do not contribute any structural information without phase.

We can also examine the phase problem in terms of sampling, and get a clue of how to solve it. In equations 1.9 and 1.12, there is a continuous electron density $\rho(\boldsymbol{r})$ on one side, and the discrete structure factor $F(\boldsymbol{S}_{h,k,l})$ on the other side. However, we can only measure the diffraction to a certain resolution d in the real experiments, i.e., $|\boldsymbol{S}|$ has a maximum of $1/d$. In the following paragraph, we will turn to an artificial 1-D crystal with periodicity A to simplify the question without loss of generality.

According to the Shannon sampling theorem, we can sample $\rho(\boldsymbol{r})$ with the interval $d/2$ without loss of information contained in $F(\boldsymbol{S})$ (Jerri, 1977). Now we can sample $\rho(\boldsymbol{r}_i)$ at a series of discrete positions, $r_i = id/2$, in the unit cell, i=1, 2, …, n ($n = 2A/d$). On the other side, we have $F(\boldsymbol{S}_j)$ at $S = j/A$, $j = \pm 1, \pm 2, \ldots, \pm n/2$.

Without considering the complication of anomalous scattering, electron density is real, which leads to the relation $F(-\boldsymbol{S}) = F^*(\boldsymbol{S})$ from equation 1.9. Thus $|F(-\boldsymbol{S})|$ is equal to $|F(\boldsymbol{S})|$ (Friedel's law). This leads to a set of $n/2$ equations

$$\left|F(S_j)\right| = \left|\sum_{i=1}^{n} \rho(r_i) e^{i2\pi r_i \cdot S_j}\right|, \qquad \left(j = 0, 1, 2, \ldots, \frac{n}{2} - 1\right) \tag{1.15}$$

Here, the total number of unknown variables, $\rho(r_i)$, is n. Solving these equations will determine the crystal structure, as well as solve the phase problem. However, the equation set is underdetermined by a factor of 2. This conclusion also holds in 2D or 3D crystals with or without considering anomalous scattering. In this view, the loss of phases in crystallography is the price paid for amplification of the structural information at the atomic level through the sampling effect on the diffraction from the periodic crystal.

Phase information from a priori *knowledge*

Since the phase problem can be viewed as the underdetermination of equation set 1.15, the solution is clear: either decrease the number of unknown variables or feed in new equations, so that the number of equations exceeds the number of variables to determine. Both require more information through additional experiments or prior knowledge, i.e. knowledge about the molecule, or more accurately, about the unit cell.

The first input is the atomicity in the crystal. The unit cell of the crystal is made up of known atoms and the electron density in each unit cell is the

summation of the electron density of the atoms in it, i.e., $\rho(\boldsymbol{r}) = \sum \rho_{atom}(\boldsymbol{r})$. Each atom's contribution to the diffraction can be described by six parameters: the atomic scattering factor, f; the three coordinates, x, y, and z; the occupancy, o; and the temperature factor, B. Thus, equation set 1.15 can be re-parameterized as

$$\left|F(\boldsymbol{S}_j)\right| = \left|\sum_{i}^{N} o_i \cdot f_i \cdot e^{i2\pi \boldsymbol{S}_j \cdot (x_i + y_i + z_i)} \cdot e^{-B_i \left|\boldsymbol{S}_j\right|^2 / 4}\right| \tag{1.16}$$

where N is the total number of scattering atoms in each unit cell. Normally the abundant hydrogen atoms are not considered since their contribution to diffraction is negligible. The term f_i is the scattering factor of the ith atom in the unit cell, which is determined by the scattering power of this type of atom. The coordinates, x_i, y_i, and z_i, define the average position of the ith atom in each unit cell. Of course, the position of each atom is not exactly the same in different unit cells, due to slightly different configurations or thermal vibration. Therefore, there is some uncertainty related to the position, which can be depicted as a Gaussian distribution around the average coordinates with variance σ^2. B_i is a reflection of the positional uncertainty in the reciprocal space. B is equal to $8\pi^2\sigma^2$.

Now we have $6N$ unknown variables, while the equation number increases cubically with the improvement of the diffraction resolution, d, i.e., it is proportional to $\left(\frac{1}{d}\right)^3$. At a certain point, the number of equations will exceed the number of variables to determine, equation 1.16 will be overdetermined, and the phase problem can be solved.

The other benefit of reformalizing equation 1.15 into 1.16 with atomic parameters is to introduce stereochemical knowledge, i.e., new equations about the bond-lengths, bond-angles, or other information between the atoms. This information will dramatically increase the overdeterminacy. Furthermore, stereochemical knowledge is derived *a priori* from the experiments, like small-molecule crystallography, with much higher accuracy. This is why the accuracy of the protein crystal structure can be much higher than the diffraction limit.

When we look at the statistics of the crystal structures in Chapters II and III, all of them have been solved with a diffraction resolution of around 2 Å. All have larger reflection numbers, i.e., numbers of $\left|F(\boldsymbol{S}_j)\right|$ measured, than six times the atom number. For example for the $^{i}R^{21}$ crystal in Chapter II, there are 35715 reflections at a resolution of 1.7 Å for 2483 atoms, corresponding to 14898 parameters to determine. The ratio is 2.4. Based upon this, we can estimate that equation 1.16 may be solved in theory when the resolution of diffraction data is beyond 2.3 Å.

In reality, diffraction data beyond 1.2 Å, the namely "atomic resolution" in literatures, is required to determine the crystal structure *ab initio* (the so called direct method) (Dauter, 2003; Uson and Sheldrick, 1999). There are two routes. The first is through the Patterson function, i.e., the inverse Fourier transformation of the intensity (Patterson, 1934),

$$P(\boldsymbol{r}) = FT^{-1}\big(I(\boldsymbol{S})\big) = FT^{-1}(F(\boldsymbol{S}) \cdot F^{*}(\boldsymbol{S})) = \rho(\boldsymbol{r}) \otimes \rho(-\boldsymbol{r}) \qquad (1.17)$$

Equation 1.17 is also termed the self-correlation function of the electron density $\rho(\boldsymbol{r})$. Combined with the atomicity assumption, the peaks in the Patterson function correspond

to the interatomic vectors, from which the position of the atoms can be deduced. The other route is to exploit the probabilistic relationship between the phases of different reflections, which is the results of the atomicity and positivity in reciprocal space (Karle, 1986). Although the direct method is a powerful tool in small molecule crystallography or in finding substructure in the macromolecule crystal, there are few successful examples of using the direct method to solve a whole protein structure with thousands of atoms (Mooers and Mattews, 2006). In the practice of protein crystallography, there are two major methods to get the initial phases, molecular replacement and experimental phasing.

A brief view of molecular replacement

Molecular replacement is based on the assumption that proteins (or protein motifs) with similar sequences share similar 3D structures. We call the proteins in the crystal under investigation the target, and choose putatively similar proteins with known structures as the probes. If the probes are put into the same orientations and positions as the targets in the real crystal, the mock crystal can diffract X-rays in a similar way which can be simulated by equation 1.16. Then the calculated $|F_c|$ can be compared with $|F_o|$, the observation from the target protein crystal. If the measurements are consistent, the calculated phase can substitute for the unknown real phase, and combined with $|F_o|$, can yield the initial structure of the target protein which can be further refined.

The key is to assemble the probes in the crystal correctly (the right molecular replacement solution). Given each probe with six freedoms, three angles for orientation

and three coordinates for position, there will be 6*N* parameters total for *N* probes to search (*N* is just 1 to several). This is a small number compared to the thousands of reflections. However, for each degree of freedom there are dozens to hundreds of choices, making a tremendous number of possibilities to search by multiplying them together. Although it is feasible to search the six parameters of each probe simultaneously, the more conventional way is to divide the procedure into orientation search and position search, then conquer the two problems separately. This can be done in either Patterson space or reciprocal space (Drenth and Mesters, 2007).

To search in Patterson space, the Patterson peaks from each probe can be calculated, which correspond to the intramolecular interatomic vectors. When the probe is put in the target crystal and turned around, the calculated Patterson peaks will rotate correspondingly. The intramolecular vectors are shorter than the intermolecular ones in general, and restricted in the sphere with the radius around the probe's diameter. So, we can focus on the Patterson map from the diffraction of the real crystal, particularly in that sphere, and rotate the probe for the best match between the two sets of Patterson peaks where the orientation of the probe is in the correct solution. Once the orientations of the probes have been determined, they can be translated around in the crystal. The calculated Patterson map from this mock crystal includes the intermolecular vectors among these peaks and their symmetry-related pairs, in addition to the intramolecular vector. The intermolecular vector will change with the translation of the probes in the unit cell until the best match is found compared with the Patterson map from the diffraction of the real crystal. Now both the orientations and the positions of the probes

are set in the unit cell. Of course, the probe's structure is not identical to that of the target protein, otherwise there would be no need for structure determination. The difference between the structures is most pronounced in the details, leading to more errors in the higher resolution data. This is why reflections beyond 3 Å are often excluded when applying this method (Drenth and Mesters, 2007; Rupp, 2010).

The two-step procedure can also be implemented in reciprocal space with the maximum likelihood method. The primary question in this method: what is the likelihood of a certain structural model, given the diffraction data? The best model should have the maximum-likelihood. However, this likelihood is difficult to quantify. According to Bayes' theorem, it is equal to the likelihood of the experimental observation given the certain model multiplied by the likelihood of the model itself. The likelihood of the model is evaluated by the consistency with the *a priori* knowledge, e.g. stereochemistry. In molecular replacement, the structures have similar stereochemistry, since they are generated by moving the same probes in the crystal. Hence the likelihood of the model is constant. The only variable term left is the likelihood of the data given the model. It can be derived from the Gaussian distribution of the structure factor, $p(F_o)$, which is an application of the central limit theorem. The mean F_o is equal to F_c calculated from the model; missing parts from the incomplete model contributes to the variance together with the errors from the model and the measurement. However, only the amplitude of the structure factor, $|F_o|$, is experimentally observed. To get the likelihood of observing a certain $|F_o|$, the phase-dependent probability, $p(\alpha)$, needs be integrated out.

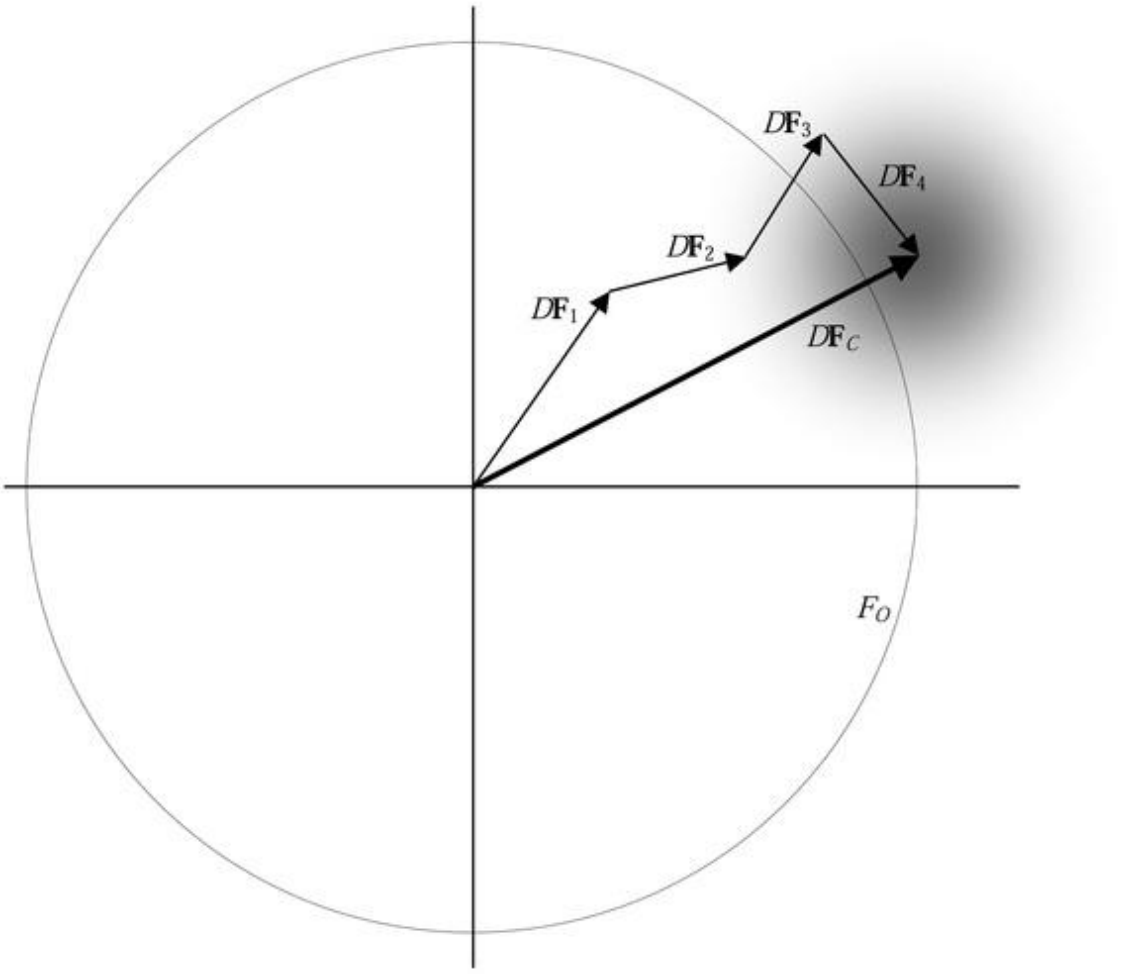


Figure 1-5. Schematic illustration of translation likelihood function for acentric structure factors. As a molecule is translated, the molecular-transform contributions from the symmetry-related copies (four in this example) will change in phase but not in amplitude. For the correct translation, the true structure factor will be found within a two-dimensional Gaussian distribution (shown in grey shading) centered on the total calculated structure factor, scaled by the factor D to obtain the centroid of the distribution. The contribution of a single structure factor to the likelihood function is obtained by integrating around a circle with a radius given by the observed amplitude, F_O, so the likelihood will be high when this circle intersects regions of high probability in the two-dimensional Gaussian. For a combined rotation/translation search, both the amplitudes and phases of the molecular-transform contributions will vary. Courtesy from *ICUr* (Read, 2001).

$$p(|F_o|) = \int_0^{2\pi} p(F_o) \cdot p(\alpha) d\alpha = \frac{1}{2\pi} \int_0^{2\pi} p(F_o) d\alpha \tag{1.18}$$

If there is no other source of phase information, $p(\alpha)$ is treated as a constant as shown in equation 1.18. This scenario is demonstrated in the likelihood function for the translation or six-dimensional search (Fig. 1-5) (Read, 2001).

In the rotation search, the probe positions are not specified. Only the amplitude contribution from each probe can be calculated. The expected $\langle F_c \rangle$ will be zero as in equation 1.13 (or the calculated structure factor from a partial solution if there is one). Based on this, the rotation likelihood function can also be derived from equation 1.18. In the rotational and translational searches, the probes are rotated and translated in the unit cell, respectively. The F_c changes accordingly, as the likelihood function does, until the likelihood reaches a maximum when the probes are in the best orientation and position given the diffraction data.

The advantage of the maximum likelihood approach is three-fold. First, the errors, both in the experimental measurements and in model building, are estimated and treated explicitly. Unlike in the Patterson method, it is not necessary to artificially truncate the reflection data to medium resolution because the errors have already been taken into account. Second, it is natural to incorporate *a priori* knowledge as part of the likelihood of the model. If new restrictions on the model are independent of the existing ones, they can be introduced into the likelihood function by multiplication as a new probability term. Third, it is easy to combine the phase information from another source. For example, if some phase distribution is derived from the experimental phasing as described below, the information can contribute to a new likelihood function in equation 1.18 as the $p(\alpha)$ term.

Experimental phasing with additional information from the substructure

The phases can also be derived from multiple measurements, either on the derivative crystals in the same lattice form (isomorphous replacement), or on the same crystal but at different wavelength for the dispersive effect of the anomalous scattering; or even from a single measurement which discerns the nuanced broken symmetry between the Friedel's pair due to the anomalous scatters in the crystal (Feil, 2002). In this way, since the phase problem appears solved solely from experiments, it is called experimental phasing.

Usually, there are only a handful of special atoms that can serve as major contributors to the differences in the measurements. For example, in isomorphous replacement, heavy atoms, like Hg, can be introduced into the crystal without changing the lattice too much. Then, the difference between the diffractions from the heavy atom derivative crystal and those from the native protein crystals is mainly ascribed to the substructure composed of heavy atoms only. If there is anomalous scattering (e.g. the selenium in the selenomethionine preparation) in the crystal, there will be measurable differences between the Friedel's pairs at certain wavelengths. In both cases, the substructure of the special atoms, i.e. the calculated F_s, accounts for the differences in the measurements, ΔF_o. For isomorphous replacement,

$$\Delta F_o = F_{o2} - F_{o1} = F_s \tag{1.19}$$

Or in the case of anomalous scattering,

$$\Delta F_o = {F_o}^+ - {F_o}^- = F_A \tag{1.20}$$

where F_o^+ and F_o^- are the Friedel's pairs. F_A is the anomalous contribution from the substructure, its amplitude is proportional to $|F_s|$, its phase is 90º more than that of F_s. Further, it can be proved that the differences between the measured F_o, is statistically equivalent to the amplitudes of F_s (Drenth and Mesters, 2007).

$$\langle \Delta^2 |F_o| \rangle = c \langle |F_s|^2 \rangle \tag{1.19}$$

where c is a constant in equation 1.19, e.g. it equals ½ in the isomorphous replacement.

The crystallographer can divide-and-conquer. The substructure can be solved with a measured difference $\Delta|F_o|$ first, using the direct method described earlier. The phased F_s can be calculated from the solved substructure. The relation between F_o and F_s in isomorphous replacement is illustrated by the Harker diagram as shown in Figure 1-6. In the complex plane, draw two circles with the radius of $|F_{o1}|$ and $|F_{o2}|$, centered at the origin and the end of vector F_s, respectively. The two intersections of these circles will be the common end points of the vectors F_{o1} and F_{o2}. Thus, they stand for two possible solutions, and both satisfy the relation of equation 1.19. In the case of the anomalous scattering, the F_s is replaced by F_A, the anomalous contribution in F_s, which is a small part of the substructure diffraction with an additional 90º in the phases. The ambiguity of the phase solution in the Harker diagram can be resolved by either multiple measurements on different derivative crystals or at multiple X-ray wavelengths in the case of anomalous scattering. Otherwise, *a priori* knowledge about the crystal composition, like solvent content, can be exploited to resolve this ambiguity later at the density modification and model building stage.

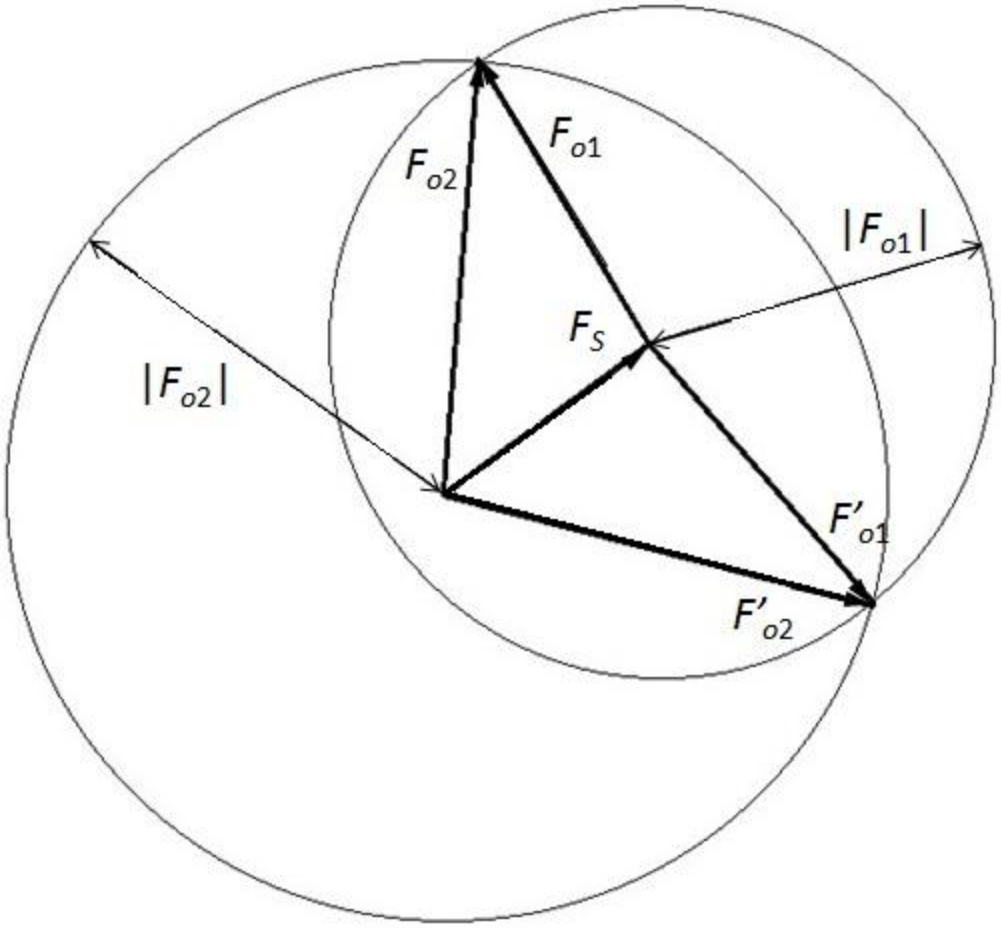


Figure 1-6. Harker diagram for phase determination in isomorphous replacement.

Succinctly, the phase information originates from prior knowledge of the composition of the crystal, including atomicity, solvent content, and stereochemistry. This structural information is compiled together during phasing, combined with the diffraction observations (the diffraction intensities measured), and eventually transformed into the electron density distribution in the biomolecules in the crystal, based upon which a 3D model can be built.

Bacteriophage and phage lysis

It has been estimated that there are 10^{31} bacteriophages on this planet, astronomically more than the number of humans on the planet. Actually, phages are the most abundant life forms on earth, and play critical roles in the global carbon cycle (Breitbart and Rohwer, 2005). Phage lysis events occur $>10^{29}$ times a day, which accounts for approximately 20% of the total biomass (Suttle, 2007).

Bateriophage literally means eaters of bacteria. They infect bacteria, reproduce themselves inside the host cell, and at the end of the infection cycle, most of them break down the cell envelope to release their progenies. Then, the released, multiplied phages search for new prey and start new life cycles. The information carried by the phages is thus propagating, and, from the perspective of the phage, hopefully perpetuated. How do they break the host's envelope, which includes two membranes and one layer of tough peptidoglycan in gram-negative bacteria, like *E. coli*? Also the timing of the lysis is critical for the phages competing with each others over billions of years of evolution. The robustness of the host cell should not be disrupted before lysis, yet once the lysis has been started, the process should be done as fast as possible. How are these principles fulfilled? The lysis cassette in phage λ is a representative model of the dsDNA phages to answer these questions (Young et al., 2000).

Holin, endolysin, and Rz/Rz1 lysis cassettes

In phage λ, there is a lysis cassette composed of holin *S*, endolysin *R*, and the overlapping *Rz/Rz1* genes under the control of the late promoter P'_R (**Fig. 1-7**). These

genes are expressed at the end of the lytic cycle to break down the host cell envelope, i.e., the inner membrane, the peptidoglycan layer of the cell wall, and the outer membrane, as depicted in **Figure 1-7B**.

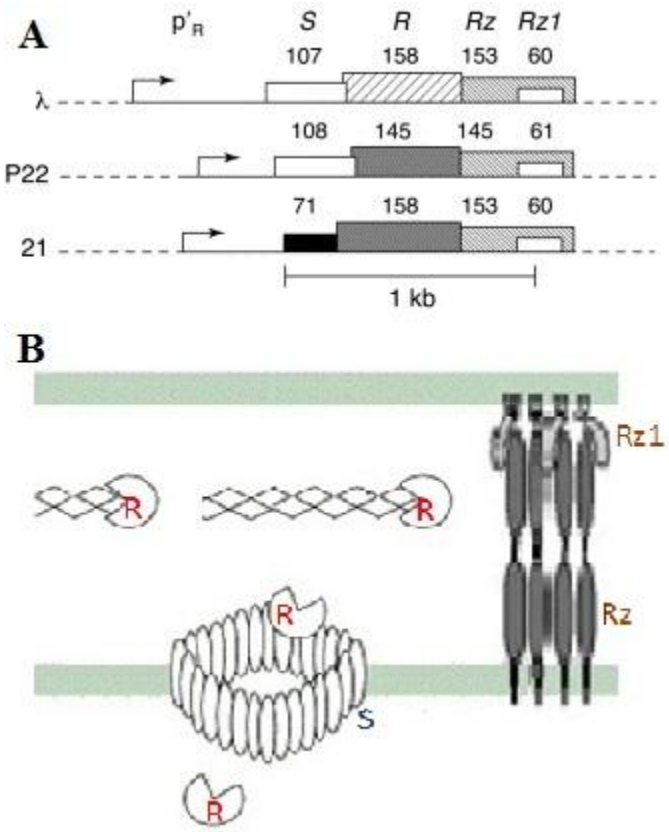


Figure 1-7. Lambdoid lysis strategy. A. Lambdoid lysis cassettes. These cassettes consist of the holin gene *S*, the endolysin gene *R*, and the *Rz/Rz1* gene, from phage λ, P22, and 21, immediately downstream of the sole late promoter, P'_R. **B. The fate of the Gram-negative envelope during λ holin-endolysin lysis.** The endolysins are produced and accumulate in the cytoplasm until the holins form holes in the inner membrane. The released endolysins hydrolyse the peptidoglycan. The outer membrane is further disrupted by the Rz/Rz1 complex. The figure was modified based on the model from Dr. Ry Young (Young et al., 2000) .

The continuous meshwork of peptidoglycan (an indispensible infrastructure for bacteria facing internal osmotic pressure) is the principal barrier to the phage progeny.

The endolysin, R, is a muralytic enzyme that degrades this barrier. However, for a successful infection, endolysin activity must be blocked throughout the latent period, typically 30 - 60 min, until, at lysis, there is a short period of acute enzymatic catalysis on the time scale of seconds (Gründling et al., 2001). Phage λ and some other canonical phages, like T4 and T7, meet these requirements passively, in that their endolysins are synthesized as fully active enzymes but are restricted to the cytoplasm, where they have no substrate, until the instant when the phage-encoded holins trigger and open holes in the membrane (Imada and Tsugita, 1971; Young, 1992). The λ holin makes a hole in the inner membrane so large that even a beta-galactosidase fusion protein greater than 480 kDa can pass through (Wang et al., 2003). However phages like 21 make smaller lesions in the membrane, so-called pinholes, that do not allow the proteins, including λ endolysin, to cross the membrane (Park et al., 2007). There must be a different mechanism. The answer is the SAR endolysins, such as those of phage 21 and P1, which have an intriguing dual-topology conferred by the unique SAR domain at their N-termini (Xu et al., 2004).

SAR endolysins

A SAR (Signal Anchor-Release) domain is an N-terminal transmembrane domain (TMD) that has the unique ability to exit the lipid bilayer completely (Xu et al., 2004). The first SAR domain was identified in Lyz^{P1}, the lysozyme of bacteriophage P1. During the latent period, Lyz^{P1} is expressed and secreted to the periplasm, where it accumulates tethered to the membrane by the SAR domain. Premature destruction of the

cell wall is avoided because in the membrane-tethered form, Lyz^{P1} is catalytically inactive, both conformationally and covalently. In the inactive form ($^{i}Lyz^{P1}$), the N-terminal domain is in a radically different fold from the active enzyme ($^{a}Lyz^{P1}$) and completely lacks its catalytic cleft. In addition, a catalytic Cys residue is occupied in a disulfide bond. Activation occurs via two dramatic transitions concomitant with the escape of the SAR domain from the bilayer (Xu et al., 2005). Covalent activation is achieved when a free thiol in the SAR domain causes a disulfide bond isomerization, releasing the catalytic cysteine. In addition, the entire catalytic domain undergoes a large conformational reorganization, unwinding 3 α-helices to form 3 β-strands in the active site. The extracted SAR domain itself contributes only the liberating thiol, remains largely helical, and makes no intimate contacts with the body of the enzyme.

This elegant regulation, with its conformational and covalent levels, ensures that the phage morphogenesis period is not shortened by premature lysis. However, many SAR endolysins do not have cysteine residues either in the transmembrane domain or the active site, and thus, must be regulated in a different way (Xu et al., 2004). One of those is R^{21}, which was identified as a SAR endolyisin together with Lyz^{P1}, and will be further characterized in Chapter II.

CHAPTER II

R^{21}: THE GENERAL REPRESENTATIVE OF SAR ENDOLYSINS*

Prior to the discovery of the SAR endolysins, all phage lysozymes that had been characterized, like T4 E, lambda R, and T7 gp3.5, were found to be soluble enzymes that accumulate fully folded and enzymatically active in the cytoplasm (Imada and Tsugita, 1971; Young, 1992). Thus, degradation of the cell wall, and subsequent lysis, could be controlled by maintaining the integrity of the membrane until it was permeabilized at the appropriate time by the phage-encoded holin. However, a survey of 196 sequenced phage genomes showed 58 endolysin genes with the Glu-8aa-Asp/Cys-5aa-Thr catalytic triad which characterizes the canonical T4 E lysozyme, a true glycosylase (Sun et al., 2009) (**Fig. 2-1**). Unexpectedly, most (43/58) have N-terminal hydrophobic domains with SAR domain characteristics, i.e., predicted N-terminal transmembrane α-helices which have a higher representation of small hydrophobic residues, like Gly and Ala, and uncharged polar residues, like Ser and Thr, compared to typical transmembrane domains (TMDs) (**Table 2-1**)(Sun et al., 2009). This special N-terminal domain make the endolysin first be produced as a membrane protein, and released into the periplasmic space later when the N-terminal domain escapes from the lipid bilayer. Thus, it plays the role of Signal-Arrest-Release, as described in Chapter I.

---

*This Chapter was rewritten based on the *Nature Structural & Molecular Biology* paper (Sun et al., 2009). The collaborators contribution was published in the paper. Additional NMR data were obtained by Qingan Sun under the instruction of Dr. Andy LiWang and Dr. Xiangming Kong.

Table 2-1 Glycosylase endolysins encoded by genomes of phages with Gram-negative hosts

Endolysin subtype	# phage genomes	Characterized enzymes
Cytoplasmic	15	T4E, P22 gp19
SAR with Cys	9	Lyz^{P1}
SAR without Cys	34	R^{21}
Total	**58**	

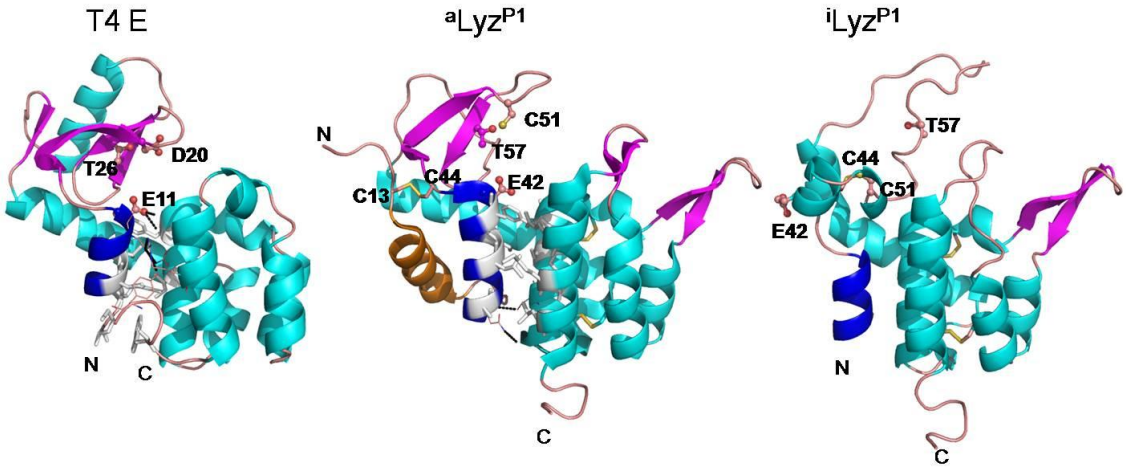


Figure 2-1 Structures of T4 E and the active ($^{a}Lyz^{P1}$) and inactive ($^{i}Lyz^{P1}$) forms of Lyz^{P1}. The protein folds are represented in cartoon format: α-helix in cyan, β-strand in magenta, coil in brown. The stabilizing α-helices of both proteins and the SAR helix of P1 are shown in blue and orange respectively. The active triads are displayed in stick-and-ball representation. The disulfide bonds are shown as yellow stick structures. Polar interactions and residues involved in hydrophobic interactions between the stabilizing helix and the C-terminal helix bundle domain are indicated by dashed lines and light gray coloring, respectively.

Thus, SAR-type endolysin regulation is the rule rather than the exception, at least among sequenced phage genomes. Moreover, among these 43 SAR endolysins, only nine have a Cys residue within the SAR domain and a catalytic cysteine, as observed in Lyz^{P1}. This suggests that, at least among the sequenced genomes of phages known to be

viable, the disulfide bond-dependent regulation of Lyz^{P1} is not the standard, and that the remaining 34 SAR enzymes must be regulated differently. One of these enzymes, R^{21} (the lysozyme of the lambdoid phage 21), has already been confirmed as an SAR endolysin (Xu et al., 2004), and was chosen as a general representative of the missed majority for detailed study.

The SAR domain of R^{21} is essential for its muralytic activity

The SAR domain of Lyz^{P1} is not an essential component of the structure or the enzymatic activity of the protein. The SAR domain is used only to tether the protein to the membrane and, after release from the bilayer, to supply a cysteine for activation of the enzyme (Xu et al., 2005). To determine whether this was also true for R^{21}, the N-terminal 26 residues of R^{21} was replaced with the cleavable secretory signal sequence PelBss (Lei et al., 1987). When the pelBssΦ$R^{21}{}_{27\text{-}165}$ chimera was expressed in *Escherichia coli*, no significant cell lysis was observed even though large amounts of the processed $R^{21}{}_{27\text{-}165}$ protein accumulated in the periplasm. Moreover, an assay of the purified protein did not detect lysozyme activity (**Fig. 2-2; Table 2-2**). The lack of *in vivo* and *in vitro* activity of the truncated R^{21} indicated that the SAR domain of R^{21} is necessary for its enzymatic function. Besides controlling the topology of the protein, the SAR domain in R^{21} plays a specific and more integral role in the catalytic activity of enzyme, compared to that in Lyz^{P1}.

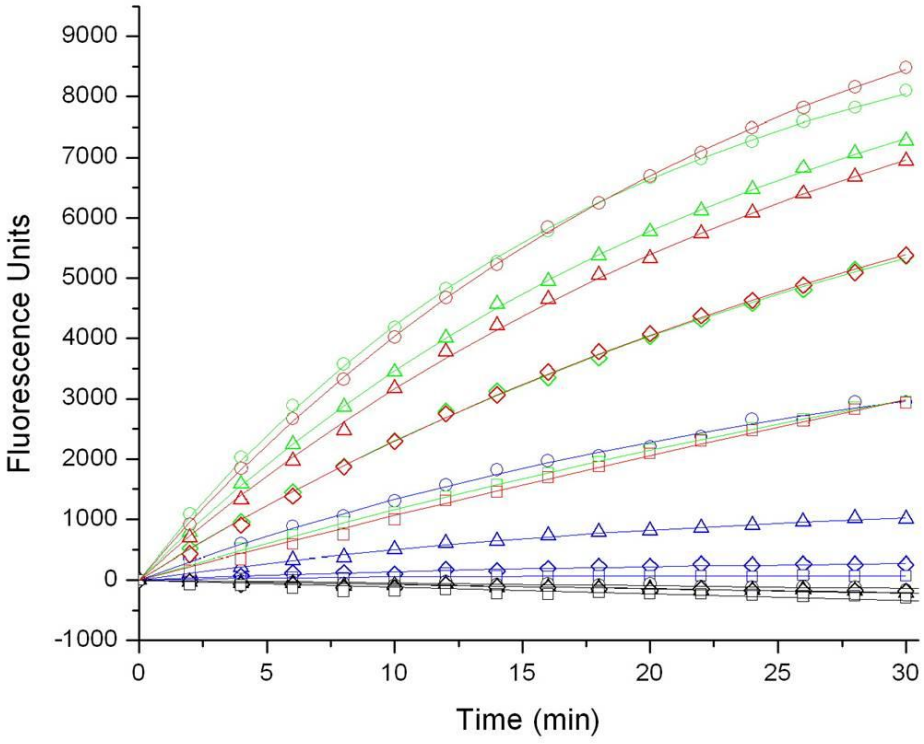


Figure 2-2. Muralytic activity of purified R^{21} and Lyz^{P1}. Lysozyme activity was assayed on purified protein *in vitro*. The relative fluorescence was read every 2 minutes for $^{a}R^{21}$ (in red), $^{i}R^{21}$ (in black), Lyz^{P1} (in green) at various concentrations: (○) 2.5 μM, (Δ) 1.25 μM, (◊) 0.63 μM, (□) 0.31 μM. Lysozyme from chicken egg white (EWL from Invitrogen) was used as the standard (in blue) which was diluted into final concentrations from 250U/ml to 31U/ml. The progress curves were fitted with first order kinetics, which indicated that $^{a}R^{21}$ has similar activity to Lyz^{P1} at about 800 U/ml/μM, while no lysozyme activity was detected for $^{i}R^{21}$.

Table 2-2. Muralytic reaction rates of $^{i}R^{21}$ and $^{a}R^{21}$

Rate (min^{-1}) [1]		Dilution[2]			
		1:1	1:2	1:3	1:4
enzyme	$^{i}R^{21}$	-4.6	-7.0	-7.2	-11.0
	$^{a}R^{21}$	500	375	260	110
	Lyz^{P1}	550	422	269	123
	EWL	160	62	17	10

[1] The reaction rates for $^{a}R^{21}$, Lyz^{P1}, EWL and were obtained by fitting the first-order reaction curve in **Fig. 2-2**. The reaction rate for $^{i}R^{21}$ was the linear slope of fitting curve.
[2] The serial dilutions were started from 2.5 μM for $^{a}R^{21}$, $^{i}R^{21}$, and Lyz^{P1}; from 250 U/ml for chicken EWL.

The structural basis of R^{21} regulation

To explore the structural basis for this novel regulation, we determined the crystal structures of the active, full-length enzyme, $^{a}R^{21}$, and the inactive enzyme, $^{i}R^{21}$, which is missing the entire SAR domain, both to high resolution (**Table 2-3**). $^{a}R^{21}$ shares the characteristic dumbbell structure of the canonical T4 lysozyme, with the 42 residues distal to the SAR sequence forming a relatively independent catalytic domain, containing a catalytic triad (Glu35, Asp44 and Thr50) connected by a long α-helix (Lys68 to Tyr89) to the cluster of C-terminal α-helices (helix α4 to α8) (**Fig. 2-3a**). Despite marginal sequence identities (33.9% of R^{21} with Lyz^{P1}, 30.3% with P22 gp*19*, 18.8% with T4 E), these 3D structural elements are conserved in all of the known structures of T4 E-like lysozymes (R^{21}, Lyz^{P1}, P22 gp*19*, and T4 E) and the geometry of the catalytic triad in $^{a}R^{21}$ (Glu35, Asp44, and Thr50) is nearly identical to those of the other three (**Fig. 2-3b**) (Bell et al., 1991; Mooers and Matthews, 2006; Xu et al., 2005).

In the structures of T4 E and Lyz^{P1} (**Fig. 2-1**), an α-helix from the N-terminal of the protein (Ile3 - Glu11 in E; Asn31 – Gly43 in Lyz^{P1}) interacts extensively with the C-terminal lobe (see gray-highlighted residues in **Fig. 2-1**) and serves to position the essential glutamate in the active site as part of the catalytic triad. In Lyz^{P1}, this stabilizing helix is 12 residues in length and separated from the SAR domain by a turn (NVRT) (Xu et al., 2005). Residues 14-26 of the SAR helix in $^{a}Lyz^{P1}$ assume an α-helical conformation that packs lateral to the stabilizing α-helix. However, in $^{a}R^{21}$, the SAR domain is folded into two anti-parallel α-helices, α1 (residues Pro3 to Gly15) and

Table 2-3. Data collection and refinement statistics of R^{21} crystals

	full-length $^{a}R^{21}$	**truncated $^{i}R^{21}_{27\text{-}165}$**
Data collection[1]		
Space group	*P* 2_1 2_1 2_1	*P* 2_1 2_1 2
Cell dimensions a, b, c (Å)	78.2, 94.8, 97.7	64.2, 109.7, 45.0
Wavelength (Å)	0.9795	0.9796
Resolution (Å)	1.95 (2.05-1.95)	1.70 (1.79-1.70)
I/sigI	19.3 (3.8)	23.9 (4.9)
Completeness (%)	99.6 (99.2)	100 (99.8)
R_{sym}	0.031 (0.268)	0.023 (0.213)
Redundancy	7.2 (5.7)	6.1 (6.1)
Refinement[2]		
# reflections	50721	35715
R_{work}/R_{free}	0.2116/0.2485	0.1991/0.2311
# atoms	5282	2483
RMSD bonds	0.005	0.004
RMSD angles	0.700	0.647

[1]The R^{21} datasets were collected at APS 23ID and processed with HKL2000.
[2]Refinement was done with Phenix.
Accession codes: The atomic coordinates and structure factors have been deposited in the PDB with the accession codes 3HDE and 3HDF for $^{a}R^{21}$ and $^{i}R^{21}$, respectively.

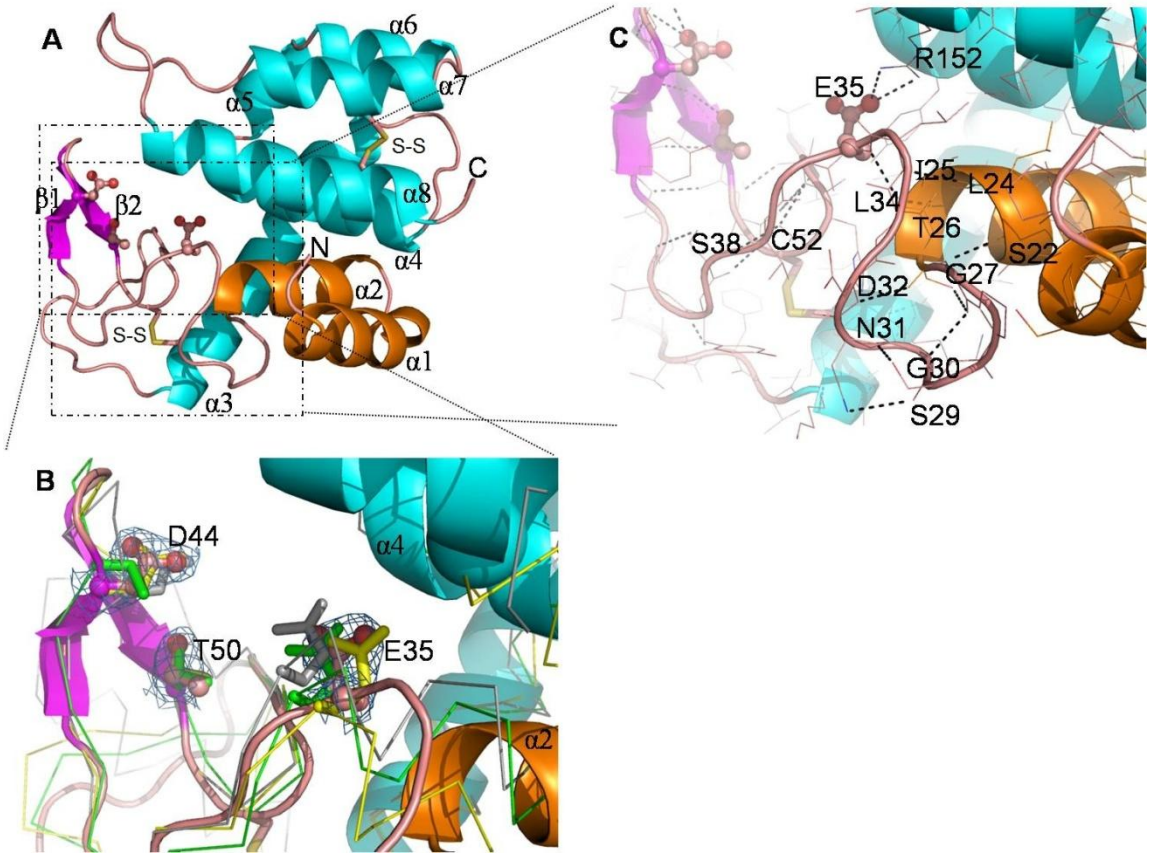


Figure 2-3. Crystal structure of $^{a}R^{21}$. a. Overall structure of $^{a}R^{21}$. The cartoon is colored according to secondary structure (helix in cyan, sheet in magenta, coil in brown), while the SAR helices (α1 and α2) are in orange. The catalytic triad and the two pairs of disulfide bonds (S-S) are shown as stick structures. **b. The catalytic triad (Glu35, Asp44, and Thr50).** The catalytic loop regions of T4 E (Glu11-Thr59, grey C-α trace), Lyz^{P1} (Glu42-Lys72, green C-α trace) and P22 gp*19* (Glu16-Ile50, yellow C-α trace) were superimposed on $^{a}R^{21}$ (Glu35 - Thr67); the RMSs are 0.66 Å, 1.03 Å and 0.40 Å, respectively. The catalytic triads of T4E, Lyz^{P1}, and P22 gp*19* are shown in stick. The 2Fo-Fc electron density map of $^{a}R^{21}$ around the triad is shown at the 1.0 σ level. **c. Hydrogen bond network around Glu35.** The atoms are shown in line (C, brown; O, red; N, blue; S, yellow). The polar contacts are indicated by dashed lines.

α2 (residues Ala17 to Thr26), connected by a sharp turn at Gly16. These two helices pack against the C-terminal helical bundle (α4 – α8) at an angle of about 45º. In this orientation, the second α-helix actually aligns close to the first two turns of the stabilizing helix when $^{a}R^{21}$ is superimposed to Lyz^{P1}. The critical Glu of $^{a}R^{21}$ (Glu35) is located not on the helix itself, but on a loop directly downstream of the SAR helix. This

loop is stabilized by a network of hydrogen bonds which includes residues of the SAR domain (**Fig. 2-3c**). The catalytic Glu35 appears to be further stabilized by the salt bridge between its Oε and the side chain of Arg152 from helix α8 of the C-terminal helical bundle, a feature well-known for T4 E (Strynadka and James, 1996) and also shared by the P22 and P1 enzymes (**Fig. 2-4**). Sequence alignment of R^{21}-like endolysins indicates that this Glu-Arg salt bridge is conserved in this family (**Fig. 2-5**).

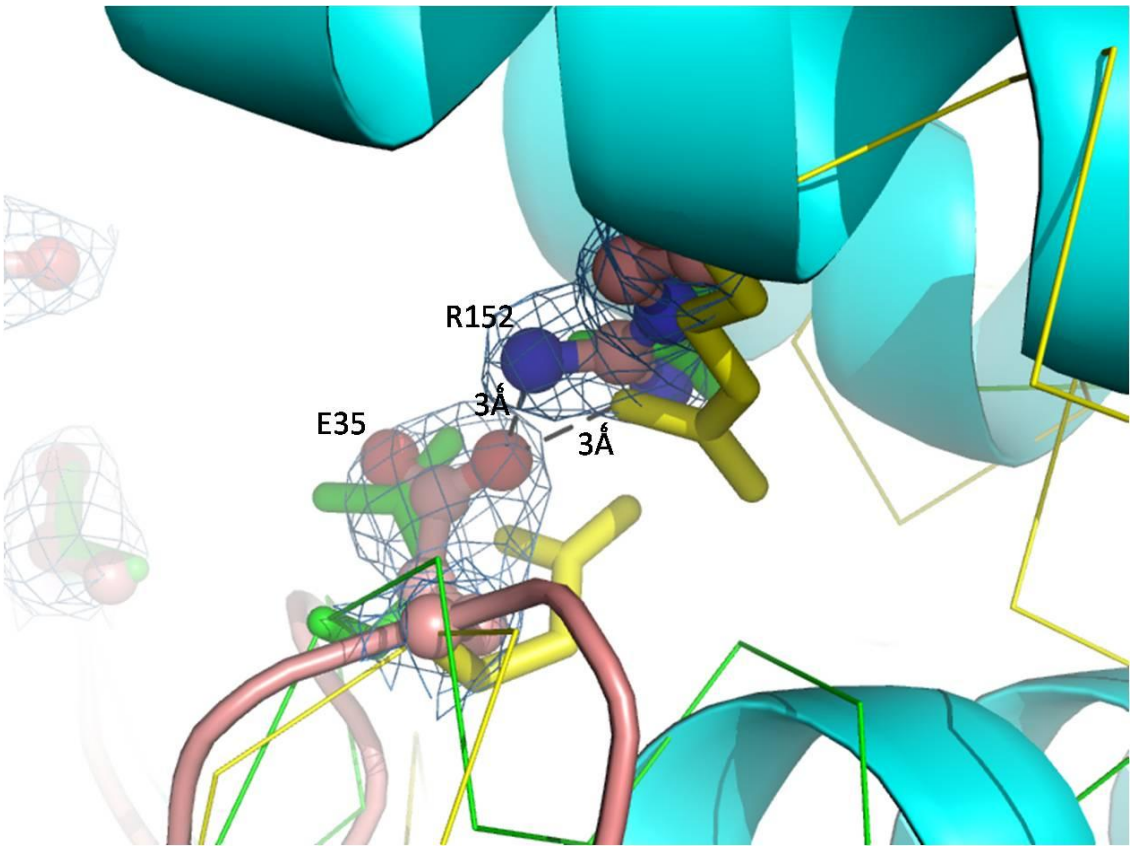


Figure 2-4. The salt bridge between Glu35 and Arg152 in $^{a}R^{21}$. Glu35 and Arg152 in $^{a}R^{21}$ are shown in ball-and-stick. The distance between Oε of Glu35 and Nη of Arg152 is 3 Å, indicated by the dashed line. The corresponding Glu and Arg in Lyz^{P1} and P22 gp*19* are shown in green and yellow stick, respectively.

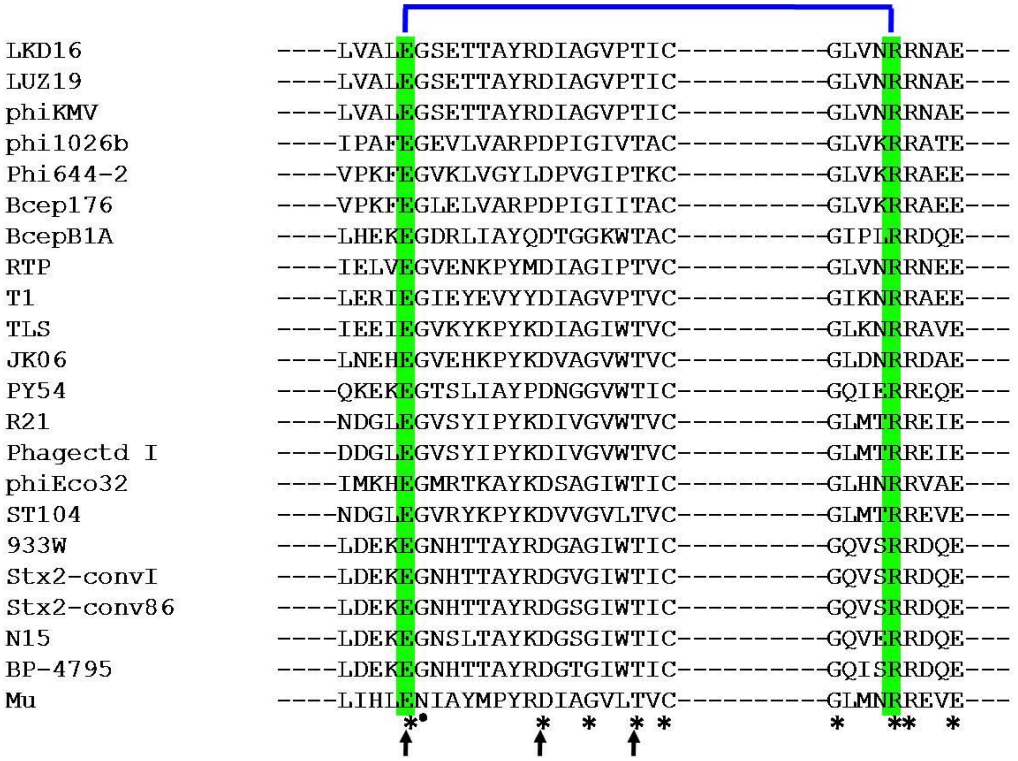


Figure 2-5. The Glu-Arg salt bridge is fully conserved in the R^{21}-like endolysin family. Salt bridge residues are highlighted in green. Only the salt bridge region is shown. The catalytic glutamate residue forms a salt bridge (blue bracket, top) with an Arg in helix 8. Catalytic (arrow), fully conserved (*), and nearly completely conserved residues (dot) are indicated. The labels of the aligned sequences correspond to the following Genbank records: LKD16: gi|158345072, LUZ19: gi|167600491, phiKMV: gi|33300856, phi1026b: gi|38707914, phi644-2: gi|134288680, Bcep176: gi|77864683, BcepB1A: gi|48697551, RTP: gi|81343992, T1: gi|45686348, TLS: gi|148734541, JK06: gi|71834140, PY54: gi|33770570, R^{21}: gi|215468, Phagectd I: gi|148609440, phiEco32: gi|167583572, ST104: gi|46358689, 933W: gi|9632511, Stx2-convI: gi|20065952, Stx2-conv86: gi|116221999, N15: gi|3192716, BP-4795: gi|157166033, Mu: gi|6010396.

Sequestration of the SAR domain induces a conformational change

Additional differences were observed in the structure of the R^{21} inactive form compared to the active enzyme (**Fig. 2-6a**). Although overall the $^{i}R^{21}$ crystal diffracted to a higher resolution than that of $^{a}R^{21}$ (**Table 2-3**), the five N-terminal residues preceding Asn31 were not visible in the electron density map. The catalytic Glu35 can be only built to Cβ, presumably due to the flexibility of this region in the inactive form.

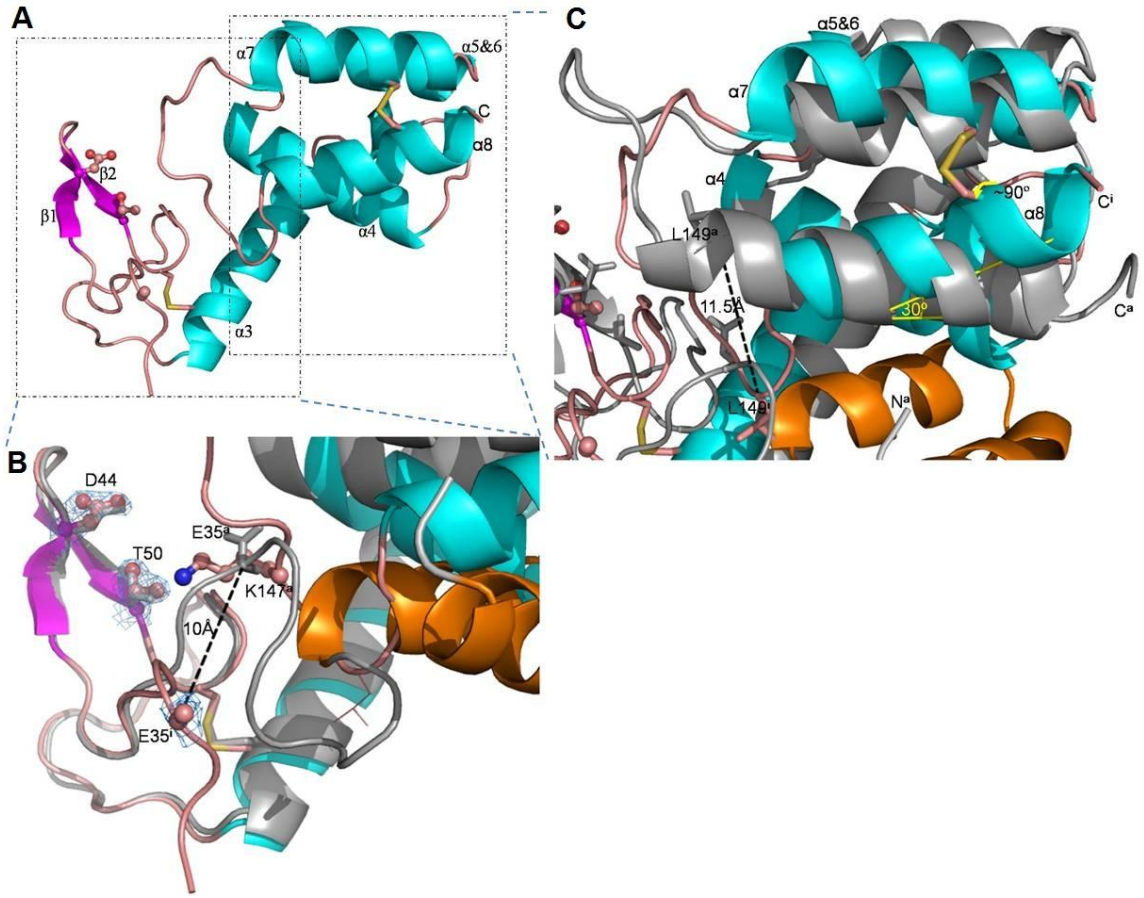
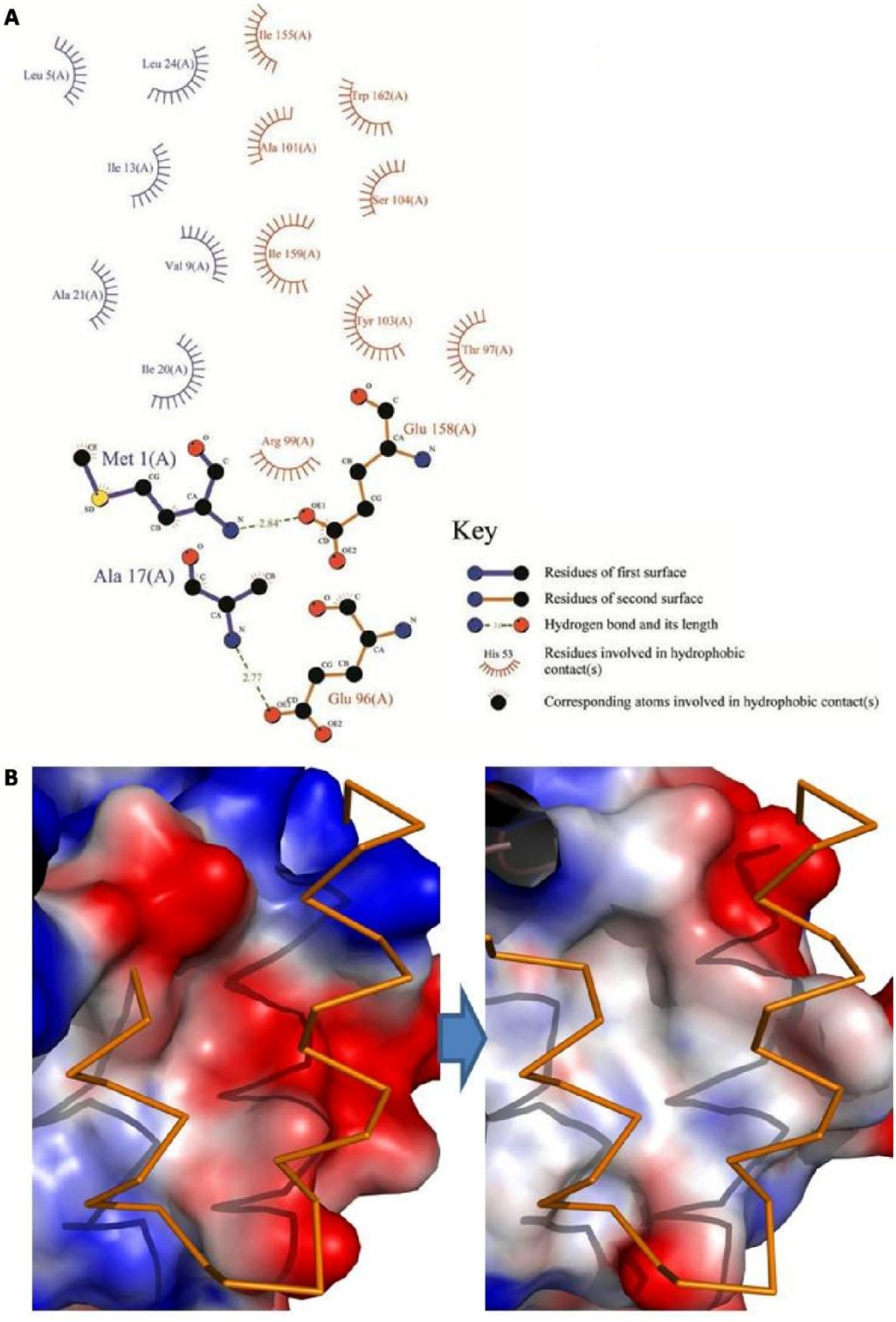


Figure 2-6. Crystal Structure of $^{i}R^{21}$. a. Overall structure of $^{i}R^{21}$. See Fig. 2-4 for cartoon features. The dashed boxes enclose the N-terminal and C-terminal regions, further analyzed in Fig. 2-7B and 2-7C, respectively. **b. Alignment of the catalytic loop regions of $^{i}R^{21}$ and $^{a}R^{21}$.** Except for Glu35, which shows a 10 Å Cα displacement (dashed line), most of the catalytic loop (Ser38 to Thr67) of $^{a}R^{21}$ (grey, SAR domain in orange) can be superimposed on the same region of $^{i}R^{21}$ (RMS = 0.36 Å for 209 atoms). The position of Glu35 in $^{a}R^{21}$ is now occupied by Lys147 (shown in stick-and-ball) from the C-terminal loop between α7 and α8 in $^{i}R^{21}$**.** The 2Fo-Fc electron density map of $^{i}R^{21}$ around the triad is shown at the 1.0 σ level.**c. Alignment of the C-terminal domains of $^{i}R^{21}$ and $^{a}R^{21}$.** $^{i}R^{21}$ helices and loops are cyan and brown, respectively. $^{a}R^{21}$ is in grey and SAR helices are orange. Beginning at Glu96, the backbone RMS between the two structures is 4.70 Å (499 atoms). The displacement of the loop between α7 and α8 is indicated by the dashed line, which is defined by the movement of the Leu149 Cα. The 90º rotation of helix α8, which is defined by the turn of Cys160, is indicated by the yellow arrow.

Whereas in Lyz^{P1}, the active and inactive forms have radically different structures (**Fig. 2-1**), the overall fold of $^{i}R^{21}$ is nearly identical to that of $^{a}R^{21}$, with every major secondary structural element preserved, except the changes to the active site,

described below. Like Lyz^{P1}, R^{21} has two disulfide bonds that provide structural stability, but unlike Lyz^{P1}, there is no difference in the disulfide bonding pattern between the active and inactive forms. In the catalytic domain, the loop region between Ser38 and Thr67, which includes two of the residues of the catalytic triad, Asp44 and Thr50, are superimposable (rmsd ~0.36 Å) in $^{a}R^{21}$ and $^{i}R^{21}$. However, the absence of the two α-helices of the SAR domain and the adjacent H-bond network has a dramatic effect on the position of the Glu35, which is displaced about 10 Å (**Fig. 2-6b**). Its position in the active enzyme is occupied by Lys147 in $^{i}R^{21}$, in the loop connecting two of the C-terminal helices, α7 and α8. Lys147 forms a hydrogen bond through its Nζ to a water molecule 3.0 Å away. This water further interacts with Tyr42, Cys52, and Gly54 in the catalytic region. Lys147 is likely to further ensure that the active site in $^{i}R^{21}$ would not be capable of binding substrate and concomitantly fold into active geometry. Given the indispensable role of the catalytic Glu demonstrated in T4 E (Rennell et al., 1991), the displacement and disorder of Glu35 account for the lack of enzymatic activity in $^{i}R^{21}$.

Figure 2-7. Interaction between SAR and the C-terminal domain in R^{21}. A. Ligplot of these interactions in $^{a}R^{21}$. Hydrogen bonds are indicated by dashed lines between the atoms involved, while hydrophobic contacts are represented by an arc with spokes radiating towards the atoms they contact. The contacted atoms are shown with spokes radiating back. Residues belonging to the SAR domain and the C-terminal helix bundle are shown as the blue side chains on the left-hand side and orange side-chains on the right, respectively (Wallace and Mao, 1984). **B. Polarity switching at the interface for the SAR domain.** The imaginary electrostatic surface (positive = blue; negative = red) contacting the helices of the extracted SAR domain is shown for $^{a}R^{21}$ (right). The corresponding surface is shown for $^{i}R^{21}$ at left, with the SAR helices super-imposed as an orange backbone ribbon trace.


A
Ile 155(A)
Leu 5(A)
Leu 24(A)
Trp 162(A)
Ala 101(A)
Ile 13(A)
Ser 104(A)
Ile 159(A)
Val 9(A)
Ala 21(A)
Tyr 103(A)
Thr 97(A)
Ile 20(A)
Glu 158(A)
Arg 99(A)
Met 1(A)
2.84
Ala 17(A)
2.77
Glu 96(A)
Key
Residues of first surface
Residues of second surface
Hydrogen bond and its length
His 53
Residues involved in hydrophobic contact(s)
Corresponding atoms involved in hydrophobic contact(s)
B

The active and inactive forms of R^{21} also differ significantly in the C-terminal domain (Glu96-Gln164; backbone rmsd ~ 3.96 Å) with helix α8, which contacts the SAR domain in the full-length structure, tilted ~30° toward the catalytic loop, rotated one-quarter turn clockwise and shorter by one turn in the inactive form (**Fig. 2-6c**). The adjacent helix, α7, is longer by one turn and is rotated counter-clockwise. Displacement of this dynamic helical turn is as large as 23.3 Å and intrudes into the space occupied by the SAR domain and the adjacent loop in $^{a}R^{21}$ (see Leu149 in **Fig. 2-6c**), in effect forming a steric barrier that prevents Glu35 from participating in the catalytic triad. In addition, the rotation of helix α8 turns Arg152 towards the inner surface of the helix α4-α8 bundle and prevents the salt bridge with Glu35. Importantly, inspection of the predicted electrostatic surfaces reveals that the interface between the SAR domain and the body of the enzyme is dominated by hydrophobic contacts (**Fig. 2-7a**). In $^{i}R^{21}$, the same surface is anionic and solvent-exposed, mainly due to rotation of helix α8 (**Fig. 2-7b**).

NMR provides structural information in solution

When we pursued the structure of R^{21}, we tried NMR and X-ray crystallography in parallel. Although the crystallography method eventually won out, the NMR experiment still provided irreplaceable information of R^{21} in solution.

First of all, both $^{a}R^{21}$ and $^{i}R^{21}$ yielded well-dispersed NMR spectra, especially for $^{i}R^{21}$. We have assigned the backbone resonances of $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$ for 135 out of 139 residues in $^{i}R^{21}$ based upon the information from the heteronuclear experiments

(**Fig. 2-8; Table A-1**) (Grzesiek and Bax, 1993). This demonstrates that these proteins are stable and homogeneous in solution at the high concentrations necessary for NMR experiments (mM range), and provided us confidence for the crystallization trials.

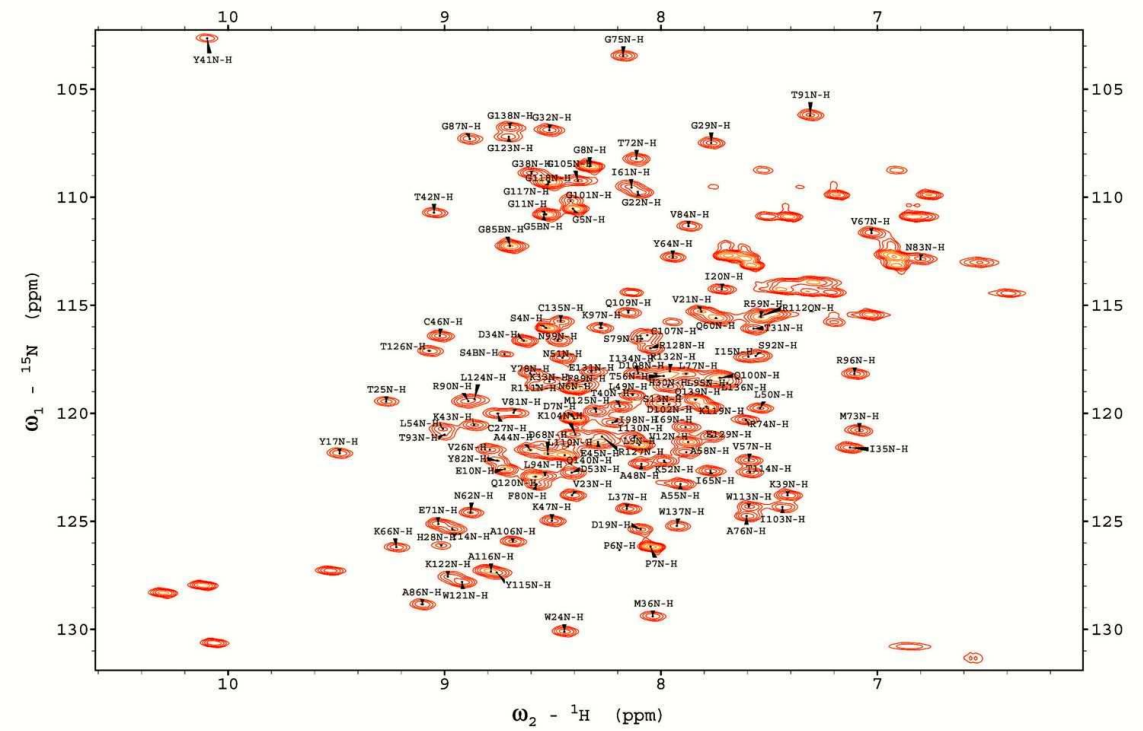


Figure 2-8. 2D ^{1}H-^{15}N HSQC spectrum of ^{15}N-labeled ^{i}R^{21}. The NMR sample was in 150 mM NaCl, 10 mM sodium phosphate, pH 5.5. The residue numbers labeled on the resonances started from Ser4 in ^{i}R^{21}, which corresponds to Ser29 in the full-length ^{a}R^{21}.

Since the two forms of R^{21} crystallized at dramatically different pH (pH 4.6 for ^{a}R^{21}, and pH 7 for ^{i}R^{21}), we measured the 2D ^{1}H-^{15}N HSQC spectra from pH 4.5 to the physiological pH 7.4. There was no obvious shift of the resonances in these spectra. The spectra were also similar for the sample with various ionic strengths, i.e., various concentrations of NaCl in buffer from 50 mM to 150 mM. These data indicate that pH and ionic strength do not affect the conformation of R^{21} in solution, and therefore not

likely in the crystal as well. Thus, the different structures observed via crystallography are not crystallographic artifacts.

When we overlaid the spectra of $^{a}R^{21}$ and $^{i}R^{21}$, most of the resonances were superimposed (**Fig. 2-9**A). This was not surprising since the crystal structures revealed that the overall folds are nearly identical and every major secondary structural element is preserved. According to the x-ray structures, the chemical environment of the backbone amide group should be similar between $^{a}R^{21}$ and $^{i}R^{21}$, which is reflected in the ^{1}H-^{15}N HSQC spectra by similar chemical shifts. However, there are many more peaks in the ^{1}H-^{15}N HSQC spectrum of $^{a}R^{21}$ than in that of $^{i}R^{21}$, more than the increased number of residues in the $^{a}R^{21}$ peptide. The additional well-dispersed signals suggested that some alternative conformations under slow exchange regime exist for $^{a}R^{21}$. To test this idea, we resorted to the zz-exchange experiment, which reveals conformational inter-conversions with exchange rate, k_{ex}, in the range 0.1 to 1000 s^{-1} (Cavanagh, 2007). **Figures 2-9 B** and **C** are the spectra from this experiment, and indicate that indeed a two-state conformational equilibrium exists. Conformational inter-conversion occurs on the time scale of seconds. This dynamic feature may be closely related to the activation from $^{i}R^{21}$ to $^{a}R^{21}$, and is worth further investigation.

Dynamic membrane topology of the SAR domain

Compared to canonical transmembrane domains, SAR domains are enriched in Gly and Ala residues, a feature that is critical for dynamic membrane topology. In R^{21}, there is a cluster of Gly residues (Gly14-Gly16). Replacing two of them with Leu

eliminates the ability of the SAR endolysin to escape from the membrane (Sun et al., 2009). Nevertheless, SAR domains still have numerous hydrophobic residues which, after extraction from the lipidic milieu, are a liability in the aqueous environment of the

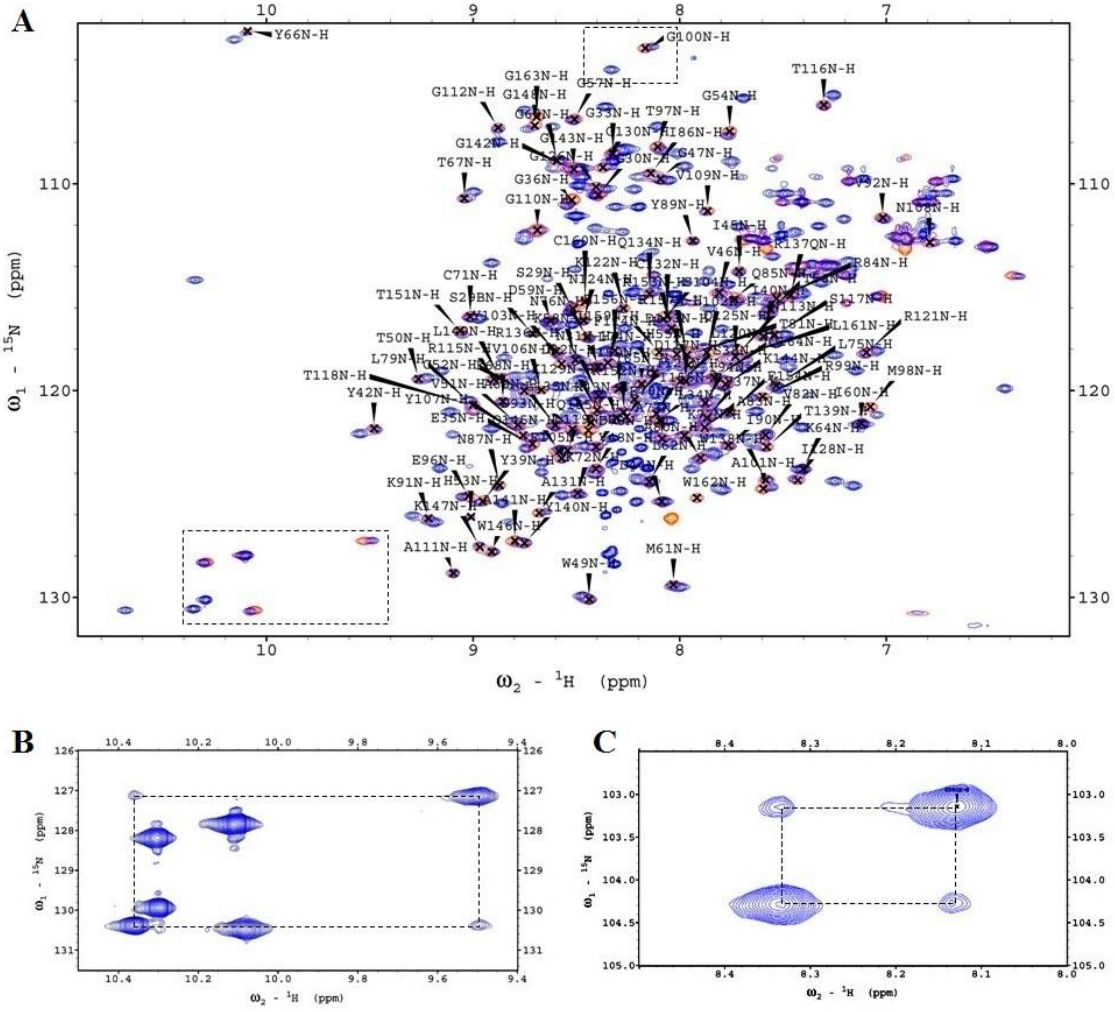


Figure 2-9. Comparison of NMR spectra between $^{a}R^{21}$ and $^{i}R^{21}$. A. Overlay of ^{1}H-^{15}N HSQC spectrum of $^{a}R^{21}$ (blue) on that of $^{i}R^{21}$(red). The backbone assignment was from $^{i}R^{21}$, the residue number is the same as the full-length $^{a}R^{21}$ sequence. The dashed boxes indicate the signature region of tryptophan ^{1}H-$^{15}N^{\varepsilon 1}$ and Gly100 that are further examined in the zz-exchange experiment. **B. ZZ-exchange spectrum in the signature region for tryptophan indole group. C. ZZ-exchange spectrum for Gly100.** The exchange peaks are linked to HSQC resonances corresponding to two alternative conformations by the dashed lines. The ZZ-exchange spectrum was acquired with a 0.8 s mixing period.

periplasm. In R^{21}, the cluster of Gly residues in the SAR domain provides a second structural function, serving as the hinge that allows the extracted SAR domain to fold into two shorter helices. Besides creating a new helix-helix interface to accommodate hydrophobic residues, this also compacts the SAR domain sufficiently so that it can fit into a hydrophobic pocket on the surface of the main body of the enzyme. Consequently, nearly all the hydrophobic residues of the SAR domain are buried in the active structure, and the most solvent-exposed surface is populated by hydrophilic residues.

A new mechanism for the regulation of SAR endolysins

Mechanistic insight for the regulation of secretory endolysins became available only recently, for the SAR endolysin of coliphage P1(Xu et al., 2005; Xu et al., 2004). After Sec-mediated export, LyzP1 is tethered to the membrane by its N-terminal SAR domain and maintained in an inactive state by virtue of a disulfide bond between the catalytic Cys51 and a second Cys residue (Cys44) in the catalytic domain, as well as by the complete disorganization of its N-terminal catalytic domain. The SAR domain escapes from the membrane, at a low constitutive rate or quantitatively when the holin triggers to disrupt the membrane. In the active, soluble form of Lyz^{P1}, a Cys residue in the SAR domain replaces the catalytic Cys in the disulfide bond, and the N-terminal catalytic domain is completely refolded, yielding a properly-arranged catalytic triad straddling the substrate-binding cleft (Xu et al., 2005). In this active structure, the SAR domain makes no significant contacts with the body of the enzyme other than the disulfide linkage. In fact, except for the disulfide linkage, the SAR domain of Lyz^{P1} is

not essential for the function of the endolysin. Even the position of the Cys residue within the SAR domain is very flexible, suggesting that the SAR domain can be arranged in a variety of orientations with respect to the catalytic domain (Xu et al., 2005).

The studies presented here describe an entirely different mode of regulation for the SAR endolysin of phage 21, the prototype of a class of SAR endolysins much more frequently represented in annotated phage genomes than the Lyz^{P1} class (**Table 2-1**). Although $^{a}R^{21}$ shares both its overall structure and the geometry of its catalytic triad and substrate-binding cleft with Lyz^{P1} and with T4 E (**Fig. 2-3b**), there is a fundamental difference. In $^{a}R^{21}$, the SAR domain, when removed from the membrane where it is presumably entirely helical, "jack-knifes" into an α-helix-turn-α-helix structure that makes many intimate contacts with the N-terminal catalytic domain and the C-terminal helical bundle and is essential to the activity of the enzyme (**Fig. 2-10**). Moreover, the overall structure of the N-terminal catalytic domain is much more preserved between the $^{i}R^{21}$ and $^{a}R^{21}$ structures, in comparison to the dramatically different folds of $^{i}Lyz^{P1}$ and $^{a}Lyz^{P1}$. Only a loop segment of the catalytic domain adjacent to the SAR sequence is differently arranged, with the Glu residue of the catalytic triad everted from the enzymatic cleft in $^{i}R^{21}$ (**Figs. 2-6B, 5**). Another striking contrast is that activation of R^{21} involves significant reorientation and restructuring of the C-terminal helical bundle, resulting in the formation of a hydrophobic binding site for the helix-turn-helix structure of the released SAR domain (**Figs. 2-6C, 2-10**). In $^{i}Lyz^{P1}$ and $^{a}Lyz^{P1}$, the C-terminal domains are very similar (Xu et al., 2005). Overall, the structural transitions are much

less dramatic in R^{21} activation than for Lyz^{P1}. Thus, paradoxically, after extraction from the membrane the SAR domain of R^{21} enters the structure of the enzyme intimately without causing fundamental restructuring. In contrast, the extracted SAR domain of Lyz^{P1} provides only its thiol group for covalent activation of the catalytic triad and can do so in a number of orientations, yet activation of Lyz^{P1} is associated with a complete refolding of the entire N-terminal catalytic domain and *de novo* creation of the substrate-binding cleft. The simplest notion is that the periplasmic domain of ^{i}Lyz*P1* is much less stable. Upon insertion into the membrane, Lyz^{P1} is remodeled by the host DsbA (Xu et al., 2004), and may be conformationally remodeled by host chaperones after release. For R^{21}, the small but distinct differences between the conformations of the bundled C-terminal helices, especially helices α4 and α8, suggest a possible induced fit mechanism for its activation, which arises the alternative conformations in the C-terminal sites, like Gly100, that we observed in the NMR experiments. Once the SAR domain has escaped from the membrane, the simplest way to minimize the exposure of its most hydrophobic residues is to pack against the non-polar face of helix α4, which could lead to dislodgement of helix α8 and further conformational adjustments. This results in the alternative minima in the energy landscape, and the protein conformation likely resonates dynamically between these minima, and the alternative conformation may be further selected by the binding of substrate. In this perspective, the newly released SAR domain activates R^{21} through the spatially adjacent C-terminal helical bundle, rather than the sequence-close N-terminal domain.

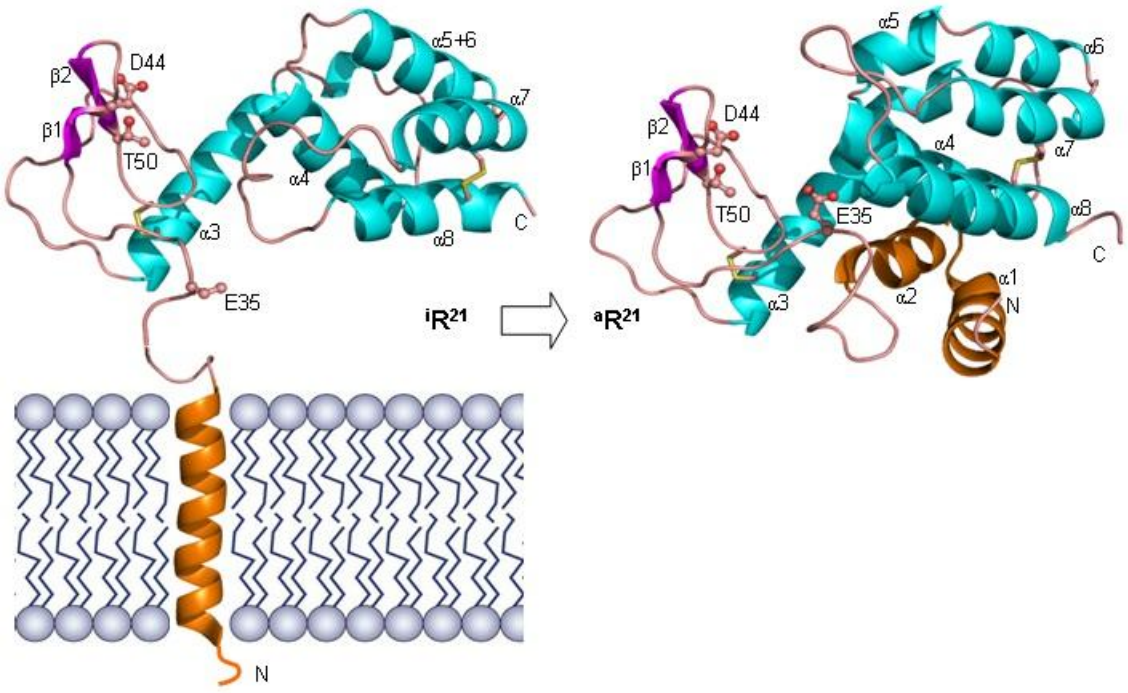


Figure 2-10. Topological and conformational dynamics of R^{21} activation. Color scheme as in **Fig. 2-1**. On the left, the SAR domain is shown as a membrane-spanning helix in the inactive, membrane-tethered form; the rest of the protein is from the x-ray structure of $^{i}R^{21}$.On the right is the crystal structure of $^{a}R^{21}$.

Comparison of the two strategies

The regulatory strategies adopted by Lyz^{P1} and R^{21} have significant differences. $^{i}Lyz^{P1}$ has two levels of negative control: covalent inactivation of its active site Cys and a N-terminal catalytic domain with radical conformational disability. In contrast, $^{i}R^{21}$ appears to lack only the correct placement of its catalytic Glu, which is sterically blocked from attaining its position by a hydrogen-bonded lysine side-chain. These striking differences suggest evolutionary trade-offs exacted to meet the conflicting dual responsibilities of the SAR endolysin: maintenance of an inactive state throughout the infection cycle and also the ability to effect degradation of the murein as rapidly as

possible, once the holin has triggered and depolarized the membrane. Moreover, in contrast to canonical endolysins, which are completely blocked from premature muralytic function by sequestration in the cytoplasmic compartment, SAR endolysins are inherently leaky, in the sense that even without holin function, enzyme is constitutively released from the membrane, and thus becoming active, during the latent period (Park et al., 2006). Lyz^{P1} appears to be better suited for negative control, in view of the Cys substitution for Asp in the catalytic triad, which has been shown to reduce enzyme activity in the T4 E context (Hardy and Poteete, 1991), and the fact that the inactive form is both covalently and conformationally crippled. R^{21} seems to be more poised for muralytic function, with its canonical catalytic triad and the minimally dysfunctional conformation of the inactive enzyme. The particular evolutionary path taken by these two SAR endolysins presumably reflect the selective pressures exerted on the lysis process. In particular, analyses based on the Minimum Value Theorem for predator-prey relationships have indicated that conditions of sparse host availability and/or poor growth conditions favor extended infection cycles, thus favoring an endolysin less prone to premature activation (Wang et al., 1996). In contrast, target-rich and/or rapid growth environments favor shorter infection cycles and more rapid release of the progeny virions, which would favor the more efficient and muralytic endolysin.

Is R21 representative of the Cys-less SAR endolysins, in terms of the structure and role of the SAR domain in the active enzyme? A compilation of the N-terminal sequences including the SAR domains and the catalytic triads of 22 Cys-less SAR endolysins suggests that there may be several different modes (**Fig. 2-11**), since the SAR

domain varies significantly in its proximity to the active site Glu. R^{21} shares similar SAR domain arrangements with endolysins from phages encoding potent human toxins, including the cytolethal distending toxin (Asakura et al., 2007), CDT, and the Shiga toxin (Plunkett et al., 1999); moreover, in the latter case, the acute dispersal of the toxin in the mammalian gut is dependent on the muralytic action of the SAR endolysin (O'Brien et al., 1989; Ochoa et al., 2007). In no case does the hydrophobic core of the SAR domain extend farther than 3 - 4 helical turns without encountering a Gly or Pro residue. These residues are tolerated in helices in the membrane but are helix-breakers in aqueous solvent (Li et al., 1996), suggesting that in most or all of these proteins, the SAR domain uses some version of the R^{21} strategy of converting the SAR sequence into a helix-turn-helix packed against the main body of the enzyme to solve the problem of what to do with a transmembrane domain out of its native lipidic environment. In some homologs, the SAR domain immediately abuts the catalytic Glu residue; in these enzymes, most of the SAR domain must constitute the stabilizing helix. Taken together with the structures described here, it seems clear that SAR domains represent structural elements of unparalleled evolutionary flexibility, both in terms of membrane topology and the capacity for integration as a regulatory component.

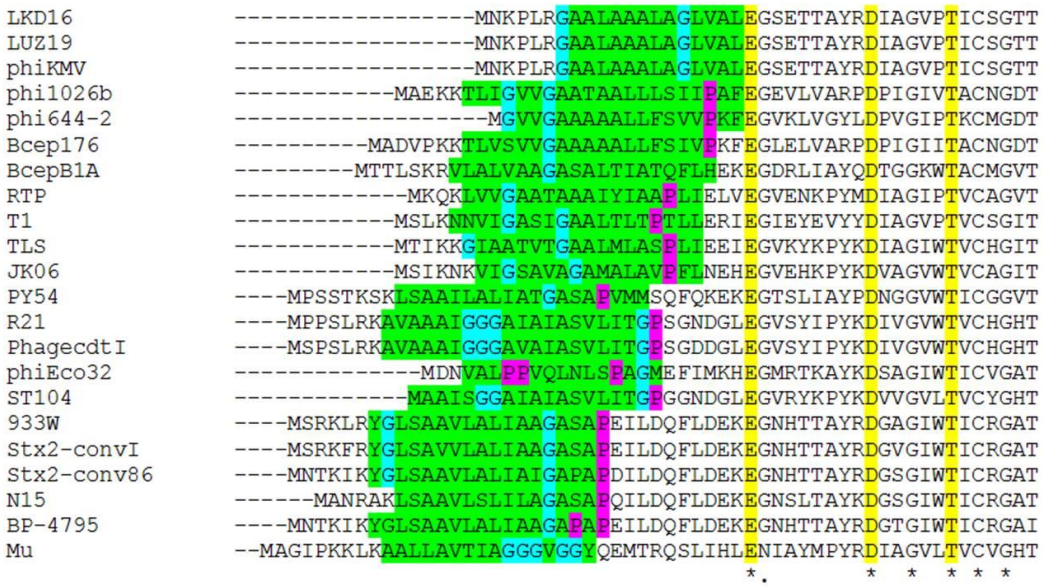


Figure 2-11. Alignments of unique SAR endolysins lacking Cys residues in the membrane domains. The N-terminal SAR and catalytic domains of R^{21}-like SAR endolysins are shown aligned by the catalytic triad (yellow). These sequences were selected from the 34 sequences identified in **Table 2-1** by the criterion of having at least one amino acid residue difference in the complete sequence. Completely conserved (asterisk) and nearly completely conserved (dot) residues are indicated. SAR domains are indicated in green, along with helix-breaking residues Gly (blue) and Pro (purple). The labels of the aligned sequences correspond to the following Genbank records: LKD16: gi|158345072, LUZ19: gi|167600491, phiKMV: gi|33300856, phi1026b: gi|38707914, phi644-2: gi|134288680, Bcep176: gi|77864683, BcepB1A: gi|48697551, RTP: gi|81343992, T1: gi|45686348, TLS: gi|148734541, JK06: gi|71834140, PY54: gi|33770570, R^{21}: gi|215468, Phagectd I: gi|148609440, phiEco32: gi|167583572, ST104: gi|46358689, 933W: gi|9632511, Stx2-convI: gi|20065952, Stx2-conv86: gi|116221999, N15: gi|3192716, BP-4795: gi|157166033, Mu: gi|6010396.

Materials and methods

Expression and protein purification

The full-length R^{21} gene was cloned into the pET11a vector between the NdeI and BamHI sites with C-terminal His-tag introduced with PCR primers. The fragment $R^{21}_{27\text{-}165}$ without a stop codon was cloned into the pET22b vector between the NcoI and XhoI sites. These two plasmids were dubbed as pETR^{21} and pETpelBΦ $R^{21}_{27\text{-}165}$, respectively.

For production of $^{a}R^{21}$, the plasmid pETR^{21} was transformed into Rosetta λ(DE3) pLysS (Novagen, Madison, WI). Tranformants were grown in Terrific Broth (BD, Franklin Lake, NJ) to $A_{600} \approx 1$ - 1.5 and induced with 0.5 mM IPTG. The culture was arrested between 60 to 90 minutes after induction, depending on the extent of lysis. The partially lysed culture was spun at 4000 rpm for 30 minutes at 4°C. Harvested cell pellets were stored at -20°C. The cells were resuspended in buffer (25 mM Tris-HCl (pH 8.0), 500 mM NaCl, 3 mM β-Mercaptoethanol, 1% Triton X-100, and 1% EDTA-free protease inhibitor cocktail (Roche Molecular)), and were disrupted by 2 passages through a French pressure cell at ~16,000 psi. After the lysate was centrifuged at 40,000 rpm (Beckman Ti50.2) for 60 - 90 min, the supernatant was collected and applied to a 5 ml HiTrap-chelating Sepharose column (Amersham Pharmacia) equilibrated with buffer A (25 mM Tris-HCl, pH 8.0, 500 mM NaCl, 2 mM β-mercaptoethanol and 10 mM imidazole). The column was washed with buffer A followed by an increased concentration of imidazole (25 mM). Bound protein was eluted with a 25-500 mM imidazole gradient and the $^{a}R^{21}$ enriched fractions were pooled, concentrated, and dialyzed against a buffer containing 25 mM sodium acetate pH 5.0, 25 mM NaCl and 1 mM DTT. Supernatant containing $^{a}R^{21}$ was concentrated and 50 µl aliquots were flash frozen in liquid nitrogen and stored at -80 °C for further use.

For production of $^{i}R^{21}$, BL21λ(DE3) cells were transformed with pETpelBΦ $R^{21}{}_{27\text{-}165}$ and grown at 37 °C in LB medium containing 100 µg/ml of carbenicillin until $A_{600} \approx 0.8$. The culture was induced with 0.5 mM IPTG and incubated for 18 h at 25 °C.

$^{i}R^{21}$ protein was purified through a Ni-affinity column and size-exclusion chromatography as described above.

A similar procedure was used for the isotopic-labeled R^{21} in NMR experiments. The BL21 λ(DE3) cells were grown at 37 ºC in minimal media (1.3% KH_2PO_4, 1% K_2HPO_4, 0.9% Na_2HPO_4, 0.24% K_2SO_4, pH 7.2~7.4) which contains $^{15}NH_4Cl$ as the only nitrogen source, and with either $^{13}C_6$-dextrose or unlabeled dextrose as the carbon source. The proteins were purified in the same ways as described above, except for the $^{i}R^{21}$ protein used in the resonance-assignment experiments, which was treated with C-terminal carboxypeptidase A (Sigma-Aldrich, St. Louis MO) for 1 week after purification at 4 ºC, then further polished by S-75 size-exclusion chromatography.

Lysozyme assay

Lysozyme activity was assayed with the EnzChek® Lysozyme Assay Kit (Invitrogen, Carlsbad CA) which measures lysozyme activity on fluorescently-labeled *Micrococcus lysodeikticus* cell walls. When the cell wall is hydrolyzed, the fluorescein is released from the cell wall resulting in an increase in fluorescence. Lyz^{P1}, $^{a}R^{21}$ and $^{i}R^{21}$ were prepared as 0.1 mM stock solutions in reaction buffer (0.1 M sodium phosphate, 0.1 M NaCl, pH 7.5, 2 mM sodium azide), and then serially diluted in a microplate to 0.3 - 2.5 μM. For standards, the chicken egg white lysozyme (EWL) from the kit was diluted from a 1000 U/ml stock with the same reaction buffer. The fluorescently-labeled *M. lysodeikticus* cell wall suspension was added as the substrate to start the reaction. The

fluorescence was measured in a POLARStar Optima plate reader (BMG LabTech) every 2 min using excitation/emission at 485/520 nm.

Crystallization and structure determination

Purified $^{a}R^{21}$ at 20~30 mg/ml concentration was used to screen a large number of commercially available crystallization conditions by the vapor diffusion method. Diffraction quality crystals of $^{a}R^{21}$ were obtained with 150 mM sodium acetate pH 4.6 and 2 M sodium formate. $^{i}R^{21}$ was crystallized in 100 mM Bis-Tris propane pH 7, 2.5 M sodium nitrate. The best crystals grew in 2 weeks and were flash frozen in mother liquor with 20% glycerol, and then stored in liquid nitrogen. Data sets were collected in the lab (Rigaku MM007HF generator with Bruker Smart 6000 CCD detector) and synchrotron beamlines at the Advanced Photon Source (APS 23-ID and 19-ID), Argonne National Laboratory, Chicago.

The initial structure was solved by single-wavelength dispersion using an iodine derivative obtained by quick soaking the crystal in the cryo with 0.5 M NaI for 30~90s. The model was refined against the native dataset iteratively until the R/R_{free} reached 0.212/0.249 at 1.95 Å for $^{a}R^{21}$ and 0.199/0.231 at 1.7 Å for $^{i}R^{21}$. The phasing and refinement were carried out with the Phenix package (Adams et al., 2002). The models were built in Coot (Emsley and Cowtan, 2004). PyMol (Delano Scientific; San Carlos CA) was used for structure analysis and rendering.

$^{a}R^{21}$ crystallized in space group $P2_12_12_1$ with unit cell dimensions of a = 78.2 Å, b = 94.8 Å, c = 97.7 Å, $\alpha = \beta = \gamma = 90°$ (**Table 2-3**) and four $^{a}R^{21}$ molecules in the

asymmetric unit (ASU). All the residues can be traced in the electron density map. The first three molecules are nearly identical (backbone RMS = 0.4~0.5 Å). The fourth molecule in the ASU is the same as the other three except for two loops, Gly27-Ser38 and Thr137-Gly146. $^{i}R^{21}$ crystallizes in space group $P2_12_12$ with unit cell dimensions of a = 64.2 Å, b = 109.7 Å, c = 45.0 Å, $\alpha = \beta = \gamma = 90°$ and two $^{i}R^{21}$ molecules in the ASU (backbone RMS =0.2 Å). The first seven and last four residues are invisible in the structure. For both $^{a}R^{21}$ and $^{i}R^{21}$, the first molecule (chain A) in the asymmetric unit is used as the representative structure. The atomic coordinates and structure factors have been deposited in the PDB with the accession codes 3HDE and 3HDF for $^{a}R^{21}$ and $^{i}R^{21}$, respectively.

NMR measurements

Samples for NMR measurements included around 1 mM R^{21}, 50~150 mM NaCl, 10 mM sodium phosphate (pH 4.5~7.4 at 25 ºC), 50 μM NaN_3, 0.5 mM DSS and 8% D_2O. The ionic strength and pH was varied in the samples to address the corresponding effects on protein conformation. NMR spectra were collected on a Varian 600 MHz spectrometer. Backbone assignments of $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$ in $^{i}R^{21}$ were obtained by using CBCANH and CBCA(CO)NH experiments (Grzesiek and Bax, 1993). Slow conformational interconversion was investigated by the zz-exchange experiments for $^{a}R^{21}$ (Tuinstra et al., 2008). The data were processed with NMRPipe and analyzed with Sparky (Delaglio et al., 1995).

CHAPTER III

D29B: A NOVEL MYCOLYLARABINOGALCATAN ESTERASE IN THE MYCOBACTERIUM CELL ENVELOPE*

Summary

Mycobacteriophages encounter a unique problem among phages of Gram-positive bacteria, in that lysis must not only degrade the peptidoglycan layer but also circumvent a mycolic acid-rich outer membrane covalently attached to the arabinogalactan–peptidoglycan complex. Mycobacteriophages accomplish this by producing two lysis enzymes, Lysin A (LysA) that hydrolyses peptidoglycan, and Lysin B (LysB), a novel mycolylarabinogalactan esterase, that cleaves the mycolylarabinogalactan bond to release free mycolic acids. The structure of LysB from mycobacteriophage D29 shows an α/β hydrolase organization with a catalytic triad common to cutinases, but which contains an additional four-helix domain implicated in the binding of lipid substrates. Whereas LysA is essential for mycobacterial lysis, a Giles ΔlysB mutant mycobacteriophage is viable, but defective in the normal timing, progression and completion of host cell lysis. We propose that LysB facilitates lysis by compromising the integrity of the mycobacterial outer membrane linkage to the arabinogalactan–peptidoglycan layer.

---

*Reprinted with permission from "Mycobacteriophage Lysin B is a novel mycolylarabino-galactan esterase" by Payne, K., Sun, Q., Sacchettini, J. C., Hatfull, G. F., 2009. *Molecular Microbiology*, 73, 367-381, Copyright [2000-2011] by John Wiley & Sons, Inc. Payne, K. in Hatfull's Lab did the genetics studies and performed the enzyme assay. Sun, Q. did the structural biology studies.

Introduction

Upon completion of lytic bacteriophage growth the integrity of the host cell wall must be compromised to release progeny phage particles (Wang et al., 2000; Young et al., 2000). Double-stranded DNA (dsDNA) tailed phages typically encode an endolysin that degrades the peptidoglycan mesh, together with a means of allowing this enzyme access to its substrate, either via a membrane-localized holin, or through the action of holin-independent SAR (signal-anchor-release) endolysins (Wang et al., 2000; Xu et al., 2005; Young et al., 2000). For phages infecting Gram-positive hosts, not only are these two requirements sufficient for cell lysis (Loessner *et al.*, 1998), but exogenously added endolysin can efficiently kill host cells, with considerable therapeutic potential (Fischetti, 2008; Loeffler et al., 2001; Schuch et al., 2002). In Gram-negative hosts, the outer membrane presents a further barrier to lysis, and phages typically encode an additional set of lysis proteins (Rz/Rz1 proteins or spanin equivalents) that are proposed to complete lysis by fusing the inner and outer membranes (Berry et al., 2008; Summer et al., 2007).

Although the mycobacteria are members of the Gram-positive *Actinomycetales,* they are unusual in that they possess a mycolic acid-rich outer membrane, which is covalently attached to the arabinogalactan-peptidoglycan complex (Hoffmann et al., 2008; Zuber et al., 2008). Mycolic acids are α-alkyl, β-hydroxy $C_{60\text{-}90}$ fatty acids with a relatively short ($C_{20\text{-}25}$) saturated α and a longer (C_{60}) meromycolate chain (the β-hydroxy branch) containing double bonds, cyclopropane rings, and oxygenated groups, depending on the *Mycobacterium* species (Hamid et al., 1993; Watanabe et al., 2001).

The outer membrane mycolic acids are linked to the cell wall via an ester linkage to the terminal pentaarabinofuranosyl components of arabinogalactan (Brennan and Nikaido, 1995; Hoffmann et al., 2008; McNeil et al., 1991; Niederweis, 2008), which in turn is covalently anchored to peptidoglycan (Brennan and Nikaido, 1995; McNeil et al., 1990). In addition to being part of the mycobacterial outer membrane, mycolic acids are constituents of trehalose dimycolate (TDM, also known as cord factor), a secreted molecule important for mycobacterial pathogenesis (Brennan, 2003). Mycolic acid synthesis is essential for viability and is the target of the major anti-tuberculosis drug isoniazid (Vilcheze and Jacobs, 2007). The mycobacterial outer membrane also plays roles in nutrient acquisition (Niederweis, 2008) and as an immune target (Karakousis *et al.*, 2004). Furthermore, it presents a potential barrier to phage-mediated lysis.

A total of 60 mycobacteriophage genomes have been sequenced (Hatfull et al., 2006; Morris et al., 2008; Pedulla et al., 2003; Pham et al., 2007) , all of which encode a putative endolysin, Lysin A (LysA) (Garcia et al., 2002; Hatfull et al., 2006). The LysA proteins are highly diverse in sequence and modular in nature (Hatfull *et al.*, 2006), containing combinations of amidase, glycosidase and peptidase motifs. The LysA of phage Ms6 has been shown to have mureinolytic activity in *Escherichia coli* (Garcia et al., 2002) and the LysA of mycobacteriophage Giles is essential for lytic growth (Marinelli *et al.*, 2008). Many of these genomes have a putative holin gene, which is closely linked to *lysA*, whereas in others there is no closely linked and easily recognizable holin gene. Holin function may be encoded from elsewhere as in phage T7 (Hausmann, 1988) or from within another open reading frame as proposed for phage 187

(Loessner *et al.*, 1999). Mycobacteriophage endolysins could contain SAR (signal-anchor-release) domains that allow secretion of the endolysin into the periplasmic space thus rendering them holin-independent (Xu et al., 2005), but none have yet been described. Some mycobacteriophages also encode a second proposed lysis protein, Lysin B (LysB) (Garcia *et al.*, 2002), which does not resemble endolysins and has no identifiable homologs in bacteriophages other than those that infect mycobacterial hosts. The LysB of Ms6 has been shown to be a lipolytic enzyme (Gil et al., 2008) and contains a G-X-S-X-G motif characteristic of serine esterases (Carvalho et al., 1999; Gil et al., 2008; Gupta et al., 2004; Longhi and Cambillau, 1999; Ollis et al., 1992) although its specific substrate is unknown.

We show here that LysB of mycobacteriophage D29 is a novel mycolylarabino-galactan esterase that cleaves the ester linkage joining the mycolic acid-rich outer membrane to arabinogalactan, releasing free mycolic acids. We propose that LysB acts at a late stage in lysis, severing the connection of the mycobacterial outer membrane to the cell wall and completing lysis of the host.

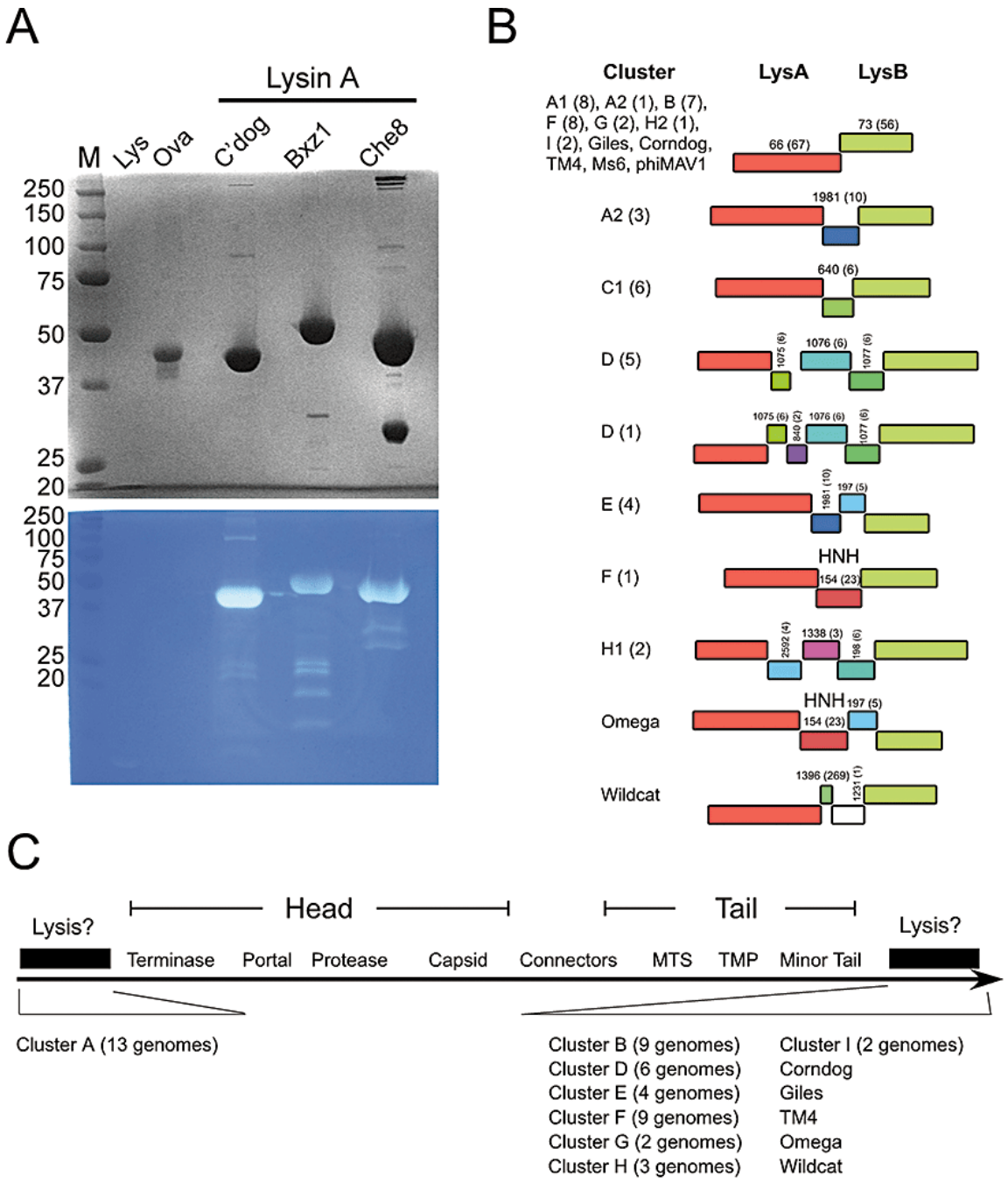
Results

Organization, conservation and location of mycobacteriophage lysis cassettes

Comparative analysis of mycobacteriophage genomes reveals them to be highly diverse with mosaic architectures (Hatfull et al., 2006; Pedulla et al., 2003). The 6858 predicted protein products can be assorted into 1523 phamilies (Phams) of related sequences and their distributions among the mycobacteriophages analyzed (Hatfull *et*

al., 2006) (our unpublished observations). While there are few Phams represented in all 60 completely sequenced genomes, one of these is the *lysA* phamily (Pham 66), and the LysA proteins are predicted from sequence comparisons to have peptidoglycan hydrolyzing activity (Garcia et al., 2002; Hatfull et al., 2006; Marinelli et al., 2008). We have confirmed this for three LysA proteins (Corndog gp69, Bxz1 gp236, Che8 gp32) all of which catalyze peptidoglycan hydrolysis in zymograms (**Fig. 3-1A**). The *lysB* gene is also implicated in lysis, primarily because of its linkage to *lysA* (**Fig. 3-1B**) and the demonstration of lipolytic activity by Ms6 LysB (Gil et al., 2008). LysB homologs are present in 56 of the 60 completely sequenced mycobacteriophage genomes and are located downstream of *lysA* and separated from it by no more than four intervening genes (**Fig. 3-1B**). Some of the intervening genes encode putative holins and exhibit holin-like function (e.g., D29 gene *11*, data not shown), whereas others (e.g., Omega gene *51*) code for putative homing endonucleases HNH motifs (**Fig. 3-1B**), or have homologs elsewhere in other mycobacteriophage genomes (e.g., Troll4 gene *37*).

Figure 3-1. Organization, location and conservation of mycobacteriophage lysis cassettes. A. Peptidoglycan hydrolysis by three LysA proteins. LysA proteins of Che8 (gp32), Bxz1 (gp236) and Corndog (gp69) were expressed, purified to near homogeneity, separated by SDS-PAGE and stained with Coomassie blue (top panel). In the lower panel, the proteins were separated through a gel matrix containing *Micrococcus luteus*, renatured, and the zymogram developed by staining with methylene blue. Peptidoglycan hydrolysis by renatured proteins within the gel produces clear areas that no longer stain with methylene blue. Lysozyme (Lys) and ovalbumin (Ova) represent positive and negative controls. **B**. Each of the 60 completely sequenced mycobacteriophages together with Ms6 and a putative *M. avium* prophage encode LysA proteins predicted to hydrolyse peptidoglycan, although they have highly modular structures (Hatfull et al., 2006). When present, the second putative lysis gene *lysB* is located downstream and is closely linked to *lysA*, with no more than four intervening genes. The 60 sequenced genomes can be grouped into clusters and subclusters according to their similarities (our unpublished observations) and the organization in each cluster (with the number of genomes represented in parentheses) is shown. All of the mycobacteriophage genes have been assorted into phamilies according to their sequence similarities, and the pham number (and number of gene members in that pham in parentheses) shown above each gene. **C**. Mycobacteriophage lysis cassettes may be positioned either to the left of the virion structure and assembly genes, as in the 13 cluster A genomes, or to their right as in other genomes, but in either case the *lysA* and *lysB* genes are closely linked.


A
Lysin A
M
Lys
Ova
C'dog
Bxz1
Che8
250
150
100
75
50
37
25
20
B
Cluster
LysA
LysB
A1 (8), A2 (1), B (7), F (8), G (2), H2 (1), I (2), Giles, Corndog, TM4, Ms6, phiMAV1
66 (67)
73 (56)
A2 (3)
1981 (10)
C1 (6)
640 (6)
D (5)
1075 (6)
1076 (6)
1077 (6)
D (1)
840 (2)
E (4)
197 (5)
F (1)
HNH
154 (23)
H1 (2)
2592 (4)
1338 (3)
198 (6)
Omega
Wildcat
1396 (269)
1251 (1)
C
Lysis?
Head
Tail
Terminase
Portal
Protease
Capsid
Connectors
MTS
TMP
Minor Tail
Cluster A (13 genomes)
Cluster B (9 genomes)
Cluster D (6 genomes)
Cluster E (4 genomes)
Cluster F (9 genomes)
Cluster G (2 genomes)
Cluster H (3 genomes)
Cluster I (2 genomes)
Corndog
Giles
TM4
Omega
Wildcat

Although *lysA* and *lysB* are closely linked, this linkage is not a simple consequence of synteny in the context of the broader genome organizations, because these presumptive lysis cassettes are situated in different chromosomal locations (**Fig. 3-1C**). Of the 53 phages with a siphoviral morphotype, the thirteen genomes constituting Cluster A [including L5, D29, Bxb1 (Ford et al., 1998; Hatfull and Sarkis, 1993; Mediavilla et al., 2000) and Che12 (Hatfull *et al.*, 2006)] have the lysis cassette situated to the left of the virion structural and assembly genes, whereas all of the other genomes have the lysis cassette positioned to the right of the structural genes (**Fig. 3-1C**). The invariable linkage of *lysA* and *lysB* regardless of genome location further supports a role for LysB in lysis.

Sequence alignment of the phamily of LysB proteins shows that they are highly diverse, and only three residues are completely conserved (**Fig. B-1**). Although the proteins vary in length [from 254 (D29 gpl2) to 451 (PG1 gp50) residues] and there are many gaps throughout the alignment, these proteins do not have modular constructions as seen in the LysA proteins (Hatfull *et al.*, 2006). However, the sequence divergence is sufficiently high that some members of the phamily have little identifiable sequence similarity (<20% identity) to each other (**Figs B-1** and **B-2**). In a search for conserved domains, 14 LysB proteins (including D29 gp12) are predicted to contain a cutinase domain (pfam01083), and Wildcat gp52 has a predicted esterase domain (COG0400). Eight proteins have a predicted peptidoglycan-binding domain (pfam01471) at their N-terminus, although this is not present in D29 LysB (**Fig. B-1**).

The three absolutely conserved residues include the putative active site serine (position 82 in D29 LysB) and a GXP motif located ~40 residues downstream (residues 117-119 in D29 LysB; **Fig. B-1**). The cutinase- and lipase-like conserved domains suggests that these enzymes are serine esterases, which typically have a Ser-Asp-His catalytic triad. Of three aspartic acids and a histidine residue previously implicated in catalysis (Gil *et al.*, 2008) through their conservation, only the aspartic acid corresponding to position 166 in D29 LysB is highly conserved among the larger group of proteins, with the single departure being the LysB protein encoded by a putative *Mycobacterium avium* prophage that has a glutamic acid residue at that position (**Fig. B-1**). The alignment does not reveal a well-conserved candidate for the histidine component of the catalytic triad (**Fig. B-1**). The GXP motif is not absolutely conserved in all serine esterases and its role in LysB functions is not clear.

D29 LysB has lipolytic activity

To further characterize the structure and function of mycobacteriophage LysB proteins, several Pham73 members were cloned and expressed. Although expression and solubility varies among these, we found that the 254-residue D29 gp12 (LysB) expressed well and was readily purified to near homogeneity and high solubility (> 10 mg ml^{-1}) (**Fig. 3-2A**). D29 LysB shares only 40% amino acid sequence identity with the previously characterized Ms6 LysB protein (Gil et al., 2008), and lacks 90 N-terminal residues present in Ms6 LysB (**Fig. B-1**). We also constructed, expressed and purified a

mutant version of the protein with a substitution of alanine for the putative catalytic serine at position 82 (S82A).

To test for lipolytic activity of D29 LysB we measured hydrolysis of *p*-nitrophenyl butyrate (*p*NPB) to release *p*-nitrophenol (Gilham and Lehner, 2005) (**Fig. 3-2B**). We observed a specific activity of 0.72 U mg^{-1} (**Fig. 3-2B**), somewhat higher than the 0.12 U mg^{-1} reported for Ms6 LysB protein (Gil et al., 2008), or for any of the seven cutinase-like proteins found in *Mycobacterium tuberculosis* (West et al., 2009). The S82A mutant is inactive, consistent with this residue being part of the catalytic triad (**Fig. 3-2B**). We also tested D29 LysB, as well as the purified lipase from *Pseudomonas fluorescens* for activity on *p*-nitrophenyl substrates with different carbon chain lengths. While D29 LysB was more active than the lipase on *p*NPB, we also observed decreasing activity of D29 LysB with longer substrates (**Fig. 3-2B**), as was reported for Ms6 LysB(Gil et al., 2008). D29 LysB thus shares lipolytic activity with Ms6 LysB (Gil et al., 2008), with serine-82 providing critical catalytic functions.

Figure 3-2. Purification and lipolytic activity of D29 LysB.

A. Phage D29 LysB was expressed from plasmid pLAM3, a derivative of vector pET21a, in *E. coli* BL21(DE3) and purified to near-homogeneity. SDS-PAGE of un-induced cells (U), whole cell lysates of induced strains (W) separated into insoluble (I) and soluble (S) fractions, and a clarified soluble lysate (L) are shown. The 30 kDa His-tagged D29 LysB was bound to a cobalt-affinity matrix, and flow-through (F), a 20 mM imidazole wash (W_{20}) sample, and fractions collected at 120 mM imidazole elutions (E_{120}) are shown.

B. Lipolytic activity of D29 LysB (filled bars; left axis scale) is shown using *p*-nitrophenyl substrates containing different lipid chain lengths (C_4, C_8, C_{12} and C_{16}); lipase activities on the same substrate are also shown (open bars; right axis scale). Specific activities are shown as Units/mg protein, with 1 Unit corresponding to the release of 1 μmol *p*-nitrophenol min^{-1}. An active site mutant (S82A), boiled sample, and a mock-purified sample show little or no activity.

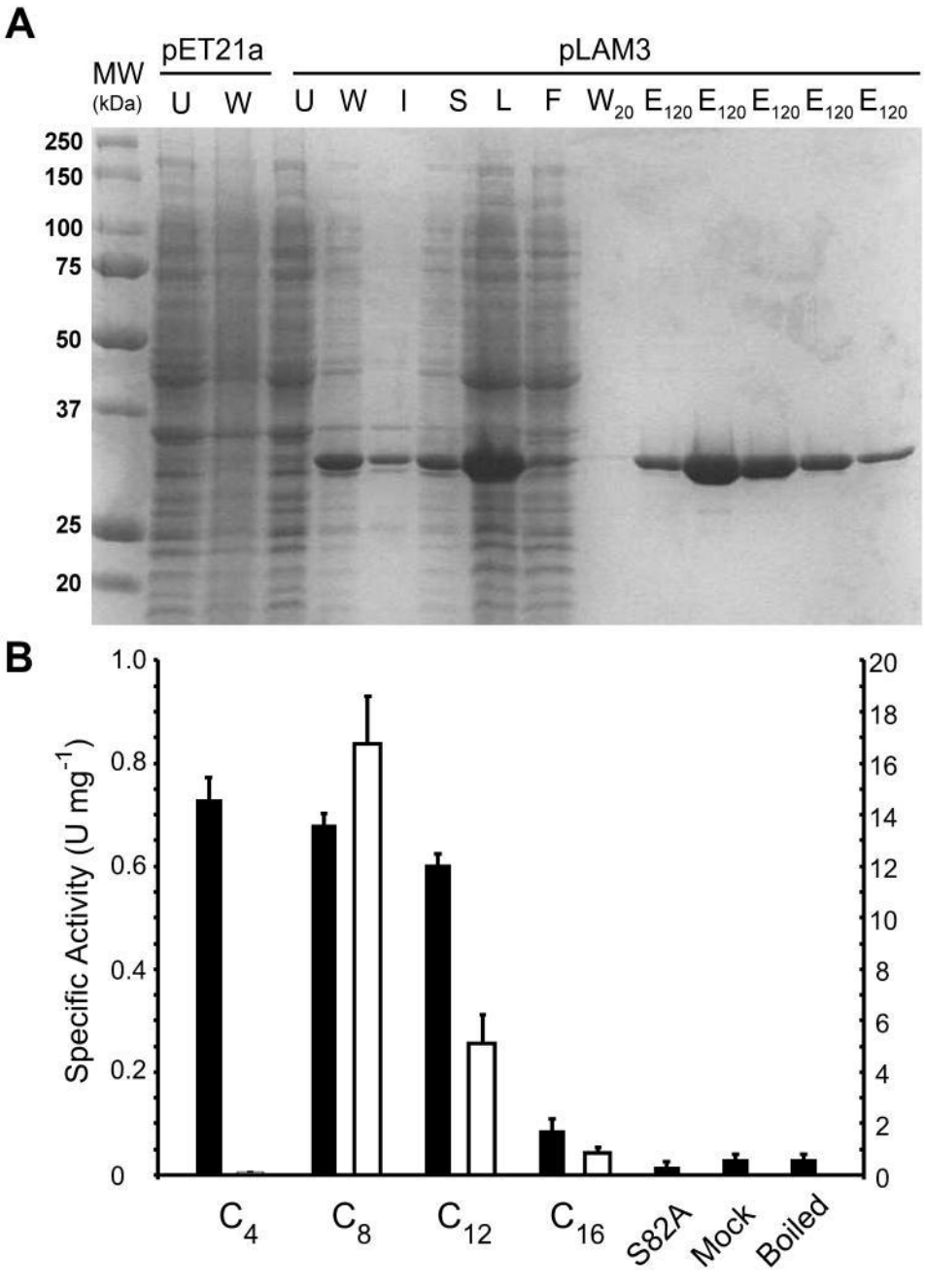

A
pET21a
pLAM3
MW
(kDa)
U W U W I S L F W20 E120 E120 E120 E120 E120
250
150
100
75
50
37
25
20
B
Specific Activity (U mg-1)
1.0
0.8
0.6
0.4
0.2
0
20
18
16
14
12
10
8
6
4
2
0
C4
C8
C12
C16
S82A
Mock
Boiled

Table 3-1. Data collection and refinement statistics of D29 LysB crystal

Data collection[a]	
Space group	$P\,4_3\,2_1\,2$
Cell dimensions	
a, b, c (Å)	92.5, 92.5, 69.6
Wavelength (Å)	0.9795
Resolution (Å)	2.0 (2.07-2.00)
I/sigI	28.0 (3.2)
Completeness (%)	99.8 (99.5)
Rsym	8.3 (31.2)
Redundancy	14.1 (12.8)
Refinement[b]	
No. reflections	19777
Rwork/Rfree	0.205/0.250
No. atoms	2094
RMSD bonds	0.004
RMSD angles	0.635

a. The native D29 LysB dataset was collected at APS 23ID and processed with HKL2000.
b. Refinement was done with Phenix.

Accession code: The atomic coordinate and structure factor have been deposited in the PDB with the accession code 3HC7 for D29 LysB.

Crystal structure of D29 LysB

The structure of D29 LysB was determined at 2.0-Å resolution in crystals of the *P* 4_3 2_1 2 space group containing a single molecule in the asymmetric unit (**Table 3-1**). Structural alignment with known protein structures using Dali server showed that D29 LysB is similar to members of the α/β hydrolase family, which includes cutinases, acetyl xylan esterases and lipases (Holm et al., 2008; Masaki et al., 2005). The five closest structural relatives of D29 LysB are all members of the cutinase family (**Table 3-2**), although they share no more than 21% amino acid sequence identity with D29 LysB.

Table 3-2. Top 5 results of structural alignment of D29 LysB in Dali database[a]

pdb	Z-score	rmsd	#aligned position	#residue	%identity	Description
2czq	20.3	2.2	169	205	21	Cutinase-like protein
1qoz	19.2	2.6	170	207	21	Acetyl xylan esterase
1bs9	18.9	2.3	164	207	21	Acetyl xylan esterase
3dd5	16.5	2.7	160	193	19	Cutinase
1ffb	16.4	2.7	157	197	20	Cutinase

a. The redundant hits of the same protein have been removed.

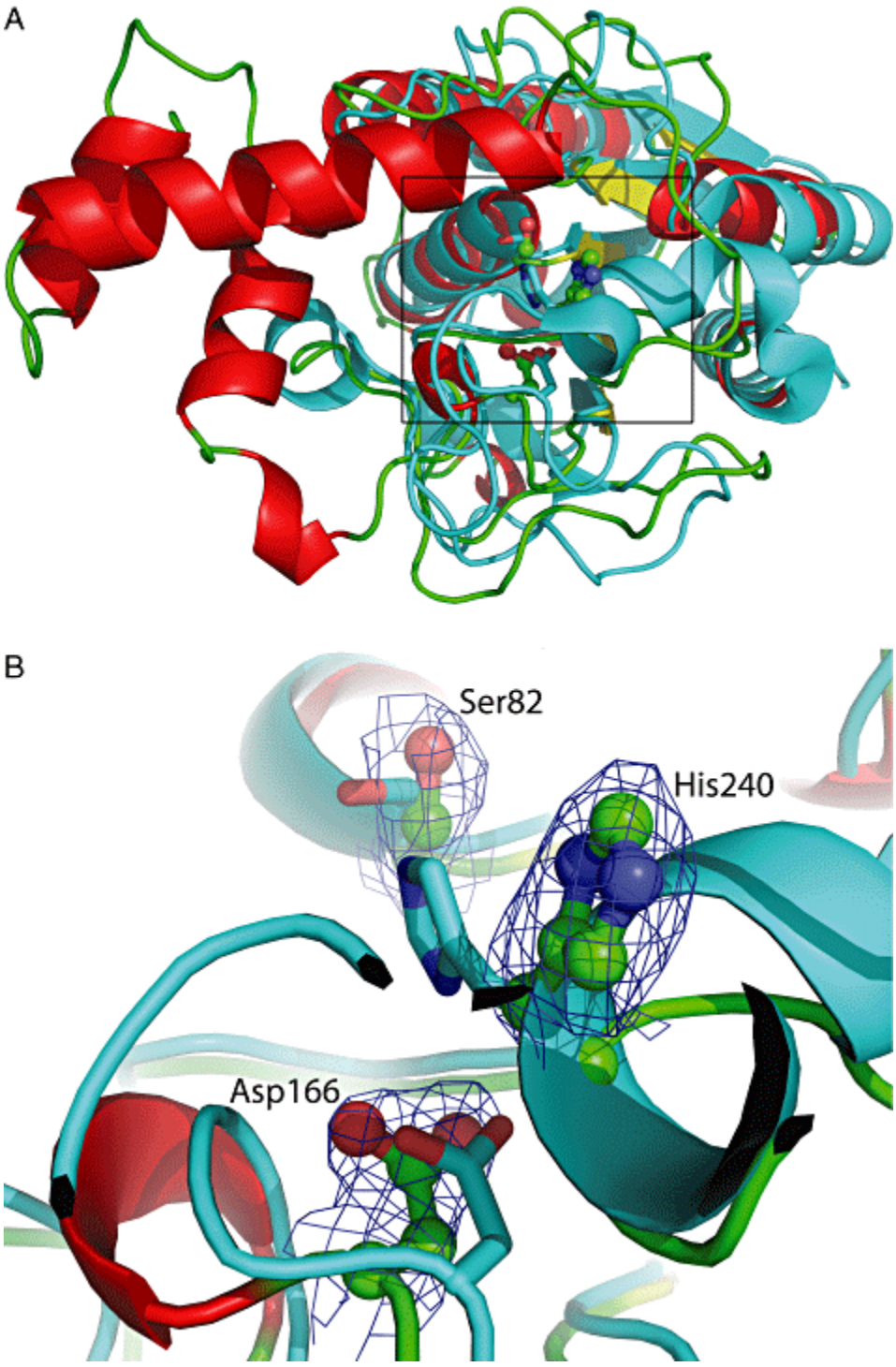
Of the 253 determined residues in D29 LysB, 170 residues (Ser2-Tyr161, Arg244-Gln253) are distributed in a doubly wound α/β sandwich, having a central parallel β-sheet flanked by two parallel α-helices on each side (**Fig. 3-3A**). This fold is typical of the α/β hydrolase family and the structure of Lysin B is remarkably similar to that of a *Cryptococcus* cutinase-like protein (**Fig. 3-3A**). Despite their low sequence

similarity, the rms differences in Cα positions are 2.45 Å and a search using the Dali server gave a high Z-score (20.3) when LysB was submitted as the query. The remainder of the protein (Ala162-Asn243) forms an 81-residue linker domain composed of four α-helices connecting the C-terminal α-helix (Arg244-Gln253) back to the side of the central β-sheet. It occupies the general location where the acyl chain is positioned in co-crystals of the *Fusarium solani* cutinase with a covalently-linked n-undecanyl-phosphonate methyl ester inhibitor (Longhi et al., 1996). Given the high B-factor and poor electron-density of the Arg231-Lys237 loop connecting the helical-linker and the C-terminal helix, it is reasonable to propose that movement of the helical-linker helps to modulate the active site to accommodate bulky and hydrophobic acyl chains. We note that the unusual glycine-rich segment corresponding to a 55-amino acid segment of PG1 LysB (and related proteins) inserted between D29 LysB residues 199 and 200, of which 19 are glycine residues, is located within this linker between the second and third helices (**Fig. 3-3A; Fig. B-1**).

Figure 3-3. Crystal structure of D29 LysB.

A. The D29 LysB structure is shown in alignment with the *Cryptococcus sp* cutinase-like protein (pdb: 2czq). D29 LysB is colored according to secondary structure (red: α-helix, yellow: β-strand, green: loops). 2czq is shown in cyan.

B. The catalytic triad of D29 LysB catalytic triad is composed of Ser82, Asp166 and His420 (ball-and-stick representations). Electron density map (2Fo-Fc) around the triad is shown in blue mesh at 1σ level. The triad of 2czq is in stick.


A
B
Ser82
His240
Asp166

The catalytic triad Ser82-Asp166-His240 is located at the edge of the central β-sheet between the α/β sandwich and the linker domain. The position of the D29 LysB catalytic triad is very similar to those in other members of the α/β hydrolase family as can be seen in the superimposition of D29 LysB on the *Cryptococcus* cutinase-like protein (**Fig. 3-3B**). Although Ser82 is invariant and Asp166 is very strongly conserved in the alignment of LysB proteins (**Fig. B-1**), the sequence alignment around His-240 varies among the LysB members, reflecting poor conservation of the flanking residues and positioning of the catalytic histidine within a loop region (**Fig. B-1**). In contrast, the GXP motif – of which the glycine and proline residues are invariant – lies immediately adjacent to the catalytic triad. In D29 LysB, the Gly117-Asn118-Pro119 segment defines the end of β4 strand and makes a 180 degree turn just beneath Asp166. Besides the turn-makers Gly117 and Pro119, Asn118 forms hydrogen bonds with Met120, Arg121 and Asp160, which may be energetically favorable to the scaffold.

D29 LysB is a mycolylarabinogalactan esterase

The D29 LysB structure confirms its enzymatic function as a serine esterase, and while it clearly has lipolytic activity, the unusual linker positioned adjacent to the active site – together with the inclusion of *lysB* genes within lysis cassettes – suggests that its substrate could be a cell wall component containing ester-linked lipids. We reasoned that an excellent candidate substrate is mycolyl-arabinogalactan–peptidoglycan (mAGP), such that LysB hydrolysis would release free mycolic acids and separate the outer membrane from the arabinogalactan-peptidoglycan layer. To test this we isolated a cell

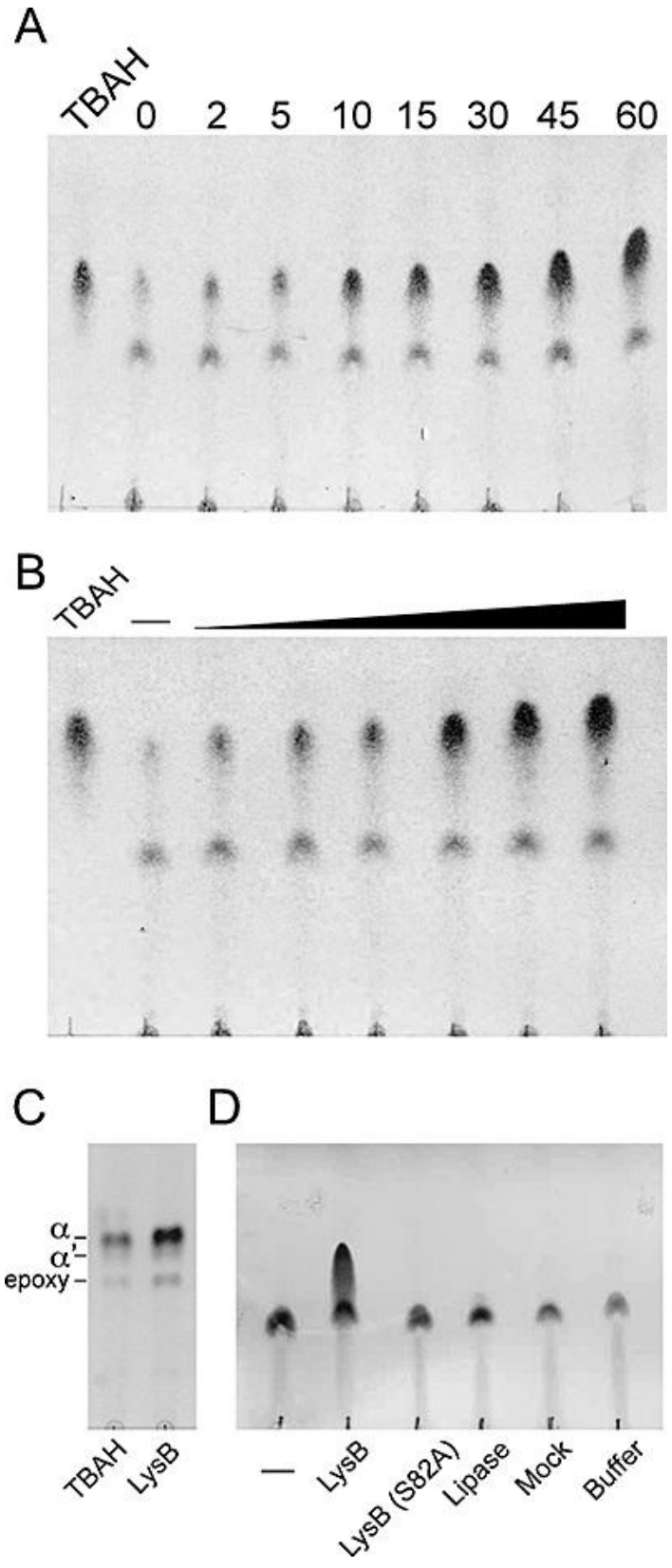
wall fraction from *Mycobacterium smegmatis* that is highly enriched in mAGP (Besra, 1998) and determined whether it is a substrate for D29 LysB hydrolysis (**Fig. 3-4A, B**). We observed both time- and enzyme concentration-dependent release of lipids that migrate similarly to mycolic acids released from the same mAGP preparation by alkaline hydrolysis of mAGP with tetrabutylammonium (TBAH) (Besra, 1998; Hamid et al., 1993; Watanabe et al., 2001). To further characterize the lipid products we methyl-esterified them with iodomethane, which yielded methyl esters of α, α' and epoxy mycolates similar to those from TBAH treatment and specific to *M. smegmatis* (**Fig. 3-4C**). Preliminary analysis of the released lipids by mass spectrometry and NMR was consistent with them being free mycolic acids (data not shown). To our knowledge, this is the first mycolylarabinogalactan esterase to be described. As expected, hydrolysis of mAGP by D29 LysB is dependent on the catalytic serine residue at position 82 and the S82A mutant enzyme exhibits no observable activity (**Fig. 3-4D**). Moreover, mAGP was not a substrate for a *Pseudomonas*-derived lipase (Arpigny and Jaeger, 1999; Gupta et al., 2004) as no release of free mycolic acids was observed (**Fig. 3-4D**).

Figure 3-4. D29 LysB is a mycolylarabinogalactan esterase.

A and **B**. Time- and enzyme-concentration-dependence of D29 LysB-mediated hydrolysis of mycolyl-arabinogalactan-peptidoglycan (mAGP). mAGP was purified from *M. smegmatis*, incubated with D29 LysB, and the lipid products separated by thin layer chromatography. The reaction was monitored as a function of time (panel **A**; incubation times as shown, in minutes) or amount of protein (panel **B**; 0.1, 0.25, 0.5, 1, 5, 10 μg D29 LysB). D29 LysB catalyzes production of a species migrating in the same position as the free mycolic acids released by tetrabutylammonium (TBAH) treatment.

C. Methylesterification of the products generated by D29 LysB cleavage of mAGP generates epoxy, α and α' mycolates.

D. Neither a D29 LysB catalytic mutant (S82A), lipase purified from *Pseudomonas fluorescens*, a mock purified sample, nor buffer release free mycolic acids from mAGP.


A
TBAH
0
2
5
10
15
30
45
60
B
TBAH
—
C
α
α'
epoxy
TBAH
LysB
D
—
LysB
LysB (S82A)
Lipase
Mock
Buffer

Giles lysB *facilitates completion of host cell lysis*

To explore the role of LysB in mycobacteriophage lytic growth, we first asked whether it is an essential function for plaque formation, using a recently described recombineering strategy to delete *lysB* (gene *32*) from the mycobacteriophage Giles genome (Marinelli et al., 2008). We targeted Giles because it is a demonstrated substrate for recombineering (Marinelli et al., 2008), we know that Giles *lysA* is essential (Marinelli et al., 2008), and we anticipated needing to complement with a *lysB* gene that is sufficiently different genetically as to avoid recombination (e.g., D29 gene *12*). It should be noted, however, that Giles LysB – cloned and purified in the same manner as D29 LysB – was also able to hydrolyse mAGP (data not shown). A 200 bp substrate containing 100 bp sequences flanking *lysB* was designed to introduce a 1146 bp internal deletion in Giles *lysB*, fusing 15 codons at the 5′ and 3′ ends of the gene to minimize effects on expression of adjacent genes as well as avoiding genetic polarity (**Fig. 3-5A**). Following co-electroporation of the 200 bp deletion substrate and Giles genomic DNA into an *M. smegmatis* recombineering strain, plaques were recovered on a lawn of *M. smegmatis* mc^2155 pKMC2 cells expressing D29 LysB. Of 22 primary plaques screened by deletion amplification detection assay (DADA)-PCR (Marinelli et al., 2008), two contained mixtures of wild-type and mutant phages (**Fig. 3-5B**).

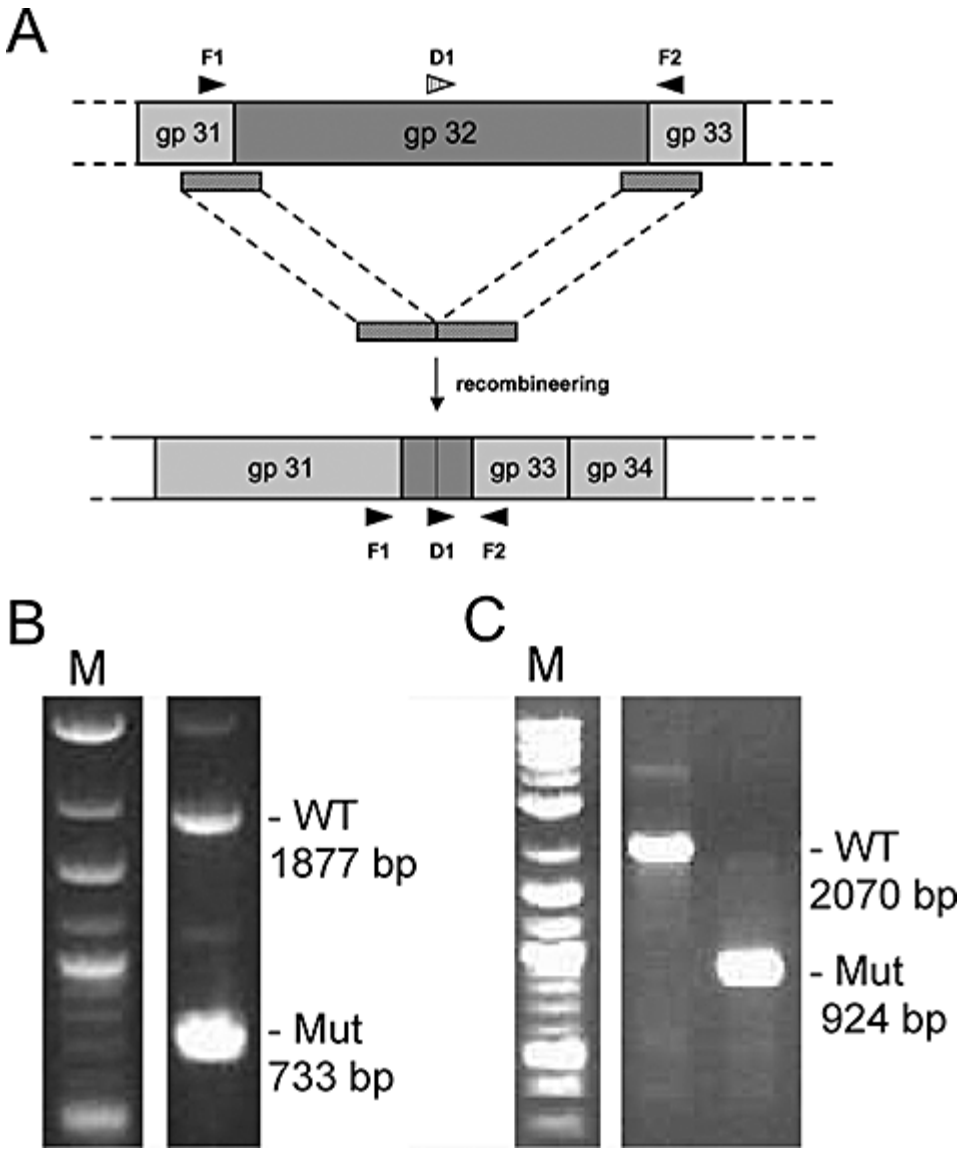


Figure 3-5. Construction of a Giles *lysB* deletion mutant. A. A 200 bp dsDNA substrate was designed that has 100 bp homology flanking a 1146 deletion in Giles gene *32*. The locations of primers F1, F2 and D1, used for PCR screening are shown. **B**. Following co-electroporation of the 200 bp substrate and Giles genomic DNA, primary plaques were recovered and screened by PCR using primers D1 and F2, designed to preferentially amplify the deletion mutant. A mixed plaque containing wild-type and mutant DNA is shown. **C**. A mixed primary plaque was diluted and plated, and isolated secondary plaques screened using primers F1 and F2. One homogenous wild-type Giles plaque (left) and one homogenous *lysB* deletion mutant plaque are shown (right).

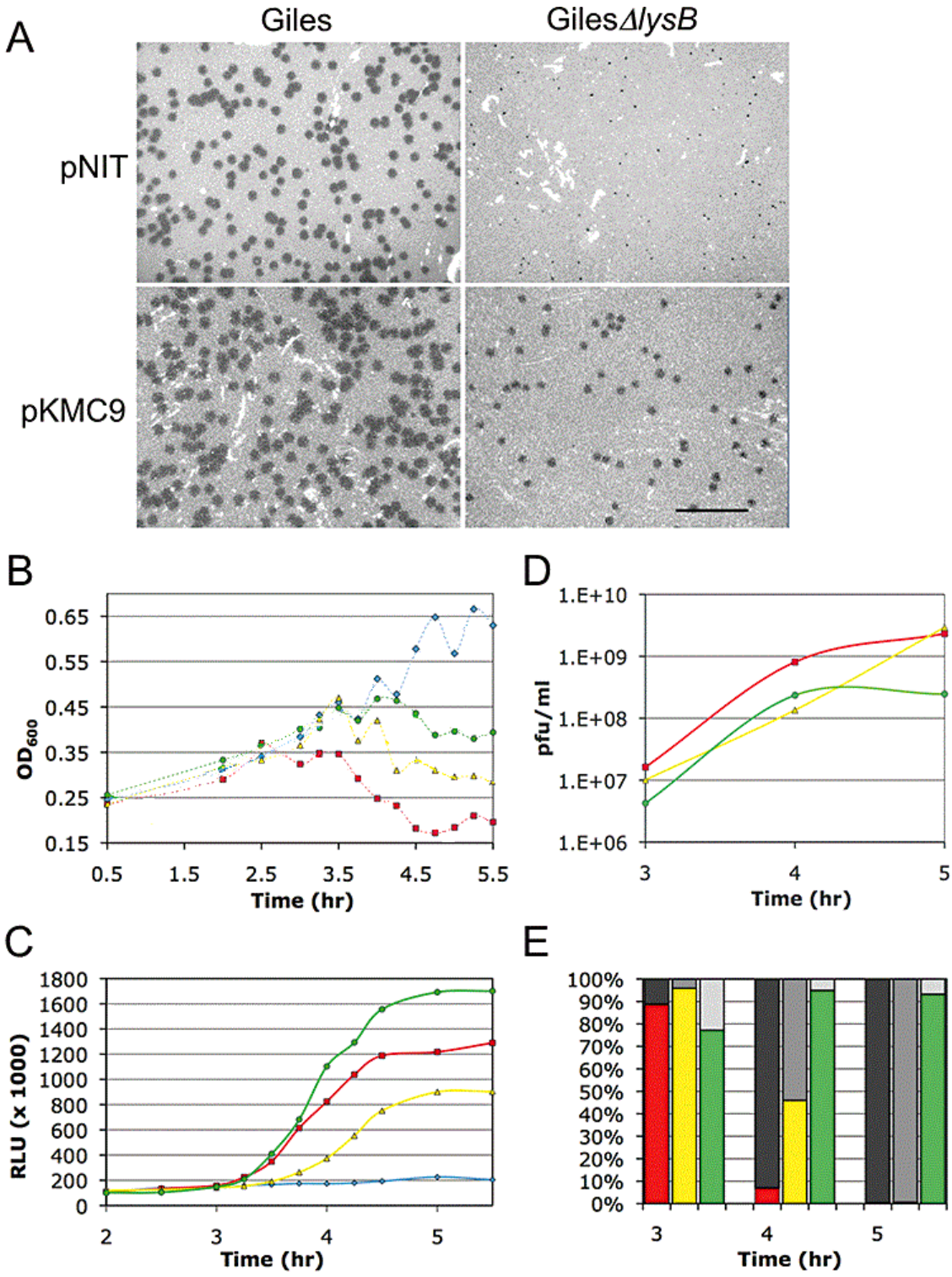
To test for viability of the Giles *ΔlysB* mutant, a mixed plaque was picked, resuspended in buffer, and plated to recover ~600 isolated plaques on both a putative complementing strain (*M. smegmatis* mc^2155pMKC2) or a non-complementing control strain. Secondary lysates representing all of the recovered particles were harvested and tested by PCR for the presence of the deletion mutant (data not shown). The mutant genotype was present in approximately equivalent proportions in both the complementing and non-complementing strains, suggesting that the mutant is viable. Thirty individual plaques from a secondary plating were then screened by PCR and a homogenous mutant plaque identified (**Fig. 3-5C**). The Giles *ΔlysB* mutant was plaque purified, and the structure of the deletion confirmed by DNA sequencing (data not shown).

The Giles *ΔlysB* mutant forms plaques at equivalent efficiencies on complementing and non-complementing strains, and titers of wild-type and mutant lysates prepared on a wild-type *M. smegmatis* host under standard conditions are similar; Giles *lysB* thus is not essential for lytic growth (**Fig. 3-6A**). However, we observed that the mutant forms somewhat smaller plaques on lawns of wild-type *M. smegmatis* compared to the parental Giles, a phenotype that is exaggerated when higher densities of plating cells are used (conditions that favor fewer bacterial doublings and fewer rounds of phage infection) (**Fig. 3-6A**). This difference is directly attributable to the loss of *lysB* since complementation with a plasmid expressing D29 *lysB* restores plaque size to near that of the wild-type parent (**Fig. 3-6A**). To eliminate the possibility that plaque size restoration results from acquisition of *lysB* by recombination with the complementing

plasmid, several plaques were picked and shown to reproduce the mutant phenotype, and by PCR contain only the mutant genotype (data not shown). Furthermore, at a relatively high cell plating density (2×10^8 cfu/plate) the average number of particles in each plaque is ~100-fold reduced in *ΔlysB* mutant relative to wild-type Giles plaques (5×10^5 and 4×10^7 pfu/ml respectively), consistent with a lysis defect. Therefore, unlike *lysA, lysB* is not essential for plaque formation but is required for efficient phage release.

To test whether the small plaque phenotype specifically results from changes in the pattern or timing of lysis, we monitored optical densities (OD) of infected *M. smegmatis* cultures (**Fig. 3-6B**). After infection with wild-type Giles the OD increases for approximately 3 hours, after which it steadily declines as the cells lyse, up to a period approximately 5 hours after infection. When infected with the *ΔlysB* mutant, the OD does not begin to decline until 3.5 hours and is incomplete even 5.5 hours after infection (**Fig. 3-6B**). In contrast, cells infected with a *ΔlysA* mutant — which does not form visible plaques on a non-complementing *M. smegmatis* strain (Marinelli et al., 2008) — cease to grow after 3-3.5 hours, with only a modest reduction in OD thereafter (**Fig. 3-6B**).

Figure 3-6. LysB is required for efficient host cell lysis. A. *M. smegmatis* mc^{2}155 containing either plasmid vector pNIT-1 or plasmid pKMC9 carrying the D29 *lysB* gene was infected with wild-type or Giles Δ*lysB* phages, as indicated. The Δ*lysB* mutant forms smaller plaques than its wild-type parent, but restored to larger plaques by complementation with D29 *lysB*. Scale bar indicates 1 cm. **B**. Lysis of phage-infected *M. smegmatis* mc^{2}155 was monitored by measuring OD_{600} of cultures infected with wild-type Giles (red), a Δ*lysA* mutant (green) or a Δ*lysB* mutant (yellow) at different times after phage addition; uninfected cells are shown in blue. Thicker lines correspond to trendlines. **C**. *M. smegmatis* mc^{2}155 cells were infected with Giles (red), a Δ*lysA* mutant (green) or a Δ*lysB* mutant (yellow) and ATP release measured at different times after infection; uninfected cells are shown in blue. ATP was measured in a luciferase assay and reported as relative light units (RLU). **D**. *M. smegmatis* mc^{2}155 cells were infected with wild-type Giles, the Δ*lysA* mutant, or the Δ*lysB* mutant and the total number of phage particles (pfu ml^{-1}) determined. Wild-type Giles is shown in red, Δ*lysA* in green, and the Δ*lysB* mutant in yellow. **E**. For each of the time points in D, the distribution (as percentage of total) of phage particles in the supernatant and cell pellet was determined. For wild-type Giles the supernatant is shown in dark grey and the cell pellet in red; the Δ*lysB* mutant supernatant is grey and the pellet is yellow; the Δ*lysA* mutant supernatant is shown in light grey and the pellet in green.


A
Giles
GilesΔlysB
pNIT
pKMC9
B
OD600
0.65
0.55
0.45
0.35
0.25
0.15
0.5
1.5
2.5
3.5
4.5
5.5
Time (hr)
D
pfu/ml
1.E+10
1.E+09
1.E+08
1.E+07
1.E+06
3
4
5
Time (hr)
C
RLU (x 1000)
1800
1600
1400
1200
1000
800
600
400
200
0
2
3
4
5
Time (hr)
E
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
3
4
5
Time (hr)

We also monitored the progression of phage infections by measuring ATP release (**Fig. 3-6C**), which we anticipated to reflect either lysis or permeabilization of the cell wall. In a wild-type Giles infection, little or no ATP release is seen until three hours after infection, followed by a steady increase to 4.5-5 hours after infection (**Fig. 3-6C**), closely mirroring the changes in OD shown in **Fig. 3-6B**. The *ΔlysB* mutant is delayed in the onset of ATP release by about 30 minutes, and fails to achieve the wild-type level even 5.5 hours after infection, which taken together with the OD changes is consistent with a lysis defect. The phenotype is reminiscent of the conditional lysis defect of phage lambda Rz or Rz1 mutants, where the Rz/Rz1 proteins are implicated in the final lysis step in which the *E. coli* outer membrane is compromised by fusion with the cytoplasmic membrane (Berry et al., 2008; Summer et al., 2007). Interestingly, the *ΔlysA* mutant shows no defect in ATP release at all, and may even release more ATP than cells infected with wild-type Giles (**Fig. 3-6C**).

Finally, we tested whether the *lysB* mutant produces the same yield of total phage particles as wild-type Giles in a lytic infection, and determined how the particles distribute between those that are released and those that remain trapped in unlysed cells (**Fig. 3-6D, E**). As shown in **Fig. 3-6D**, both the *ΔlysA* and *ΔlysB* mutants show no major defect in the production of phage particles. However, by 4 hours after infection, although >90% of wild-type particles are present in the culture supernatant, about 45% of the *ΔlysB* particles remain associated with unlysed cells. In contrast, <10% of *ΔlysA* particles are released into the supernatant even 5 hours after infection (**Fig. 3-6E**). These

observations are consistent with a strong defect in cell lysis in the *ΔlysA* mutant, and a milder defect in the *lysB* mutant.

Discussion

We have shown here that the mycobacteriophage D29 LysB protein is a novel mycolylarabinaogalactan esterase required for completion of lysis of host mycobacterial cells, and a model for the role of the LysA and LysB proteins as well as the reaction catalyzed by LysB is presented in **Figure 3-7**. The mycolic acid-rich mycobacterial outer membrane presents an unusual problem for phages of Gram-positive bacteria, which typically complete lysis through simple endolysin-mediated degradation of the peptidoglycan layer. The mycolylarabinoglactan linkage is not common among bacteria, since mycolic acids are found primarily in the *Corynebacterineae* suborder of the *Actinomycetales*, which includes *Corynebacteria*, *Gordonia*, *Nocardia*, *Rhodoccocci* and *Mycobacteria*. Few phages infecting non-mycobacterial members of the *Corynebacterineae* have been characterized, although these would be good candidates for also encoding mycolylarabinogalactan esterases. We note that neither phage P2101 of *Corynebacterium glutamicum* (Chen *et al.,* 2008) nor BFK20 of *Brevibacterium flavum* (Bukovska *et al.,* 2006) encodes a LysB relative, although unlike the *Mycobacteria*, mycolic acids are dispensable for growth of *C. glutamicum* (Portevin *et al.,* 2004) and therefore may not pose a barrier to efficient lysis.

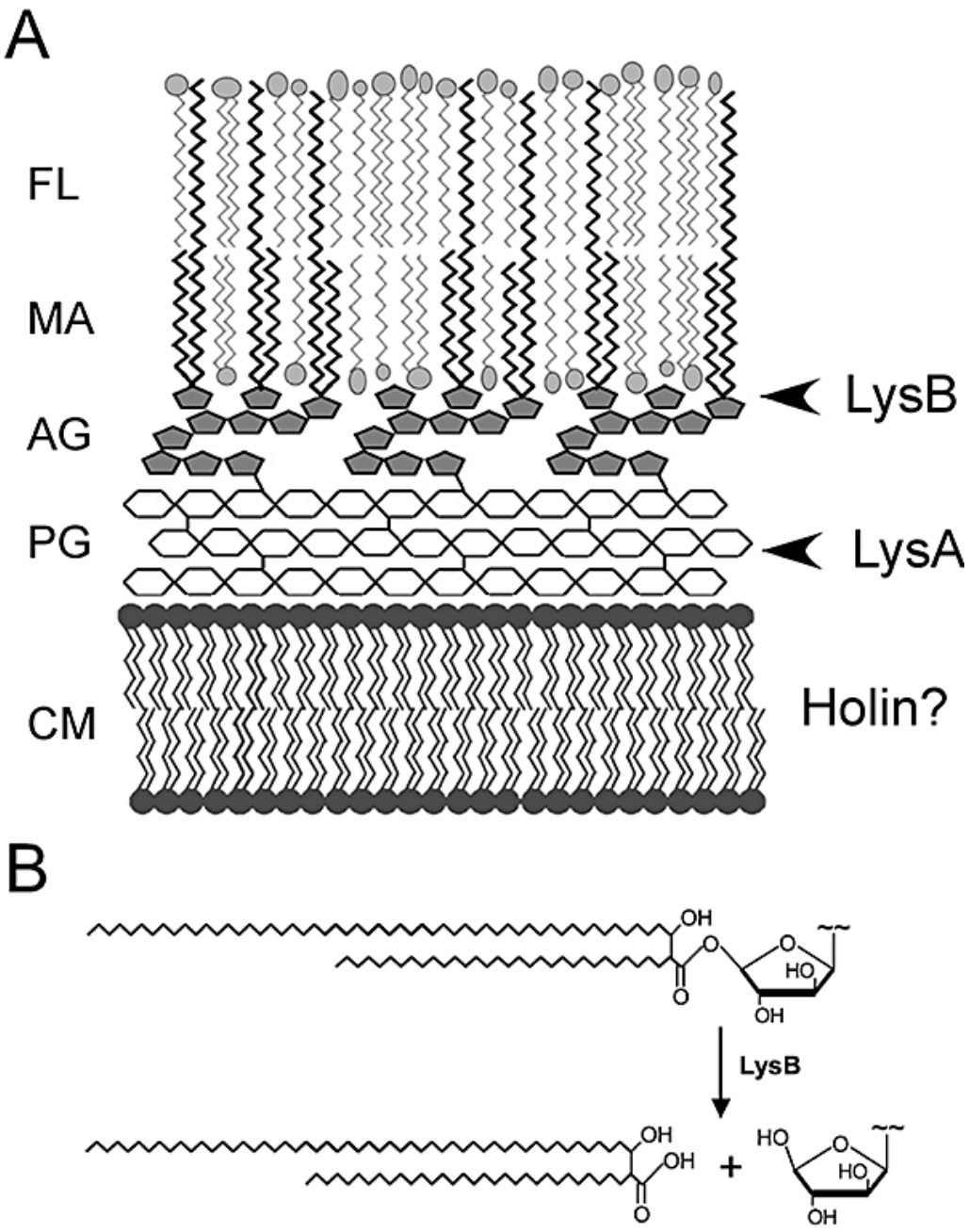


Figure 3-7. A model for mycobacteriophage lysis of mycobacteria. A. Mycobacterial cell walls are unusual in that the cytoplasmic membrane (CM) is surrounded by a peptidoglycan layer (PG) to which a network of arabinogalactan (AG) is covalently attached. A mycobacterial outer membrane consisting of mycolic acids (MA) and free lipids (FL) is covalently attached via an ester linkage of mycolic acids to arabinogalactan. We propose that LysA – assisted by holins encoded by at least some of mycobacteriophages – perform an essential step in lysis involving degradation of the peptidoglycan layer, and the lysis is completed through LysB-mediated cleavage of the outer membrane from arabinogalactan. **B**. Diagram illustrating LysB cleavage of the ester bond linking mycolic acids and arabinogalactan.

While removal of LysB function in mycobacterial Giles results in an apparently mild plaque phenotype that can be rescued by expression of D29 LysB, the problems in lysis are made more apparent by measuring changes in optical density of the culture, ATP release, and phage particle release. The ATP release by the Giles *ΔlysA* mutant is in contrast to its lysis defect as reflected in topical density measurements. A simple explanation for this observation is that the peptidoglycan layer — which most probably remains intact in the *ΔlysA* mutant infection — provides no barrier to ATP release at all, whereas both the cytoplasmic membrane and the mycobacterial outer membrane must be compromised for complete ATP release. Presumably, the cytoplasmic membrane is permeabilized through the action of a Giles holin (which has yet to be identified), and the differences observed between the *ΔlysA* and *ΔlysB* mutants are consistent with an attached mycobacterial outer membrane providing a significant barrier to ATP release. While it is difficult to eliminate the possibility of secondary effects of the deletion mutation in the *ΔlysB* mutant on expression of other phage genes (including a putative holin), the complementation studies (**Fig. 3-6A**) are consistent with loss of LysB function as the primary cause of the phenotypes observed.

At this time it is unknown how LysB is localized to its substrate, and no signal peptide motifs have been identified (Gil *et al.,* 2008). It would be interesting to see if LysB activity is dependent on a mycobacteriophage holin, similar to endolysins of other bacteriophage and presumably the LysA proteins of mycobacteriophage. However, few mycobacteriophage holins have been predicted bioinformatically, and even fewer functional studies have been conducted; Ms6 gp4, a predicted holin, was lethal when

over-expressed in *E. coli* (Garcia *et al.,* 2002); and experiments in our lab indicate s a potential holin function for D29 gp11 (L. Marinelli and G. F. Hatfull, unpublished observations).

The question arises as to whether phages Che12, Rosebush, Qyrzula and Myrna that lack *lysB* encode alternative enzymes performing analogous functions. These phages do not form a closely related group (**Fig. B-2**), although comparison of each with their closer phage relatives is informative. For example, in Che12 the surrounding genes are syntenic with those in phages L5, D29, Pukovnik and Bxz2 and it appears as though Che12 has simply lost *lysB* (**Fig. B-3A**). The organizations in Rosebush and Qyrzula depart from their close relatives in the inclusion of a gene immediately downstream of *lysA* that is lacking in other Cluster B phages, all of which do encode LysB (**Fig. B-3B**). This predicted protein is small (178 aa) and is a distant relative of the D29 putative holin (gp11). In Myrna, neither of the two orfs (*244*, *245*) immediately downstream of *lysA* are related to any other mycobacteriophage proteins (**Fig. B-3C**). Myrna gp244 does have similarity to the N-terminal segment of the *Rhodococcus* protein (RHA1_ro08121) that contains both peptidase and muramidase motifs in its C-terminus, and Myrna gp244 may play a yet undefined role in lysis. We also note that although *Proprionobacterium acnes* phage PA6 (Farrar *et al.*, 2007) and *Streptomyces* phage R4 (our unpublished data) encode proteins with sequences similar to mycobacteriophage LysA proteins, neither encodes a LysB relative. LysB thus appears to be restricted to those phages infecting hosts with mycolic acid outer membranes.

The release of free mycolic acids from LysB-treated cell walls strongly suggests that mAGP is the substrate for the enzyme. The cell wall preparations may contain additional lipid components, but other predominant mycolic acid-containing constituents – such as trehalaose dimycolate (TDM; cord factor) and trehalaose monomycolate (TMM) – are not cell wall linked and are not major components of the mAGP preparations (Besra, 1998; Brennan and Nikaido, 1995). Preliminary data suggest that D29 LysB can also hydrolyze TDM (A. Ojha, K. Payne and G. F. Hatfull, unpublished observations), but it seems unlikely that this has physiological relevance for lysis because TDM is not covalently attached to mycobacterial cells. We propose that cleavage of the mycobacterial outer membrane from the peptidoglycan-arabinogalactan layer is the primary role of LysB.

The only known physiological circumstances in which free mycolic acids are released from mycobacterial cells is during maturation of *M. tuberculosis* biofilms (Ojha *et al.*, 2008), although the source of biofilm mycolic acids is more likely to be TDM than mAGP (A. Ojha and G. F. Hatfull, unpublished observations). While seven cutinase-like proteins (culps) encoded by *M. tuberculosis* have been expressed and characterized (West et al., 2009), they are not evidently related to LysB and have not yet been tested for mAGP hydrolysis. Culp1 and Culp4 have optimal enzyme activity on short (butyrate) *p*-nitrophenyl substrates, and only Culp6 has significant activity on longer carbon chain substrates (West et al., 2009). Also, the Culp4 homolog in *M. smegmatis* has been shown to have phospholipase A activity, which is not found in cutinases

(Parker *et al.*, 2007). It remains to be seen whether there are any host-encoded enzymes that share mycolylarabinogalactan esterase activity with LysB.

The D29 LysB structure shows it is a member of the α/β hydrolase family with an α/β sandwich and an active site containing Ser, Asp and His. The five closest related structures are all cutinases, although there is no greater than 21% amino acid sequence identity with any of them. The LysB catalytic mechanism is expected to be similar to that for other serine-esterases, although the strong conservation of the GXP motif and its juxtaposition to the active site may play a critical role in its stabilization. However, the D29 LysB also contains an unusual four-helix domain that projects from close to the active site and is lacking in other cutinase-like protein structures. While the specific role of the large linker is not known, its absence in all other cutinase-like proteins and lipases and its proximity to the active site would suggest a role in binding of the D29 LysB substrate. This linker contributes to the generally larger size of LysB proteins (254-451 residues) compared with other cutinases (~225 residues) (Carvalho et al., 1999; Longhi and Cambillau, 1999).

Acquisition of *lysB* by mycobacteriophages throughout their evolution likely confers a substantial selective advantage over those without it by providing faster and more complete lysis. While considered more similar to Gram-positive bacteria, the existence of the mycobacterial outer membrane composed of mycolic acids and free lipids (**Fig. 3-7A**) presents a second barrier analogous to the outer membrane of Gram-negative bacteria (Hoffmann et al., 2008; Zuber et al., 2008). Phages that infect Gram-negative bacteria have the help of spanin proteins and their Rz/Rz1counterparts (Berry et

al., 2008; Krupovic et al., 2008; Summer et al., 2007) to span the periplasmic space and link the inner and outer membrane, enhancing the completion of lysis. We propose that mycobacteriophage have developed an alternative solution to compromise the mycobacterial outer membrane, by severing its linkage to the underlying arabinogalactan-peptidoglycan layer.

Exogenously applied phage-encoded endolysins have been shown to have effective antimicrobial activity against Gram-positive bacterium pathogens including *Streptococcus pneumoniae* and *Bacillus anthracis* (Fischetti, 2008). They are, however, ineffective against Gram-negative bacteria since the outer membrane blocks access to the peptidoglycan targets. The mycobacteria are likely to be similarly intractable to exogenously added endolysins because of their mycolic acid-rich outer membrane. While we do not yet know whether mycobacteriophage LysB proteins can reach their mycolylarabinogalactan targets from the outside, this possibility raises the intriguing idea that mycobacterial pathogens such as *M. tuberculosis* may be rendered susceptible to endolysin treatment through co-treatment with LysA and LysB proteins.

Experimental procedures

Bacterial strains and growth

M. smegmatis mc^2155 and the recombineering strain containing pJV53 expressing the Che9c genes *60* and *61* under an acetamide-inducible promoter have been described previously (Parish et al., 1997; Snapper et al., 1990; van Kessel and Hatfull, 2007). All *M. smegmatis* strains were cultured in Middlebrook 7H9 medium or grown on

Middlebrook 7H10 agar supplemented with 10% Albumin Dextrose Complex (ADC), 0.2% succinate, 0.05% Tween-80, 1 mM Ca_2Cl, carbenicillin (50 μg/ml), cyclohexamide (10 μg/ml), and kanamycin (20 μg/ml) as required. *E. coli* was grown in LB supplemented with carbenicillin (50 μg/ml) and kanamycin (20 μg/ml) as needed, with *E. coli* GC5 cells (Stratagene) used for cloning and *E. coli* BL21(DE3) cells (Stratagene) for protein overexpression.

Plasmid constructions

Plasmids pET21a (Novagen), pLAM12, pJV53, and pNIT-1 have been previously described (Pandey et al., 2009; van Kessel and Hatfull, 2007). Plasmid pLAM3 was constructed by amplifying the D29 gene *12* (*lysB*) with primers bearing NdeI and HindIII restriction sites and cloning into pET21a to add a C-terminal His tag. To create the inducible D29B active site mutant (pKC20), the pLAM3 plasmid was subjected to site-directed mutagenesis using the QuikChange XL Site-Directed Mutagenesis Kit (Stratagene) to change Ser82 to Alanine. Plasmid pKMC2 was made by amplifying D29 *lysB* from pLAM3 with primers containing NdeI and NheI restriction sites and cloning this into pLAM12. Plasmid pKMC9 was constructed by amplifying D29 *lysB* with primers containing NdeI and HindIII restriction sites and cloning this into pNIT-1. Plasmids were purified using QIAprep Spin Miniprep Kit (QIAGEN). Oligonucleotides were supplied by Integrated DNA Technologies, Inc. and are listed in **Table B-1**.

Expression and purification

E. coli BL21 λ(DE3) carrying plasmids pLAM3 or pKC20 (or vector pET21a) were grown to OD_{600} between 0.4-0.6 in L-Broth (LB) containing carbenicillin (50 μg/ml) and induced with 1 mM IPTG for 4 hours. Cells were resuspended in TWEB (50 mM Tris-HCl pH 8.0, 300 mM NaCl), sonicated, centrifuged, and the clarified lysates passed over TALON Co^{2+} resin (Clonetech). The resin was washed sequentially with TWEB containing 10 mM and 20 mM imidazole, and bound protein eluted with five volumes of 120 mM imidazole. Eluted fractions were concentrated using Vivaspin concentration columns (molecular weight cutoff 10 kDa; Sartorius Inc.) followed by dialysis against storage buffer (50 mM Tris pH 8.0, 50 mM NaCl, 50% glycerol) and stored at −20°C.

D29 LysB was purified for crystallization similarly. *E. coli* BL21 λ(DE3)pLAM3 cells were induced for 18 hours at 25°C, pelleted, resuspended in Buffer A (50 mM Tris, 500 mM NaCl, 5 mM imidazole, pH 8) and disrupted in a French pressure cell three times at 10,000 psi for 3 passages. The cell lysate was cleared by centrifugation at 15,000 rpm for 30 min and applied to a 5 ml HisTrap FF column (GE Healthcare, Piscataway, NJ), and eluted with a linear gradient to 100% Buffer B (50 mM Tris, 500 mM NaCl, 500 mM imidazole, pH 8). Fractions were pooled and dialyzed against buffer containing 20 mM Tris, 50 mM NaCl, pH 8. D29 LysB was concentrated to at least 10 mg/ml.

Zymogram assays

Zymograms were performed as described previously (Piuri and Hatfull, 2006) by

incorporation of 0.2% lyophilized *Micrococcus luteus* cells as a source of peptidoglycan into the gel matrix. Zymograms were developed by renaturation overnight at 37°C in 25 mM Tris (pH 7.5), 1% Triton X-100, and 0.1 mM $ZnSO_4$, stained with 0.5% methylene blue with 0.01% KOH before destaining with water.

Lipolytic enzyme assays

Enzymatic assays for lipolytic activity were adapted from those described previously (Gilham and Lehner, 2005). One milliliter of *p*-nitrophenyl substrates (50 μM) (Sigma Inc.) were incubated with 1 μg of D29 LysB, D29 LysB S82A, lipase (*Pseudomonas fluorescens*, Sigma), or 5 μl of a mock purified sample (derived from pET21a containing cells) in buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.1% Triton X-100) at room temperature for 30 minutes. Release of *p*-nitrophenol was determined by measuring absorbance at 420 nm (A_{420}).

Assays for hydrolysis of M. smegmatis *mAGP*

Mycolyl-arabinogalactan-peptidoglycan was isolated from *M. smegmatis* as described previously (Besra, 1998). *M. smegmatis* cells were grown, collected, and washed three times in phosphate buffered saline (PBS; pH 7.4), and resuspended in PBS+2% Triton X-100 (PBSX). Cells were disrupted by extensive sonication, centrifuged to collect the insoluble cell wall fraction, resuspended in PBSX and agitated overnight at 4°C. After centrifugation the pellet was resuspended in PBS+2% SDS and

incubated at 100°C for 60 minutes; this was done three times. After three rounds of extraction, the pellet was washed once each with H_2O, 80% acetone in H_2O, and acetone. Following evaporation of the acetone, the mAGP-enriched cell wall material (mAGP) was resuspended in PBS+0.1% Triton X-100 (final concentration 10 mg/ml) and frozen aliquots stored at −80°C.

The mAGP material was chemically hydrolyzed by addition of an equal volume of 15% tetrabutylammonium (TBAH) to 1 mg mAGP resuspended in PBS+0.1% Triton X-100 and incubated at 100°C overnight. Enzymatic assays were performed by incubation of 1 mg mAGP in 100 µl with varying concentrations of protein at 37°C. TBAH- and enzyme-treated samples were prepared for analysis by thin layer chromatography by addition of an equal volume of dichloromethane and incubation for 15 minutes at room temperature. The lipid-rich lower dichloromethane layer was removed, extracted once with 0.25 M HCl and once with water, and lipids collected by evaporation. Samples were resuspended in dichloromethane, spotted onto silica-aluminum TLC plates, and separated by chromatography in chloroform/methanol (97:3); lipids were identified by spraying with 5% molybdophosphoric acid (in ethanol) and charring for 15 minutes at 110°C.

Methyl-esterification of lipids was performed by resuspension of enzyme reaction products in 15% TBAH, addition of an equal volume of dicholoromethane and 1/10 volume iodomethane. Reactions were incubated with shaking at room temperature for 15 minutes and the lower dichloromethane layer recovered and extracted with HCl

and water as described above. Lipids were separated by TLC in 95:5 petroleum ether/ethyl ether and recognized by charring as above.

Bacteriophage recombineering of electroporated DNA

A *lysB* deletion mutant of phage Giles was constructed as described previously (Marinelli et al., 2008). The targeting substrate was generated using a 100 bp oligonucleotide with 50 bp of homology upstream and downstream of the deletion site, which was then expanded to a 200 bp dsDNA by PCR. The deletion was deigned to remove 1,146 bp between Giles genome coordinates 28,384 and 29,529. Phage Giles DNA (350 ng) was co-electroporated with 200 ng of the 200 bp substrate into induced electrocompetent *M. smegmatis* mc^2155 pJV53 cells, recovered at 37°C for 2 hours and plated on top agar lawns with *M. smegmatis* mc^2155 pKMC2 cells in the presence of 0.2% acetamide; approximately 100 plaques were recovered. Individual plaques were picked into 100 µl phage buffer (10 mM Tris-HCl, pH 7.5; 10 mM $MgSO_4$, 68.5 mM NaCl; 1mM $CaCl_2$) and analyzed by Deletion Amplification Detection Assay (DADA)-PCR (Marinelli et al., 2008) using 1 µl sample and primers GilesB-DiagR3 and GilesB-DADAPCR3 (**Table B-1**). Primary plaques containing both wild-type and mutant alleles were diluted and plated with 300 µl of *M. smegmatis* cells containing pLAM12 or pKMC2, either with or without acetamide inducer. To test for viability of the mutant, lysates from plates containing approximately 600 plaques were harvested and tested by PCR with primers GilesB-DiagR3 and GilesB-DiagF. Individual plaques from the

secondary plating were picked and tested using the same two primers. A mutant derivative was plaque purified and the sequence confirmed by DNA sequencing.

To test for complementation of the Giles *ΔlysB* small plaque phenotype, mc^2155 cells containing pKMC9 or pNIT-1 vector control were grown to OD_{600} 0.6 and 10 µl of dilutions of either wild-type Giles or Giles*ΔlysB* phage containing ~10^2 pfu were added to 500 µl of cells, adsorbed for 30 minutes at room temperature, and plated as top agar lawns in 0.35% MBTA with 1 mM $CaCl_2$ on 7H10 plates with 0 to 0.4% ε-caprolactam.

Lysis assays

M. smegmatis cells grown in 7H9 supplemented with ADC, carbenicillin, cyclohexamide, and calcium were grown to OD_{600} 0.3-1.0. For the ATP release assay, cells were diluted to OD_{600} = 0.03, and infected with phage particles at an m.o.i. of 10. After adsorption for 30 min, cultures were shaken at 37°C, and 100 µl samples removed at different times. ATP release was measured by addition of 100 µl of ENLITEN rLuciferase/Luciferin reagent (Promega), and luminescence recorded for a 10 second interval in a Monolight 2010 luminometer. For the OD assay, cells were diluted to an OD_{600} of 0.25 and infected with phage particles at an m.o.i. of 10. After adsorption for 30 min, cultures were shaken at 37°C and 1 ml samples removed at different times. OD was measured at 600 nm using a Beckman Coulter DU 530 Spectrophotometer.

To determine the number of phage released into the supernatant or retained in unlysed cells, *M. smegmatis* cells were grown as above and diluted to an OD_{600} of 0.25. These were infected at an m.o.i. of 0.1 with adsorption for 30 min followed by

incubation with shaking at 37°C. One milliliter samples were removed at different times and separated by centrifugation into supernatant and pellet fractions. The pellet was resuspended in 1 ml phage buffer and sonicated. Both pellet and supernatant were serially diluted and 5 ul samples were spotted onto top agar lawns containing *M. smegmatis* in 0.35% MBTA with 1 mM $CaCl_2$ on 7H10 plates.

Crystallization and structure determination

Purified D29 LysB at 10 mg/ml concentration was used to screen a large number of commercially available crystallization conditions by vapor diffusion method. After optimization upon the hit condition in the initial screen, diffraction quality crystals were obtained with 2 mg/ml protein in 50 mM Tris pH 8, 20% PEG1000, 100 mM $Ca(OAc)_2$. The best crystals grew in 4 days and were flash frozen in mother liquor with 25% glycerol, and then stored in liquid nitrogen. Data sets were collected with a Rigaku MM007HF generator with a Bruker Smart 6000 CCD detector, and with synchrotron beamlines at the Advanced Photon Source (APS 23-ID), Argonne National Laboratory, Chicago. The data were processed with HKL2000 and the anisotropic scaling was further performed on the diffraction anisotropy server (Strong et al., 2006).

The initial structure was solved by SIRAS combining the native crystal with an iodine derivative obtained by quick soaking of the crystal in the cryo with 0.5 M NaI for 2 min. The model was refined against the native dataset iteratively until the R/Rfree reached 0.205/0.252 at 2.0 Å. The iodine sites were found with Shelxd/e program. The phasing and refinement of the protein structure were carried out with the Phenix

package. The model was built in Coot. 20.3 PyMol was used for structure analysis and rendering.

Research history and future direction

When we began this collaboration with the Hatfull lab on mycobateriophage lysine, very little was known about Lysin B. There were no reference papers on it, and Mrs. Kimberly Payne was still Miss Kim Clemens. In fact, D29B was just one of several clones with a collection of Lysin A genes from the Hatfull lab including D29A, Che8A, CorndogA, and Bxz1A. In contrast to the *lysine A* genes, lysin B is unique to the mycobacteriophage. Bioinformatics studies were intriguing, but confusing. Initial hits in the crystallization screens gave us a good starting point. Although there was no homologous structure to use as a template for molecular replacement, and the selenomethione protein did not result in diffractable crystals, the phase problem was overcome by the iodide derivative data.

Of the known structures the most structurally similar proteins belonged to the cutinases (**Table 3-2**). However, the additional helix bundle in the linker domain (Glu178-Arg231) constrains the active site into a small cavity instead of an open surface, though the position of the active triad (Ser82-Asp166-His240) is well conserved. When the D29 LysB structure is aligned to the *Fusarium solani* cutinase structure, which includes a covalently-linked n-undecanylphosphonate methyl ester inhibitor, the linker domain occupies the general position where the acyl chain of the substrate resides (**Fig. 3-8**) (Longhi et al., 1996). The loop (Arg231-Lys237) connecting the linker domain

back to the α/β sandwich has poorly defined electron density, which indicates the high flexibility of this region. We speculate that it serves as a hinge, and the linker domain may move like a lid (Derewenda et al., 1992).

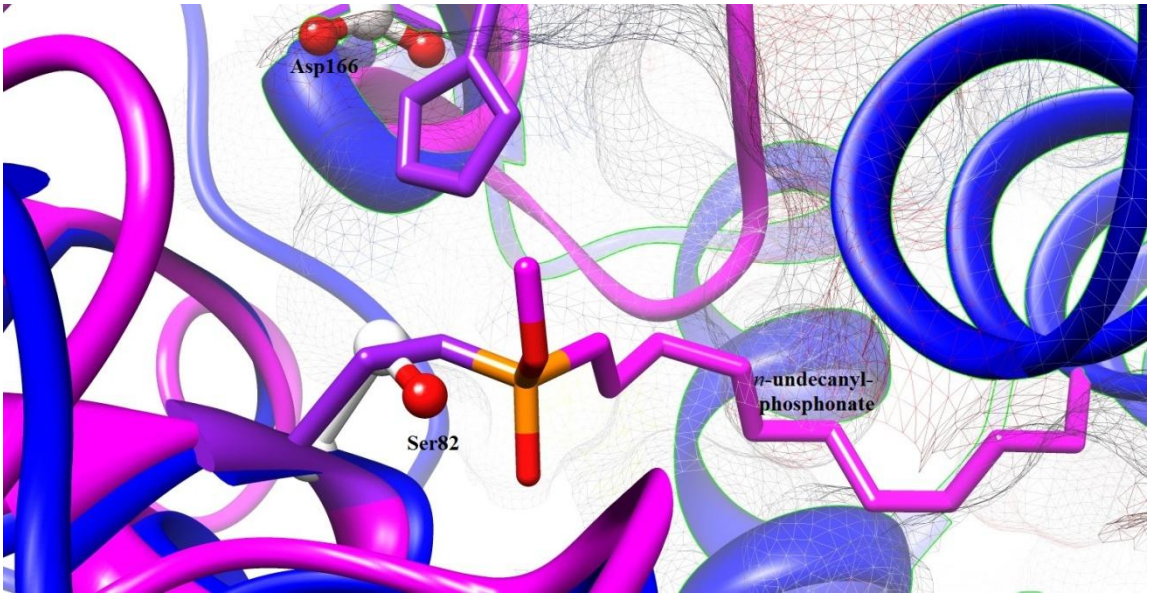


Figure 3-8. Conflict between the linker domain and the putative substrate acyl chain.
The D29 LysB is in blue ribbon, and the *Fusarium solani* cutinase is in magenta (pdb: 1XZM). The surface of the D29 LysB linker domain is shown in mesh. This image was made with Chimera (Pettersen et al., 2004).

To catch D29 LysB in action, we tried soaking the protein crystals with various substrates, such as the artificial p-nitrophenyl substrates of the lipolytic assay and the mAGP extracted from the mycobacterium. These efforts resulted in either cracked crystals or no substrate electron density. This may corroborate our speculation of the large movement of the linker domain involved in substrate recognition and binding,

although further experiments are required to prove this. Our attempts at co-crystallization have not yet yielded suitable crystals.

It may be possible to analyze D29 LysB by NMR as the size of the protein falls in the practical range for NMR. Future structural research on D29 LysB may be accomplished using this technique.

CHAPTER IV

PREDICTION OF PLASMODIUM L4 STRUCTURE REVEALS ANTIMALARIAL DRUG MECHANISM*

Summary

Azithromycin (AZ), a broad-spectrum antibacterial macrolide that inhibits protein synthesis, also manifests reasonable efficacy as an antimalarial. Its mode of action against malarial parasites, however, has remained undefined. Our *in vitro* investigations with the human malarial parasite *Plasmodium falciparum* document a remarkable increase in AZ potency when exposure is prolonged from one to two generations of intraerythrocytic growth, with AZ producing 50% inhibition of parasite growth at concentrations in the mid to low nanomolar range. In our culture-adapted lines, AZ displayed no synergy with chloroquine (CQ), amodiaquine, or artesunate. AZ activity was also unaffected by mutations in the *pfcrt* (*P. falciparum* chloroquine resistance transporter) or *pfmdr1* (*P. falciparum* multidrug resistance-1) drug resistance loci, as determined using transgenic lines. We have selected mutant, AZ-resistant 7G8 and Dd2 parasite lines. In the AZ-resistant 7G8 line, the bacterial-like apicoplast large subunit ribosomal RNA harbored a U438C mutation in domain I. Both AZ-resistant lines

*Reprinted with permission from "In vitro efficacy, resistance selection, and structural modeling studies implicate the malarial parasite apicoplast as the target of azithromycin" by Sidhu, A. B., Sun, Q., Nkrumah, L. J., Dunne, M. W., Sacchettini, J. C., Fidock, D. A., 2007. *J. Biol. Chem*, 282, 2494-2504, Copyright [2011] by American Society for Biochemistry and Molecular Biology. David Fidock's lab performed the genetics and pharmacology experiments. Sun, Q. did the structural biology studies.

revealed a G76V mutation in a conserved region of the apicoplast-encoded *P. falciparum* ribosomal protein L4 (PfRpl4). This protein is predicted to associate with the nuclear genome-encoded *P. falciparum* ribosomal protein L22 (PfRpl22) and the large subunit rRNA to form the 50 S ribosome polypeptide exit tunnel that can be occupied by AZ. The PfRpl22 sequence remained unchanged. Molecular modeling of mutant PfRpl4 with AZ suggests an altered orientation of the L75 side chain that could preclude AZ binding. These data imply that AZ acts on the apicoplast bacterial-like translation machinery and identify *Pfrpl*4 as a potential marker of resistance.

Introduction

Licensed antibiotics with antimalarial activity offer considerable promise as suitable partners for novel combination therapies directed against drug-resistant *Plasmodium falciparum* infections. Clinical studies have demonstrated good clinical cure rates with treatments that use tetracycline, its more potent derivative doxycyline, or clindamycin (CLD), in combination with fast-acting antimalarials such as quinine (Baird, 2005; Lell and Kremsner, 2002; White, 1996). Whereas the mode of action of these antibiotics has until recently remained elusive for *Plasmodium*, against bacteria they act by binding to the 70S ribosome (comprised of the 50S and 30S subunits), thereby inhibiting protein synthesis. Tetracycline and doxycycline bind to the 30S subunit (comprising the 16S rRNA and ribosomal proteins), whereas clindamycin binds to the 50S subunit (comprising the 23S rRNA, the 5S rRNA, and ribosomal proteins including L4 and L22) (Poehlsgaard and Douthwaite, 2005). The World Health

Organization recommends doxycycline as malaria prophylaxis for travelers to endemic areas. Tetracycline and doxycycline, however, manifest toxicity in children below the age of eight and in pregnant women, effectively precluding their use in the most important malaria patient populations (Kain et al., 2001).

Another antibiotic that appears particularly promising for antimalarial combination therapy is azithromycin (AZ), a 15-membered azalide that has been broadly used for the treatment of streptococcal, staphylococcal, chlamydial, and gonorrheal infections. Notably, AZ displays a good safety profile for young children and pregnant women (Girard et al., 1987). This semisynthetic, second-generation antibiotic was derived from its parent compound erythromycin by introducing a methyl-substituted nitrogen group at position C9 of the lactone ring, resulting in several improved features. These include a broader spectrum of activity, more favorable pharmacodynamics, and a longer elimination half-life. This extended half-life has been attributed to AZ concentrating inside lysosomes in phagocytic cells, as a result of protonation of its two basic amine groups (Retsema et al., 1987), followed by AZ leaching out into the bloodstream over a period of several days to a week. This provides effective therapeutic coverage with AZ for several days post completion of treatment.

Against bacteria, AZ acts by binding reversibly to the 50S ribosomal subunit, in the vicinity of the peptidyltransferase center in domain V and the entrance of the polypeptide tunnel (Poehlsgaard and Douthwaite, 2005). This binding site is formed by the association of domains II and V of the bacterial 23S rRNA and includes the ribosomal proteins L4 and L22 that line the protein exit tunnel (Ban et al., 1999;

Douthwaite et al., 2000; Gabashvili et al., 2001). AZ binding interferes with transpeptidation/translocation of nascent polypeptides exiting the polypeptide tunnel, leading to premature detachment of incomplete peptide chains and subsequent cell death (Hansen et al., 2002; Retsema and Fu, 2001).

AZ activity against malarial parasites was initially discovered by *in vitro* testing of chloroquine (CQ)-sensitive and CQ-resistant *P. falciparum* lines (Gingras and Jensen, 1992, 1993). Further studies revealed a slow onset of parasite killing, similar to tetracycline, doxycycline, and clindamycin (Yeo and Rieckmann, 1995). *In vitro* screening for potential drug combinations was suggestive of additive to synergistic interactions with AZ-CQ as well as AZ-quinine combinations in some CQ-resistant *P. falciparum* strains (Ohrt et al., 2002). Against CQ-sensitive strains, the effect was additive. *In vivo* studies, initially conducted in mice infected with blood stage *P. berghei* parasites, also showed good treatment efficacy with AZ monotherapy. This efficacy was enhanced by combining AZ with the fast-acting blood schizonticidal agents CQ, quinine, artemisinin or halofantrine (Andersen et al., 1995; Gingras and Jensen, 1993). AZ curative blood schizonticidal efficacy was also confirmed in *Aotus* monkeys infected with CQ-resistant *P. falciparum* (Andersen et al., 1995). Evidence that AZ was clinically effective in treating *P. falciparum* malaria in humans came from pilot studies showing that AZ monotherapy was an effective prophylactic agent (Anderson et al., 1995; Kuschner et al., 1994). A more recent treatment trial in India reported close to 100% cure rates in patients treated with AZ plus CQ, with this combination proving to be far more effective than monotherapy with the individual agents (Dunne et al., 2005a). High

levels of treatment efficacy were also observed in a recent trial from Thailand that studied regimes of combing AZ with quinine or artesunate (Noedl et al., 2006). A monotherapy trial in India also found AZ to be well tolerated and highly effective in treating *Plasmodium vivax* malaria (Dunne et al., 2005b).

The proven antimalarial efficacy observed with AZ has made it important to define its mode of action against malarial parasites. One pressing question is whether AZ, a weak base, acts inside the parasite digestive vacuole (DV), a lysosomal-like acidic compartment (pH 5.0 – 5.4) where CQ and other weak base 4-aminoquinolines concentrate and act by interfering with heme detoxification (Banerjee and Goldberg, 2001; Foley and Tilley, 1998; O'Neill et al., 1998). Such an activity might account for earlier reports of AZ-CQ *in vitro* synergy (Ohrt et al., 2002). Another possibility is that AZ acts by inhibiting the bacterial-like protein synthesis machinery present in the *Plasmodium* apicoplast, an organelle of cyanobacterial origin that has been shown in the related Apicomplexan parasite *Toxoplasma gondii* to be indispensable for parasite growth (Fichera and Roos, 1997; Pfefferkorn and Borotz, 1994; Ralph et al., 2004). To address these questions, we have investigated the efficacy of AZ alone or in combination with fast-acting antimalarials, the effect on AZ susceptibility of point mutations in DV transmembrane transporters, the selection of AZ-resistant parasite lines, and the molecular mechanisms that can casuse AZ resistance. Our results directly implicate the apicoplast 50S ribosomal subunit as the target of AZ in *P. falciparum*.

Experimental procedures

Parasite culture and in vitro *antimalarial drug assays*

P. falciparum parasites were cultured *in vitro* in human red blood cells as described (Fidock et al., 1998). Antimalarial susceptibilities were measured in 96-well plates over 48 or 96 h of exposure with an initial parasitemia of 0.5 or 0.2% synchronized ring stage parasites, respectively. Drug IC_{50} values (producing 50% inhibition of [^{3}H]hypoxanthine uptake) were derived by linear regression from the dose-response curves (Desjardins et al., 1979; Fidock et al., 1998). In the 48 h drug assays, parasites were exposed to drug for 30 h, and 0.5 μCi of radiolabeled [^{3}H]hypoxanthine was added to each well for the last 18 h. In the 96 h drug assays, media was replaced after 48 and 72 h, and [^{3}H]hypoxanthine was added at the 72 h time point (0.5 μCi/well). Media replacements contained appropriate amounts of drug to ensure that drug concentrations remained constant in each well throughout the duration of the assay. To test the interaction of AZ with other drugs, each drug was tested alone and in combination at fixed ratios of its IC_{50} value, as calculated separately for each assay (AZ, test drug ratios of 0:1, 1:3, 1:5, 1:1, 3:1, 5:1 and 1:0). Fractional IC_{50} (FIC_{50}) values from 96 h assays were calculated as previously described (Ohrt et al., 1997) and used to plot isobolograms. Two-tailed unpaired Student's *t* tests were performed for statistical comparisons. AZ dihydrate (CP-62,993-03) was obtained from Pfizer. Thiostrepton, tetracycline, doxycycline, and erythromycin were purchased from Calbiochem, whereas triclosan and CQ were purchased from Sigma.

Selection of AZ-resistant P. falciparum *lines*

P. falciparum lines were selected for resistance to AZ using the following regime. First, we exposed Dd2 and 7G8 parasites (5×10^{10} each) to 38 and 115 nM AZ, respectively (*i.e.*, 30 and 90 ng/ml of AZ dihydrate), for 3 days with twice daily feeding. We then increased the drug concentration to 380 nM for both lines and maintained this level for 4 days. For the following 15 days, Dd2 and 7G8 parasite lines were again maintained at 38 and 115 nM AZ, respectively. Giemsa-stained thin blood films were examined every 2 to 3 days to check for viable asexual-stage parasites. Once a week, 30–40% of the culture was replaced with freshly drawn red blood cells. For both lines, AZ-resistant ring stage parasites were observed by day 21. Once the parasitemias reached 2–3%, parasites were phenotypically characterized for their drug response profiles using [^{3}H]hypoxanthine incorporation assays. Dd2 and 7G8 AZ-resistant parasite lines (termed AZ-R^{Dd2} and AZ-R^{7G8}, respectively) were subsequently cloned by limiting dilution (Goodyer and Taraschi, 1997) and cultured in the presence of 640 and 1280 nM AZ, respectively.

DNA constructs and sequencing

A 1.0 kb DNA fragment encompassing the open reading frame of the *P. falciparum* apicoplast-encoded ortholog of the *rpl*4 ribosomal protein gene (denoted *Pfrpl*4; GenBank™ accession number X95276) was PCR amplified from the wild-type and AZ-resistant parasite lines using primers p1083 (5′-AATGGTAACATATCTATTTTGGGAATAGA) and p1084 (5′-

AATTTCCACCTCCTTTATTTAATAGTA). This fragment was used in nested PCR with internal primers p1112 (5′-TTatcgatAATATTATTATTTTAAATAATAATACA; ClaI site in lowercase) and p1111 (5′-AcgtacgAATTAAAAATATTAAATAATTATACATAAT (BsiWI site in lowercase) to amplify a full-length 0.6 kb coding sequence fragment lacking only the ATG start and TAA stop codons. Amplified products were cloned into pCR2.1-TOPO vector (Invitrogen) and sequenced using T7 and M13 reverse primers (Invitrogen).

For the transfection studies, sequence-verified *Pfrpl*4 inserts were subcloned into the ClaI and BsiWI restriction sites of the expression plasmid pLN-ACP-V5. This plasmid, based on pLN (Nkrumah et al., 2006), contains a 180 bp sequence encoding the N-terminal bipartite leader sequence of the fatty acid synthase acyl-carrier protein (ACP), used to target proteins to the apicoplast (Waller et al., 2000), as well as a 3′ sequence encoding a C-terminal V5 epitope tag (GKPIPNPLLGLDST). The final construct, termed pLN-ACP-*Pfrpl*4-V5 (**Fig. C-1**), contained mutant or wild-type *Pfrpl*4 fused at its 5′ end to the ACP leader sequence and at its 3′ end to the V5 tag. This fusion gene is regulated by a 1.0 kb *calmodulin* 5′-untranslated region and a 0.8 kb *hsp86* 3′-untranslated region. Related constructs were made that replaced the 3′ V5 tag with a GFP-Mut2 (green fluorescent protein) sequence, creating the pLN-ACP-*Pfrpl*4-GFP plasmid. From these plasmids, we derived pLN-SP-*Pfrpl*4-V5 and pLN-SP-*Pfrpl*4-GFP by replacing the full-length ACP apicoplast targeting sequence with a 57 bp signal peptide (SP) sequence from the ACP gene. All of these constructs were also made with a 0.9 kb *hsp86* 5′-untranslated region fragment in place of the *calmodulin* 5′-untranslated

region. This resulted in 16 constructs, which were all transfected into Dd2 parasites. The *calmodulin* promoter series of constructs were also transfected into 3D7 and GC03 parasites. Transfections were carried out as previously described (Fidock and Wellems, 1997) and recombinant lines were selected using 2.5 μg/ml blasticidin hydrochloride.

The 0.9 kb full-length *P. falciparum* nuclear genome-encoded ortholog of the *rpl*22 ribosomal protein gene (*Pfrpl*22; PlasmoDB ID number PF14_0642) was PCR amplified using p1429 (5′-TAATGATATTGCGCAAGTGTTGATTTGT) and p1432 (5′-GTACATAAATATATACACATTTAATTTTTAACAAACGGA). The apicoplast large subunit (LSU) rRNA gene (GenBank accession number X95275) was PCR amplified as a 2.9 kb fragment with primers p986 (5′-GCGAAATAGAGCATAAGGAAAGTTCG) and p800 (5′-AGCTAATGGTGAGATTTGAACT). PCR fragments were gel purified and fully sequenced on both strands using internal primers.

Structural modeling

Sequences and structures of ribosomal L4 proteins were retrieved from the RCSB Protein Data Bank. Multiple sequence alignments were carried out at the SDSC workbench using *CLUSTAL W* (Thompson et al., 1994). The model of the PfRpl4 conserved loop from Lys^{57} to Pro^{97} was built with MODELLER (Sali and Blundell, 1993). As templates, we used the *Escherichia coli* and *Deinococcus radiodurans* L4 proteins, which share 39 and 32% sequence identity with PfRpl4 in this loop region. The AZ molecule in the *D. radiodurans* ribosome structure (Schlunzen et al., 2003) also

served as a ligand for modeling. The loop of the PfRpl4 G76V mutant was also built using MODELLER with all three wild-type protein structures as templates. The quality of these models was evaluated with DOPE potential (Melo et al., 2002), and their superimposition and rendering were carried out with PyMol.

Results

AZ displays extraordinary increases in potency against P. falciparum *strains exposed for a second generation of growth as compared with thiostrepton, triclosan, and CQ*

A previous study with two *P. falciparum* lines observed an 80- to 220-fold increased potency when the duration of AZ exposure was prolonged from 48 h to 96 h (Yeo and Rieckmann, 1995). We extended this observation to other multidrug resistant and drug sensitive lines and compared the effect to that of other antimalarial agents that share similar modes of action or are in clinical use. The lines chosen were Dd2 (Indochina; resistant to CQ, QN, pyrimethamine and sulfadoxine), 7G8 (Brazil; resistant to CQ), and GC03 (a CQ-sensitive progeny of the HB3 × Dd2 genetic cross (Wellems et al., 1990)). Comparative antimalarial agents included thiostrepton (which targets the GTPase site of the apicoplast LSU rRNA (McConkey et al., 1997; Rogers et al., 1997)); triclosan (a topical antimicrobial that inhibits enoyl ACP reductase, a component of the fatty acid type II biosynthesis pathway in the *Plasmodium* apicoplast (Perozzo et al., 2002; Waller et al., 1998)), and CQ (which binds to heme moieties in the parasite DV and prevents their detoxification (Uhlemann et al., 2005)). A comparative profile of responses to each compound for these parasite lines is presented as IC_{50} values in **Figure**

4-1. After 48 h, Dd2, 7G8, and GCO3 displayed AZ IC_{50} values of 3.5, 15.7 and 18.1 μM, respectively, whereas after 96 h these values decreased to 103, 190 and 77 nM, respectively (**Table C-1**). This translated into a substantial 34-, 83-, and 234-fold increase in AZ potency resulting from a second generation of parasite exposure. In comparison, these lines had thiostrepton IC_{50} values of 377, 521, and 917 nM, respectively, after 96 h, corresponding to more modest 4.5-, 1.3-, and 2.5-fold increases in susceptibility. In 96 h assays with these same lines, the triclosan IC_{50} values were 4186, 424 and 1526 nM, and CQ IC_{50} values were 134, 113, and 24 nM, respectively. This corresponded to 1.0-2.0 fold increases in triclosan or CQ susceptibility. These results indicate that AZ becomes a highly active antimalarial that manifests its potency after a second generation of parasite intraerythrocytic growth. This increase in potency far exceeds other agents that also presumably target apicoplast pathways, including in the case of thiostrepton presumably the same ribosomal subunit involved in protein synthesis.

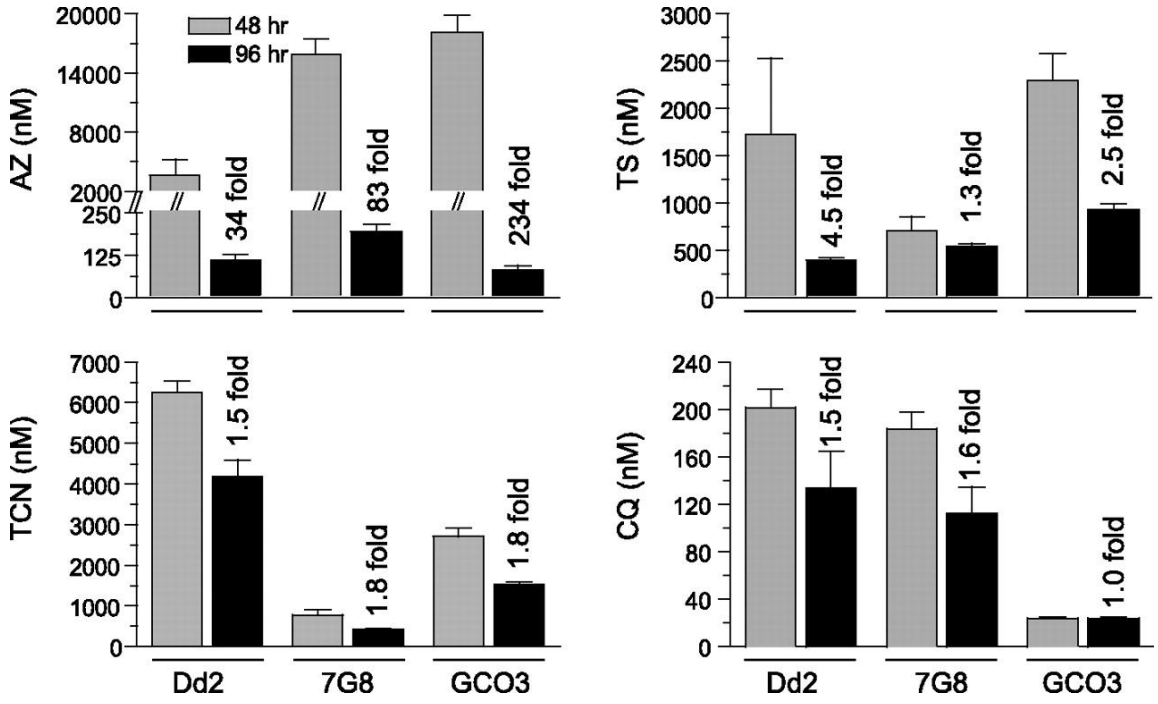


Figure 4-1. Antimalarial activity of AZ, thiostrepton (TS), triclosan (TCN), and CQ against *P. falciparum* Dd2, 7G8 and GCO3 lines.

The data represent the mean±SEM IC_{50} values calculated from at least five independent [^{3}H]-hypoxanthine drug assays involving drug exposure for either 48 h or 96 h (corresponding to one or two generations of the parasite's asexual blood stage cycle). The fold increases in parasite sensitivity to these compounds from 48 h to 96 h are indicated. Note that a 2-segment Y-axis is represented for the AZ data to adequately present the wide range of values.

AZ is not synergistic with CQ, amodiaquine, or artesunate

To investigate the antimalarial activity of AZ in combination with CQ, 96 h combination drug assays were performed using the CQ-resistant *P. falciparum* lines Dd2, 7G8, and 3BA6 (the latter is a CQ-resistant progeny of the HB3 × Dd2 cross). From these, FIC_{50} values were determined and isobolograms were derived and plotted. As illustrated in **Figure 4-2**, this combination appeared essentially additive for Dd2 (mean FIC_{50} value: 1.39, 95% confidence interval (CI), 1.10–1.69) and 3BA6 (mean

FIC_{50} value: 1.38, 95% CI, 1.22–1.55) and was also additive for 7G8 (mean FIC_{50} value: 1.05, 95% CI, 1.02–1.08).

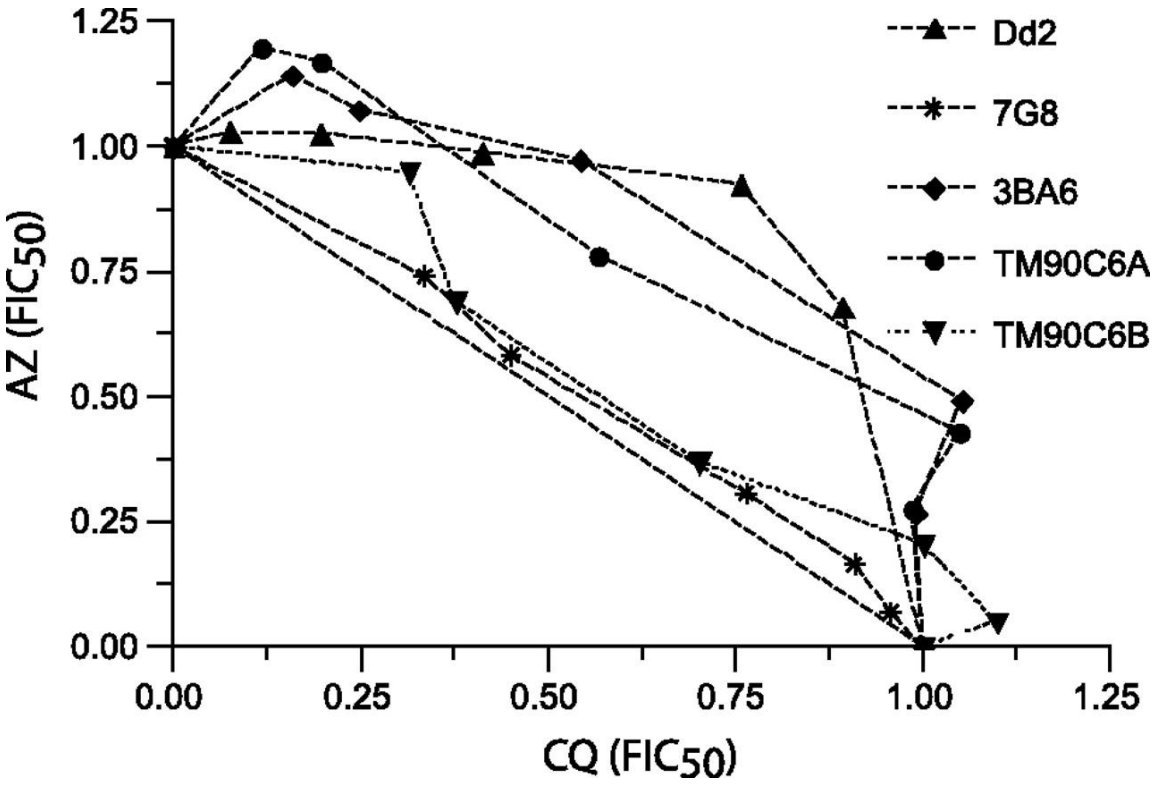


Figure 4-2. Isobologram plots of AZ plus CQ tested against *P. falciparum* culture-adapted lines. FIC_{50} values were determined for AZ and CQ tested at different concentration ratios (0:1, 1:3, 1:5, 1:1, 3:1, 5:1, and 1:0). Values for the Dd2, 7G8, and 3BA6 lines were calculated from 96 h assays, whereas for the TM90C6A and TM90C6B lines these were calculated from 72 h assays. For each line, isobolograms were performed at least 5–7 times for each drug combination.

We also assayed the TM90C6A and TM90C6B Thai isolates that have been more recently adapted to *in vitro* culture (Looareesuwan et al., 1996). AZ-CQ combination assays performed with these lines also found no evidence for synergy (TM90C6A mean

FIC_{50} value, 1.35, 95% CI, 1.25–1.45; TM90C6B mean FIC_{50} value, 1.15, 95% CI, 1.04–1.25; **Fig. 4-2**).

In view of the decline of CQ as a first-line antimalarial and the increasing interest in the use of amodiaquine or artesunate as alternative antimalarials (notably for combination therapy), we also assayed AZ-amodiaquine and AZ-artesunate combinations with the drug-resistant Dd2 and 7G8 lines. In 96 h combination assays, mean FIC_{50} values for AZ-amodiaquine were 1.28 (95% CI, 1.01–1.56) and 1.23 (95% CI, 0.87–1.58), whereas mean FIC_{50} values for AZ-artesunate were 1.62 (95% CI, 1.02–2.22) and 1.34 (95% CI, 1.14–1.54) against Dd2 and 7G8 parasite lines, respectively (data not shown). Thus, these assays found no evidence of *in vitro* synergy between AZ and either of these drugs.

Parasite susceptibility to AZ is not influenced by mutations in the DV transmembrane proteins PfCRT and PfMDR1

We assessed whether *P. falciparum* susceptibility to AZ *in vitro* could be influenced by point mutations in the DV transmembrane proteins PfCRT or PfMDR1 (the latter is also known as Pgh1). This was for several reasons. First, genetic studies have demonstrated that mutant forms of these proteins can affect parasite *in vitro* susceptibility to a wide range of antimalarials, including CQ, quinine, mefloquine, halofantrine, and the anti-influenzal agent amantadine (Cooper et al., 2002; Fidock et al., 2000; Johnson et al., 2004; Reed et al., 2000; Sidhu et al., 2005; Sidhu et al., 2002; Sidhu et al., 2006). Second, there is a published report of AZ-CQ synergism that raised

the possibility of their sharing a related mode of action (Ohrt et al., 2002). Third, AZ, like CQ, is a weak base that could be predicted to accumulate in the acidic *P. falciparum* DV (where CQ also accumulates and exerts its antimalarial property (Saliba et al., 1998)). We thus took advantage of existing recombinant lines that had been genetically modified to express mutant forms of either protein (Lakshmanan et al., 2005; Sidhu et al., 2005; Sidhu et al., 2002) and tested these in 96 h assays with AZ. Results showed AZ IC_{50} values in the range of 48-194 nM for *pfcrt* or *pfmdr1* genetically modified lines, with no significant changes observed between these lines and their parental or recombinant control lines (**Figure 4-3;** haplotypes and IC_{50} values are reported in **Table C-2**). These results demonstrate that mutations in these DV transporter proteins do not noticeably modulate parasite susceptibility to AZ.

AZ-resistant P. falciparum *lines can be selected* in vitro

To assess whether AZ resistance could be selected *in vitro*, we subjected Dd2 and 7G8 parasites to moderate levels of AZ (see "**Experimental Procedures**") and selected resistant lines. These lines, termed AZ-R^{Dd2} and AZ-R^{7G8}, were assayed for their degree of AZ resistance in 96 h assays, in parallel with their parental lines Dd2 and 7G8. This produced AZ IC_{50} values of 1.9 and 4.0 μM, respectively, as compared with IC_{50} values of 124 and 228 nM, respectively, for the parental Dd2 and 7G8 lines, respectively, corresponding to highly significant ($p < 0.001$) 16- and 17-fold decreases in AZ susceptibility in the selected lines (**Figure 4-4**; values reported in **Table C-3**). The

fact that resistant lines could be rapidly selected highlights the need to survey for resistance in AZ-containing anti-malarial combination therapies.

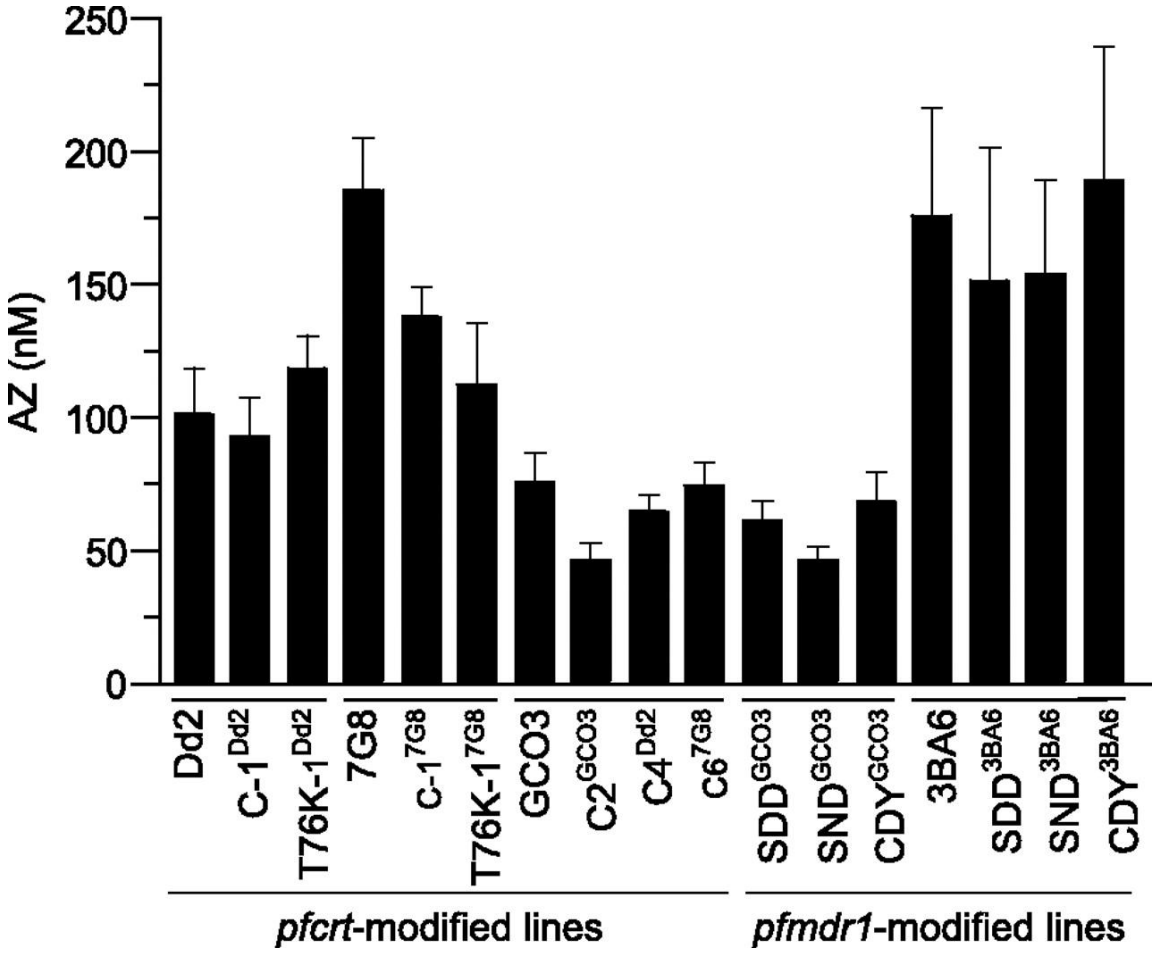


Figure 4-3. Response of *pfcrt*- and *pfmdr1*- genetically modified parasite lines to AZ
The data represent the mean±SEM IC_{50} values from 96 h drug assays performed on six separate occasions in duplicate. The PfCRT and PfMDR1 haplotypes of these genetically modified parasite lines are shown in Supplemental **Table C-2**. The AZ IC_{50} values for Dd2, 7G8 and GCO3 lines were taken from assays performed and represented in **Figure 4-1**.

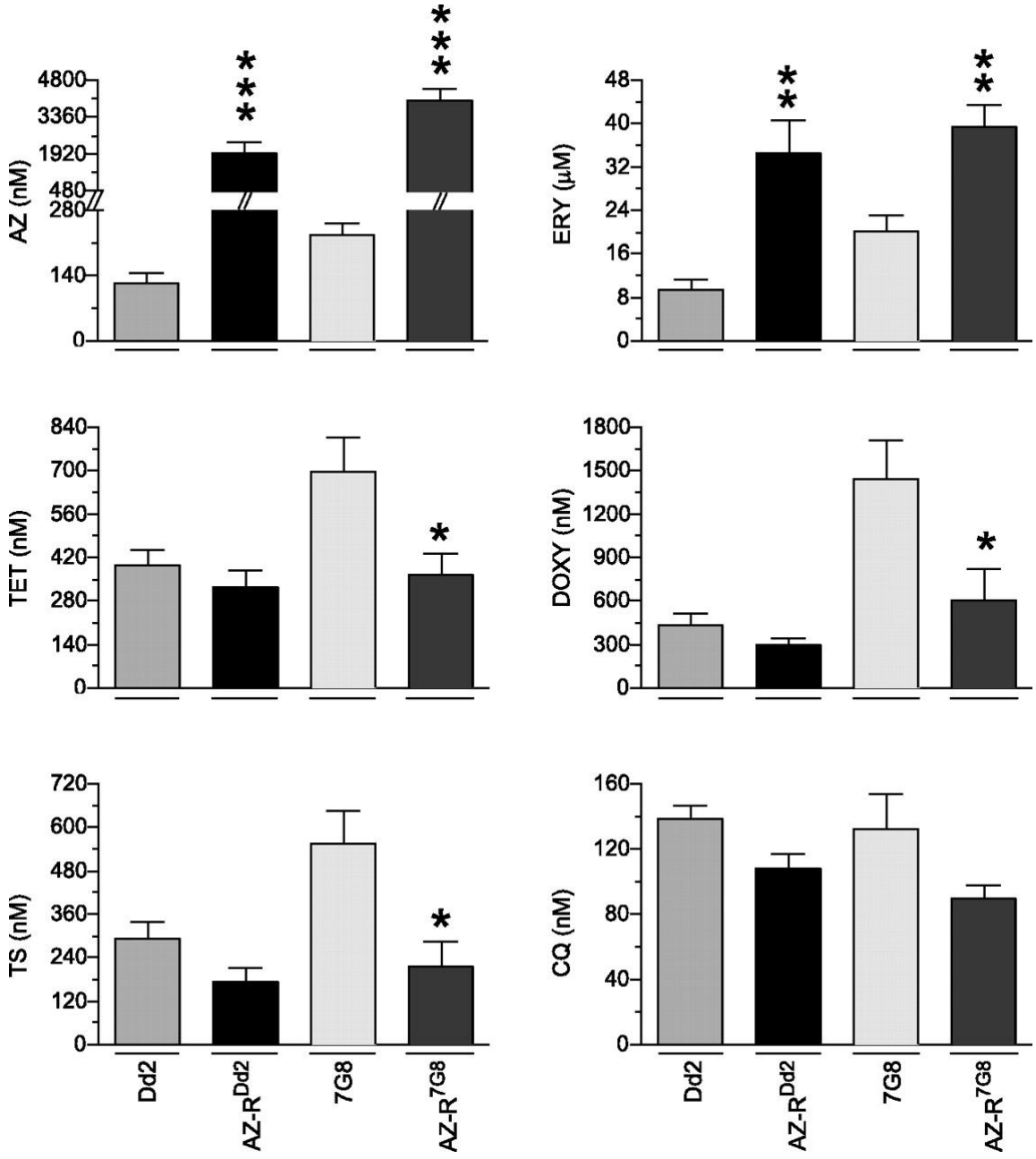


Figure 4-4. Phenotypic response profile of the AZ-selected parasite lines AZ-R^{Dd2} and AZ-R^{7G8}
Mean±SD IC_{50} values of AZ, erythromycin (ERY), tetracycline (TET), doxycycline (DOXY), thiostrepton (TS) and CQ were determined from 96 h drug assays on AZ-selected mutant lines and parental controls. Values were calculated from five independent assays performed in duplicate. IC_{50} values were compared between AZ-R^{Dd2} and AZ-R^{7G8} and their respective parental lines and Mann-Whitney *U*-tests were used to assess for statistical significance (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). Note that the AZ bar chart is represented using a 2-segment Y-axis to cover the data range.

As noted earlier, AZ and its progenitor erythromycin are chemically very closely related, and cross-resistance between the two antibiotics has been reported in several bacterial species including *Streptococcus pneumoniae*, *Haemophilus influenza*, and *Neisseria gonorrhoeae* (Canu et al., 2002; Clark et al., 2002; Ng et al., 2002; Tait-Kamradt et al., 2000a). Accordingly, we performed 96 h assays to investigate whether our *P. falciparum* AZ-resistant mutant lines would also be resistant to erythromycin. Indeed, this was observed but to a smaller extent, with significant ($p < 0.01$) 4- and 2-fold increases in erythromycin IC_{50} values in AZ-R^{Dd2} and AZ-R^{7G8}, respectively (**Figure 4-4**).

We also screened for possible changes in the parasite's susceptibility to tetracycline, doxycycline, thiostrepton, and CQ. Tetracycline and doxycycline were chosen because of their clinical use in combination with quinine, primarily for the treatment of uncomplicated, CQ-resistant *P. falciparum* malaria. As compared with its parental line 7G8, the mutant AZ-R^{7G8} line displayed significant ($p < 0.05$) decreases in tetracycline, doxycycline, and thiostrepton IC_{50} values, indicating that AZ-R^{7G8} had become hypersensitive to these antimalarial drugs. No such change in susceptibility to these drugs was observed in the AZ-R^{Dd2} line compared with Dd2. The CQ response was unchanged in both mutant AZ-resistant lines (**Figure 4-4**).

To assess the stability of the AZ-resistant phenotype, both AZ-selected lines were maintained in the absence of drug pressure for 1 month and the AZ responses then retested. This revealed no alteration in IC_{50} values as compared with parasites

continuously cultured in the presence of AZ (IC_{50} values for AZ-R^{Dd2} and AZ-R^{7G8} were 2.1 and 4.2 μM in the absence of drug pressure *versus* 1.9 and 4.0 μM in the presence of sustained drug pressure, respectively, in 96 h assays). These findings indicated that the selected AZ-resistance phenotype was stable over time.

Identification of mutations in the apicoplast LSU rRNA and Pfrpl4 genes in AZ-resistant P. falciparum

To examine the molecular basis of AZ resistance, we PCR amplified and sequenced the apicoplast LSU rRNA (the *P. falciparum* ortholog of bacterial 23S rRNA) and *Pfrpl4* genes as well as the nuclear *Pfrpl22* gene (all are present as a single copy). We selected these potential targets based on their similarity with bacterial genes and published reports that mutations in these sequences can confer resistance to AZ and related macrolides in pathogenic bacteria (Gardner et al., 1993; Leclercq, 2002; Vester and Douthwaite, 2001; Wilson et al., 1996).

In examining the *P. falciparum* LSU rRNA (which bears 70% identity to the *E. coli* 23S rRNA (Gardner et al., 1993)), we first focused on A1875, which corresponds to the *E. coli* A2058 nucleotide in the peptidyltransferase center of domain V, and A706, which is equivalent to *E. coli* A754 in domain II. Mutations in these positions can confer macrolide resistance in bacterial species including *E. coli*, *Helicobacter pylori*, *S. pneumoniae*, and *Mycobacterium* (Vester and Douthwaite, 2001). We also examined G1878 (which corresponds to *E. coli* G2061) as a mutation at this position can confer resistance to clindamycin and moderately decrease the susceptibility to AZ in *T. gondii*

(Camps et al., 2002). These nucleotides remained unchanged in our AZ-resistant *P. falciparum* lines, as did the entire domains II (hairpin loop 35) and V that in bacteria form a binding pocket for macrolides including AZ and that can also harbor other mutations conferring macrolide resistance (Ban et al., 1999; Hansen et al., 2002; Schlunzen et al., 2003; Tu et al., 2005; Yonath, 2005).

Upon sequencing the full-length 2.7 kb apicoplast LSU rRNA gene, we identified a T to C mutation at nucleotide position 438 in the AZ-R^{7G8} line. No mutation was observed in this gene in AZ-R^{Dd2}. Sequence analysis confirmed that both the Dd2 and 7G8 parental lines carried the same T^{438} nucleotide, indicating that the AZ-R^{7G8} line had acquired this novel LSU rRNA U438C mutation during the period of AZ selection. Intriguingly, this mutation was found in region I, away from domains II and V that contains the majority of the mutations associated with AZ resistance (Ban et al., 2000).

During this study, we also identified a ribosomal *rpl*22 ortholog in *P. falciparum* (*Pfrpl*22; corresponding to PF14_0642 and originally annotated as encoding a hypothetical protein) based on the similarity of its encoded product with bacterial L22 proteins and an apicolast-directed Rpl22 protein in *T. gondii* (see sequence alignment in **Fig. C-2**). Sequence analysis of the full-length *Pfrpl*22 gene from the AZ-R^{Dd2}, AZ-R^{7G8}, Dd2, and 7G8 parasite lines showed no sequence change.

Analysis of the *Pfrpl4* gene, however, identified a single point mutation, resulting in a glycine (Gly, codon GGA) to valine (Val, codon GTA) substitution at codon 76 in both the AZ-R^{Dd2} and AZ-R^{7G8} lines that was not present in either of the parental Dd2 or 7G8 lines. This G76V mutation is present in the conserved *P.*

falciparum motif ^{71}VQKGLGKAR^{79}, which bears close similarity to the bacterial motif ^{61}RQKGTGRAR^{69} (conserved residues underlined) present in the *E. coli* L4 protein (**Table 4-1**) (Chittum and Champney, 1994; Farrell et al., 2004; Nagai et al., 2002; Pihlajamaki et al., 2002; Tait-Kamradt et al., 2000a). The fact that both the Dd2 and 7G8 parasite lines were independently selected for AZ resistance and underwent the same mutation in their *Pfrpl4* gene provides compelling evidence that this G76V mutation can contribute to *P. falciparum* resistance to AZ. For both the *Pfrpl4* and LSU rRNA genes, the sequences were determined from PCR-amplified products and the electropherograms were unambiguous, implying that the mutant lines had a homogeneous population of mutant apicoplast genomes with no evidence of the wild-type equences. At this time, we cannot exclude the possibility that undetected mutations occurred elsewhere in the apicoplast or nuclear genomes in our AZ-pressured lines.

Table 4-1. Alignment of the ribosomal protein L4 conserved motif from *P. falciparum* and other selected species*

Organism	Phenotype	Amino acid sequence									Residues
P. falciparum	Sensitive	V	Q	K	G	L	G	K	A	R	71-79
T. gondii	Sensitive	Q	Q	K	G	S	G	K	A	R	78-86
D. radiodurans	Sensitive	G	Q	K	G	T	G	N	A	R	60-68
S. pneumoniae	Sensitive	R	Q	K	G	T	G	R	A	R	66-74
E. coli	Sensitive	R	Q	K	G	T	G	R	A	R	61-69
P. falciparum AZI-R^{Dd2}	Resistant	V	Q	K	G	L	**V**	K	A	R	71-79
P. falciparum AZI-R^{7G8}	Resistant	V	Q	K	G	L	**V**	K	A	R	71-79
S. pneumoniae	Resistant	R	Q	K	**C**	T	G	R	A	R	66-74
S. pneumoniae	Resistant	R	Q	K	**T**	**P**	**S**	R	A	R	66-74
S. pneumoniae	Resistant	R	Q	K	**V**	**P**	G	R	A	R	66-74
S. pneumoniae	Resistant	R	Q	**Q**	G	T	G	R	A	R	66-74
S. pneumoniae	Resistant	R	Q	K	G	T	**R**	R	A	R	66-74
E. coli	Resistant	R	Q	**N**	G	T	G	R	A	R	61-69

* P: Plasmodium; T: Toxoplasma; D: Deinococcus; S: Streptococcus; E: Escherichia. L4 mutations in AZ-resistant; organisms are indicated as bold and boxed residues.

Apicoplast-encoded PfRpl4 cannot be redirected to this organelle from episomally replicating transgene plasmids

As an initial approach to assess whether the *Pfrpl*4 G76V mutation was sufficient for AZ resistance, we transfected mutant and wild-type *Pfrpl*4 alleles into several AZ-sensitive *P. falciparum* lines (3D7, Dd2, and GCO3). The native *Pfrpl4* gene is located on the apicoplast genome, yet our transfection strategies were based on expression of this gene from plasmid DNA that was replicated in the nucleus. In an

attempt to target recombinant PfRpl4 to its host apicoplast organelle, these alleles were ligated to a DNA fragment encoding the complete apicoplast targeting sequence of ACP as this sequence was previously shown to direct trafficking of nuclear-encoded gene products to the apicoplast (Waller et al., 2000). This bipartite targeting sequence is composed of a SP and a transit peptide, and the apicoplast targeting sequence algorithm PlasmoAP (Foth et al., 2003) predicted the presence of a putative transit peptide sequence in PfRpl4. Accordingly, we also generated constructs that fused just the SP sequence from the ACP gene to mutant or wild-type *Pfrpl*4. The recombinant ACP-*Pfrpl*4 and SP-*Pfrpl4* mutant or wild-type gene fusions (with a V5 epitope tag or GFP at the 3′ end) were cloned into expression plasmids under the control of *calmodulin* or *hsp86* promoters, making a total of 16 constructs (Crabb et al., 1997). Parasite lines transfected with sequence-verified plasmids expressing ACP-*Pfrpl4*-GFP or ACP-*Pfrpl*4-V5 revealed no detectable protein expression (data not shown). Lines transfected with plasmids expressing SP-*Pfrpl4*-GFP fusions yielded cultures that expressed recombinant PfRpl4-GFP fusions. However, fluorescence microscopy revealed cytosolic localization of these fusions proteins (**Fig. C-1**), indicating that the combined SP and *Pfrpl*4 putative transit peptide sequence were insufficient for apicoplast targeting. These experiments were repeated on several occasions with the same outcomes, independent of which promoter (*calmodulin* or *hsp86*) was used to drive the transgene or whether the *Pfrpl*4 sequence was mutant or wild-type. We speculate that the inability to traffic PfRpl4 to the apicoplast might reflect tight regulation of the apicoplast ribosomal

translation machinery that is incompatible with PfRpl4 overexpression in transfected parasites.

Modeling of mutant PfRPL4 supports a role for G76V in AZ resistance

Recent elucidation of the crystal structure of the large ribosomal subunit (50S) of *D. radiodurans* has provided evidence favoring simultaneous binding of two AZ molecules in the peptide exit tunnel of the 50S ribosomal subunit (Schlunzen et al., 2003). One AZ molecule (AZ-1) interacted exclusively with domains IV and V of the 23S rRNA, whereas a second molecule (AZ-2) interacted with the ribosomal proteins L4 and L22 and the 23S rRNA domain II. To date, AZ is the only macrolide reported to interact directly with these ribosomal proteins. The major source of interactions with AZ-2 involved the formation of hydrogen bonds with residues Gly60 and Thr64 in the *D. radiodurans* L4 loop (corresponding to residues Val71 and Leu75 in PfRPL4; **Table 4-1**). This loop (Arg46-Pro86) shares 32% sequence identity between *D. radiodurans* and the corresponding loop (K57-P97) in PfRPL4. Using the *D. radiodurans* structural data, we modeled the mutant and wild-type PfRPL4 loops as a complex with AZ **(Figure 4-5)**. The wild-type PfRPL4:AZ molecular model is supportive of AZ interactions with PfRPL4 residues Val71 and Leu75 (**Figure 4-5A**).

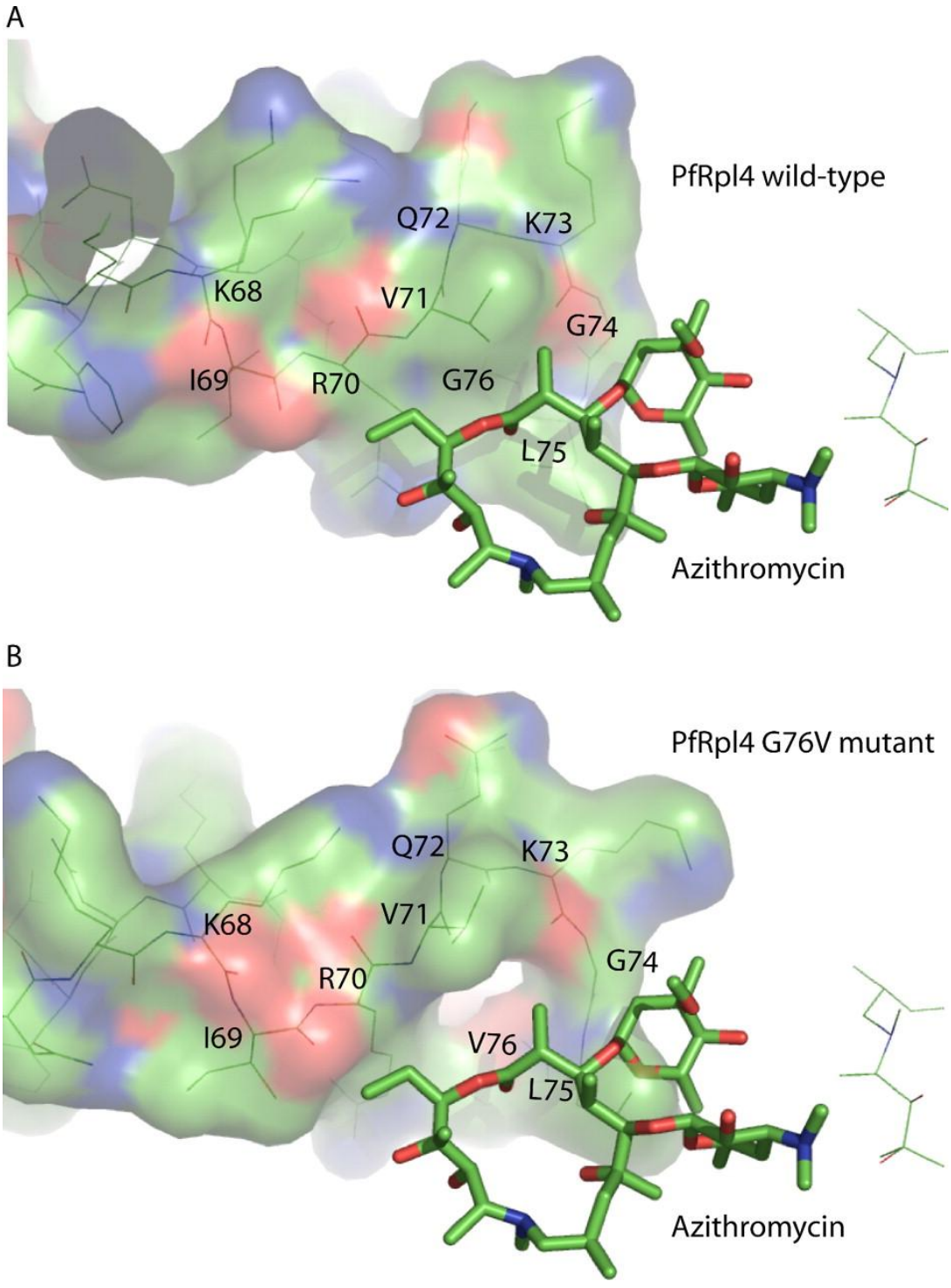


Figure 4-5. Structural models of (A) wild-type and (B) the G76V mutant *Pfrpl4* with AZ
These models, derived using the *D. radiodurans* crystal structure of protein L4 as a template, reveal steric clash between the side chain of 76V in the G76V mutant and AZ, consistent with this mutation conferring AZ resistance.

Superimposition of the mutant PfRPL4 loop structure predicted that the proximal G76V mutation causes a conformational shift in the side chain of Leu75, which would be expected to sterically interfere with the AZ-2 binding pocket **(Figure 4-5B)**. Of all the possible rotamer positions of the Val76 side chain, none were predicted to allow the L75 residue to stay in the wild-type conformation. Thus, these structural models suggest that the G76V mutation causes a conformational change in the neighboring Leu75 residue that creates spatial conflict with AZ binding to this molecule.

Discussion

Our investigations into the antimalarial properties of AZ reveal this to be a slow-acting drug that becomes highly potent when extending parasite exposure from one to two generations, attaining IC_{50} values as low as 50 nM (**Table C-2**). AZ activity was not synergistic with CQ, amodiaquine, or artesunate. Furthermore, AZ potency was unaffected by point mutations in the DV transmembrane proteins PfCRT or PfMDR1 that can reduce susceptibility to several antimalarials that act inside the DV, suggesting that this is not the primary site of action of AZ. Drug pressure studies resulted in the selection of AZ-resistant parasite lines that were found to harbor point mutations in the apicoplast-encoded *Pfrpl4* and LSU rRNA genes involved in apicoplast protein synthesis. These AZ-resistant parasites also displayed cross-resistance to erythromycin, a known inhibitor of the bacterial 50S ribosomal subunit (Canu et al., 2002; Chittum and Champney, 1994; Farrell et al., 2004; Nagai et al., 2002; Pihlajamaki et al., 2002; Tait-Kamradt et al., 2000a). Collectively, these data provide compelling evidence that the

antimalarial properties of AZ are a result of its binding to the apicoplast 50S ribosomal subunit and inhibiting protein synthesis in this organelle. These data complement separate studies directly implicating the apicoplast as the target of tetracycline and doxycycline in *P. falciparum* and clindamycin in *T. gondii* (Camps et al., 2002; Dahl et al., 2006).

In our study, AZ potency increased 30–230-fold upon prolonged exposure (from 48 to 96 h). These data are consistent with an earlier study (Yeo and Rieckmann, 1995) and provide a marked contrast to the comparator drugs thiostrepton and triclosan (that are both also thought to act in the apicoplast (McConkey et al., 1997; McLeod et al., 2001; Perozzo et al., 2002; Rogers et al., 1997)) and CQ (that acts in the DV (Saliba et al., 1998)), which showed only minor gains in potency (**Fig. 4-1**). This "delayed death" phenotype was first described with *T. gondii* tachyzoites exposed to clindamycin and AZ (Fichera et al., 1995). These antibiotics were found to not affect replication in the first generation of exposure, yet drastically reduced tachyzoite division in the second generation (Camps et al., 2002; Fichera et al., 1995). One proposed explanation for this phenotype was that these antibiotics render the parasites unable to establish a parasitophorous vacuole upon host cell reinvasion (Fichera et al., 1995; Fichera and Roos, 1997). Alternatively, their effect on apicoplast protein synthesis might lead to insufficient levels of an as yet unidentified, apicoplast-encoded parasite factor required for the successful development of the progeny of drug-treated parasites (Dahl et al., 2006; Fichera et al., 1995; Ralph et al., 2004).

Our *in vitro* studies reveal essentially additive effects between AZ and CQ, which is in contrast to earlier reports of *in vitro* synergy between these drugs with some *P. falciparum* lines (Ohrt et al., 2002). This discrepancy may be attributable to differences in methodology or analysis, or differences between the lines employed and for how long they had been adapted to *in vitro* culture. Our lack of synergy, reported with five lines tested on five separate occasions, provides evidence that these drugs have separate modes of action. AZ and CQ may nevertheless manifest some synergistic effects when used clinically. This is evidenced by a recent clinical trial in India in an area with a high prevalence of CQ-resistant malaria, which found that AZ-CQ combination therapy was significantly more efficacious in treating falciparum malaria than monotherapy with either agent (97% cure rates with AZ-CQ *versus* 33 and 27% with these individual agents, respectively (Dunne et al., 2005a)).

Direct evidence that AZ resistance in *P. falciparum* is mediated by changes in the bacterial-like apicoplast ribosome comes from our resistanct mutant *in vitro* selection studies, which identified mutations in the apicoplast-encoded *Pfrpl4* and LSU rRNA genes. A key role for *Pfrpl4* is suggested by our finding that both the drug-pressured 7G8 and Dd2 lines underwent a G76V mutation in this gene. Similarly placed mutations in genes encoding L4 have been strongly implicated in macrolide resistance in several bacterial species (**Table 4-1**). For example, two studies observed G69C or G71R mutations in this gene in *S. pneumoniae* strains selected for AZ resistance (Canu et al., 2002; Tait-Kamradt et al., 2000a). These mutations correspond to residue Gly74 and Gly76 in PfRPL4 (**Table 4-1**). Other AZ-resistant *S. pneumoniae* strains also show

insertions, deletions, or multiple mutations in this L4 conserved site (Farrell et al., 2004). Similar L4 mutations have also been identified in macrolide-resistant strains of *Streptococcus pyogenes*, *H. influenzae*, *Staphylococcus aureus* and *E. coli* (Chittum and Champney, 1994; Franceschi et al., 2004; Prunier et al., 2003). Whereas many of those studies only reported an association, genetic complementation studies have provided convincing evidence that L4 mutations can confer AZ resistance. For example, genetic transformation of the AZ-sensitive *S. pneumoniae* R6 strain with the mutant *rpl4* alleles G69C, ^{67}QKG^{69} to QSQKG, ^{69}GTG^{71} to TPS, ^{67}QKGT^{70} to QT, and ^{64}PWRQ^{67} to PQ produced 8-, 15-, 60-, 16-, and 16-fold decreases, respectively, in the susceptibility of these transformed strains to AZ (Tait-Kamradt et al., 2000a; Tait-Kamradt et al., 2000b; Wolter et al., 2005). In the converse approach, complementation of erythromycin-resistant *E. coli* mutants with plasmids expressing wild-type *rpl4* or *rpl22* genes resulted in decreased resistance to erythromycin (Chittum and Champney, 1994). Our study also adopted a complementation approach by trying to express the apicoplast-encoded *Pfrpl*4 gene (both wild and mutant types) in different AZ-sensitive parasite strains. However, despite multiple and varied attempts, we could not obtain recombinant parasite lines that targeted recombinant PfRpl4 to the apicoplast. An alternative and more definitive approach would be to perform allelic exchange of the endogenous *Pfrpl*4 gene with a mutant *Pfrpl*4 sequence that is being replicated inside the apicoplast. This experiment can only be undertaken once there is a suitable selectable marker for transfection of the *P. falciparum* apicoplast.

We note that one of the AZ-resistant lines, AZ-R^{7G8}, also harbored a LSU rRNA mutation (U438C) in domain I. This is away from domains II and V that form part of the macrolide binding pocket in the bacterial 50S ribosomal subunit and that house the majority of the 23S rRNA mutations associated with AZ resistance (Ban et al., 2000). Nonetheless, recent reports document the presence of 23S rRNA domain I mutations (A138G, A260G, A373T, T389C, or A449C) in macrolide-resistant clinical isolates of *S. pneumonia* (Reinert et al., 2005; Reinert et al., 2003). This might contribute to the fact that our AZ-R^{7G8} mutant had a higher AZ IC_{50} than AZ-R^{Dd2}. The AZ-R^{7G8} line also displayed significantly increased *in vitro* sensitivity to thiostrepton, tetracycline, and doxycycline. Thiostrepton, a peptide antibiotic, inhibits protein translation by blocking GTPase activity in the 23S rRNA domain II (Egebjerg et al., 1989) and studies have shown that an A1067 mutation (*E. coli* numbering) can confer thiostrepton resistance in *E. coli* and *P. falciparum* (Rogers et al., 1997; Rosendahl and Douthwaite, 1994). Tetracycline and doxycycline block protein synthesis by binding to the 30S subunit that inhibits attachment of tRNA to the A-site of ribosome (Oehler et al., 1997; Rasmussen et al., 1991). Photolysis experiments performed with radiolabeled tetracycline and *E. coli* ribosomes, however, provide evidence that this antibiotic can also interact with nucleotides in the central loop of the 23S rRNA domain V(Steiner et al., 1988). Domain II and V, however, were invariant in AZ-R^{7G8}, suggesting that the LSU rRNA domain I mutation we observed in this line was responsible for the observed hypersensitivity to thiostrepton, tetracycline, and doxycycline. A possible explanation is provided by structural studies. Analysis of the structure of the large ribosomal subunit from

Haloarcula marismortui complexed with its repertoire of ribosomal proteins showed that the three-dimensional catalytic structure of the 23 S rRNA is maintained by interactions of all six domain subunits with each other through extensive hydrogen bonding and other non-covalent interactions (Ban et al., 2000). Cryo-electron microscopy studies have also shown that mutant L4 proteins can exert long-range structural effects on the 30S subunit in erythromycin–resistant *E. coli* ribosomes (Gabashvili et al., 2001). We posit that the hypersensitivity of AZ-R^{7G8} to thiostrepton, tetracycline, and doxycycline might be due to the LSU rRNA domain I U438C mutation causing structural changes that are transmitted to the antibiotic binding sites on the 50S and 30S rRNA subunits. We also note that this increased antibiotic sensitivity, if found to occur frequently in AZ-resistant parasites, would promote the potential use of these antibiotics as partner drugs for combination therapy. Our findings of mutations in *pfrpl4* and the LSU rRNA are consistent with published data from pathogenic bacteria (including *S. pneumoniae and H. influenzae*) that directly implicate mutations in the 23 S rRNA or the genes encoding ribosomal proteins L4 and L22 as causal determinants of macrolide resistance (Edelstein, 2004; Franceschi et al., 2004; Vester and Douthwaite, 2001). Furthermore, our finding that resistance to AZ could be rapidly selected *in vitro* is an important consideration in the testing of future AZ-containing antimalarial combination therapies.

Other reported mechanisms of macrolide resistance in bacteria are methylation of the A2058 nucleotide in the 23 S rRNA domain V by a methyltransferase encoded by the *erm* (erythromycin ribosome methylation) genes, and drug efflux mediated by an efflux pump encoded by the *mef* (macrolide efflux) genes (Vester and Douthwaite, 2001;

Weisblum, 1995). Methylation of this critical nucleotide acts to reduce macrolide binding to the ribosome, and tends to predominate over point mutations as the major mechanism of macrolide resistance in bacterial species such as *Streptococcus* and *Staphylococcus* that harbor multiple rRNA operons. Our analysis of the recently released *P. falciparum* (HB3 strain) apicoplast genome (GenBank accession number DQ642846, sequenced by the Broad Institute, Cambridge, MA) reveals that this 29-kb genome has only one functional copy of the LSU and small subunit rRNA genes. This resolves earlier questions about the possible presence of a second LSU gene in the apicoplast (Wilson et al., 1996). Our analysis also shows the absence of any putative apicoplast gene coding for an adenosine methyltransferase or an active drug efflux pump. Thus, at this time it would appear less likely that rRNA methylation or drug efflux mechanisms could operate to confer AZ resistance in *P. falciparum.*

L4 and L22 are the early assembly proteins of the 50 S ribosomal subunit and play a pivotal role in scaffolding the ribosomal domains (Gabashvili et al., 2001). The crystal structure of L4 from *Thermotoga maritima* defines a loop of 55 residues that displays the highest degree of phylogenetic conservation across the L4 sequences (corresponding in PfRpl4 to residues 47–101) and that represents the main rRNA binding site (Worbs et al., 2000). RNA-protein cross-linking studies in *E. coli* have shown that this conserved L4 loop interacts with 23 S rRNA nucleotides 284–350 and 580–670 in domains I and II, respectively (Gulle et al., 1988). Cryo-electron microscopy of erythromycin-resistant *E. coli* ribosomes indicates that sequence changes in L4 or L22 can alter the spatial organization of nucleotides in 23 S rRNA domains II, III, and V, and

the conformation of the C-terminal ends of these proteins that form a segment of the polypeptide exit tunnel. These structural rearrangements can significantly change the diameter of the peptide tunnel, affecting macrolide accessibility to the peptidyl-transferase center in domain V and preventing the macrolide from inhibiting protein synthesis and egress from the peptide tunnel (Gabashvili et al., 2001; Gregory and Dahlberg, 1999). Structural analysis of the entire 50 S ribosomal subunit of *H. marismortui* containing L4 and L22 has localized binding of AZ and its progenitor erythromycin to a hydrophobic drug-binding pocket formed by critical 23 S nucleotides A^{2058}, A^{2059}, and C^{2611} (*E. coli* numbering). One key interaction is the formation of a hydrogen bond between the desosamine sugar at position C-5 of the macrolide lactone ring and A^{2058} (Tu et al., 2005). A similar pattern of interactions between AZ and these rRNA nucleotides was also identified in crystallography studies of the *D. radiodurans* large subunit ribosome. This work also identified an additional binding site for AZ, formed by ribosomal proteins L4 and L22 and domain II of the 23 S rRNA (Schlunzen et al., 2003). AZ was observed to directly interact with ribosomal proteins L4 (notably with the Tyr^{59}, Gly^{60}, Gly^{63}, and Thr^{64} residues) and L22 (the Arg^{111} residue) through hydrogen bonding and hydrophobic interactions, respectively (Schlunzen et al., 2003).

Our modeling of wild-type and mutant PfRpl4 , based on structural data from *D. radiodurans* that identified AZ binding to L4 (Schlunzen et al., 2003), predicts a dramatic impact of the G76V mutation on the neighboring Leu75 residue as a result of its reorientation in the mutant protein. This leads us to propose that the reoriented L75 residue can create a steric clash with AZ, resulting in reduced binding. An alternative

explanation is that this G76V mutation prevents AZ from binding to the nascent polypeptide exit tunnel in the 50S ribosome by constricting the macrolide-binding site. In either event, this study provides compelling evidence that AZ acts upon the *P. falciparum* apicoplast and highlights the importance of screening for possible mutations in the LSU rRNA, *Pfrpl4* and *Pfrpl22* genes in ongoing clinical trials designed to assess the efficacy of AZ-containing antimalarial combination therapies.

Thinking outside the crystallography box

To address the molecular mechanism of AZ resistance from the G76V mutation, we first resorted to crystallography. However, the *pfrpl4* cannot be expressed exogenously. Various vectors and *E. coli* strains were tried. None of these yielded a sufficient quantity of soluble pfrpl4 for crystallization trials. **Table 4-2** shows the status statistics of the target proteins for structure determination from the worldwide structure genomics center. Only 7.7% of 184512 cloned targets can be crystallized, 3.7% have yielded diffraction, and eventually 4592 targets (2.5%) have their structures determined by crystallography. Given these statistics, our difficulties are unsurprising.

Table 4-2. TargetDB Status Statistics*

Status	Total Number of Targets	(%) Relative to "Cloned" Targets	(%) Relative to "Expressed" Targets	(%) Relative to "Purified" Targets	(%) Relative to "Crystallized" Targets
Cloned	184512	100.0	-	-	-
Expressed	115773	62.7	100.0	-	-
Soluble	34734	18.8	30.0	-	-
Purified	44826	24.3	38.7	100.0	-
Crystallized	14122	7.7	12.2	31.5	100.0
Diffraction-quality Crystals	6862	3.7	5.9	15.3	48.6
Crystal Structure	4592	2.5	4.0	10.2	32.5

*The data was from http://targetdb.pdb.org, last updated on Jan 21, 2011.

Is it possible to get useful structural information without a crystal? We examined the *pfrpl4* gene more closely. A multiple sequence alignment (**Fig. 4-6**) shows a relatively conserved region of PfRpl4 between Lys57 and Pro97, which contains the motif in which we are interested, ^{71}VQKGLGKAR^{79}. Thus we turned to homology modeling for the solution. The structures of L4 from *E. coli* (pdb: 2AWB) and *D. radioduran* (pdb:1NWY) were chosen as multiple templates to model that loop in Modeller. The sequence identity between PfRpL4 and *E. coli* L4 is the highest in that region (39%). The resolution of 2AWB is 3.5Å. The sequence identity between *D. radioduran* L4 and PfRpl4 is 32%. 1NWY has two azithromycin molecules in the structure. The resolution is 3.3 Å, but only the C_α information was deposited in the pdb. The model of the mutant L4 loop G76V was built in the same way but without azithromycin in the homologous templates.

Figure 4-6. multiple sequence alignment of L4 sequences using *CLUSTAL W*.
2AWB, L4 from *E. coli*; 1DMG, L4 from *Thermotoga maritime*; 1NWY, L4 from *Deinococcus radiodurans*; 1K9M, L4 from *Haloarcula marismortui*; *pf* L4, L4 from *Plasmodium falciparum* apicoplast.

To assess the accuracy of the homologous model, the DOPE (Discrete Optimized Protein Energy) potentials were calculated during the modeling (Shen and Sali, 2006).

Both models yielded negative values (**Fig. 4-7**), which indicated that the models satisfied the stereochemistry constraints. The two models, pfL4 and G76V, were aligned in pyMol. The RMS between them is 1.68 Å. We found that the conformational change in G76V causes a collision between Leu75 and the second azithromycin (**Fig. 4-5**), which we ascribed to the azithromycin resistance of the mutant.

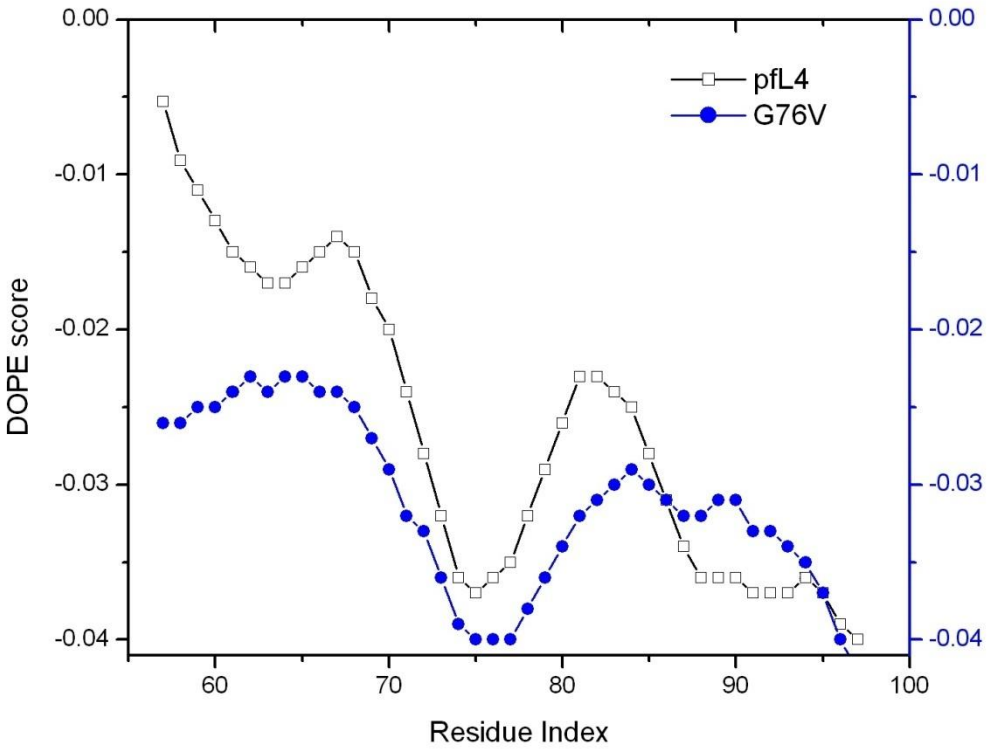


Figure 4-7. DOPE profiles of the homologous models pfL4 and G76V.
The DOPE scores were calculated residue-by-residue for the homologous model of pfL4 (in black) and G76V (in blue).

This modeling structure was based upon a strong, yet reasonable hypothesis that Pfrpl4 shares the common feature with its *D. radiodurans* homolog, which was supported by sequence conservation (**Table 4-1**), and their similar ribosomal environments. This model has illustrated, but not proved the molecular mechanism of AZ potency and how resistance arises in the mutants. This mechanism is consistent with both the observations in genetics and pharmacology, and the stereochemistry knowledge up to date. Also, the power of science resides in its capability to predict. From our model of PfRpl4, we propose that the mutation of Leu75 into a less bulky amino acid, say alanine, would abolish the resistance effect to AZ. This could be tested in future genetic experiments.

CHAPTER V

SUMMARY

Chapter I is the introduction of the main method, X-ray crystallography, and the major topic, phage lysis, of this thesis. In the crystallography section, the focus is on the flow of information: how to transform the atomic structural information to the macroscopic observation; why the structural information is initially lost with the undetermination of the phases; and how the phase information is rescued with *a priori* knowledge. In brief, crystallography is a magnifier, amplifier, sampler, and compiler as described. As to the phage lysis, the classic λ strategy is introduced. At the beginning of each following chapters, there is also a short background more specifically related to the respective project.

Chapter II addresses the activation mechanism of R^{21}, the lysozyme of coliphage 21, which has an N-terminal signal-anchor-release (SAR) domain that directs its secretion in a membrane-tethered, inactive form and its release and activation in the periplasm. Two crystal structures, combined with genetics studies, show that the SAR domain, once extracted from the bilayer, refolds into the body of the enzyme and effects muralytic activation by repositioning one residue of the canonical lysozyme catalytic triad. The NMR experiments not only support the crystal model, but also provide preliminary data for the possible dynamics involved.

Chapter III explains how the phages overcome the unique cell wall of mycobacteria, which is composed of an arabinogalactan-peptidoglycan and mycolic

acid-rich lipid layer. The crystal structure allowed us to determine that the gp12 from mycobacteriophage D29, D29 LysB, has an α/β hydrolase organization of the catalytic triad common to cutinases, but contains an additional four helix domain implicated in the binding of lipid substrates. Further biochemical and genetic studies led to the proposal that LysB facilitates lysis by compromising the integrity of the mycobacterial outer membrane linkage to the arabinogalactan-peptidoglycan layer.

Chapter IV strays from the crystallography and phage lysis, and discusses medicine in malaria. Azithromycin (AZ), a broad-spectrum antibacterial macrolide that inhibits protein synthesis, also manifests reasonable efficacy as an antimalarial. In the absence of a crystal structure, homology modeling of *Plasmodium falciparum* ribosomal protein L4 helps to interpret the pharmacologic observance, and suggests an action mode of azithromycin in malaria.

REFERENCES

Adams, P.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). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D: Biol. Crystallogr. *58*, 1948-1954.

Anderson, S.L., Berman, J., Kuschner, R., Wesche, D., Magill, A., Wellde, B., Schneider, I., Dunne, M., and Schuster, B.G. (1995). Prophylaxis of Plasmodium falciparum malaria with azithromycin administered to volunteers. Ann. Intern. Med. *123*, 771-773.

Andersen, S.L., Ager, A., McGreevy, P., Schuster, B.G., Wesche, D., Kuschner, R., Ohrt, C., Ellis, W., Rossan, R., and Berman, J. (1995). Activity of azithromycin as a blood schizonticide against rodent and human plasmodia in vivo. Am. J. Trop. Med. Hyg. *52*, 159-161.

Arpigny, J.L., and Jaeger, K.E. (1999). Bacterial lipolytic enzymes: classification and properties. Biochem. J. *343 Pt. 1*, 177-183.

Asakura, M., Hinenoya, A., Alam, M.S., Shima, K., Zahid, S.H., Shi, L., Sugimoto, N., Ghosh, A.N., Ramamurthy, T., Faruque, S.M.*, et al.* (2007). An inducible lambdoid prophage encoding cytolethal distending toxin (Cdt-I) and a type III effector protein in enteropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. *104*, 14483-14488.

Baird, J.K. (2005). Effectiveness of antimalarial drugs. N. Engl. J. Med. *352*, 1565-1577.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science *289*, 905-920.

Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P.B., and Steitz, T.A. (1999). Placement of protein and RNA structures into a 5 A-resolution map of the 50S ribosomal subunit. Nature *400*, 841-847.

Banerjee, R., and Goldberg, D.E. (2001). The *Plasmodium* food vacuole. In Antimalarial chemotherapy, P.J. Rosenthal, ed. (Totowa, NJ, Humana Press), pp. 43-63.

Bell, J.A., Wilson, K.P., Zhang, X.J., Faber, H.R., Nicholson, H., and Matthews, B.W. (1991). Comparison of the crystal structure of bacteriophage T4 lysozyme at low, medium, and high ionic strengths. Proteins *10*, 10-21.

Berry, J., Summer, E.J., Struck, D.K., and Young, R. (2008). The final step in the phage infection cycle: the Rz and Rz1 lysis proteins link the inner and outer membranes. Mol. Microbiol. *70*, 341-351.

Besra, G.S. (1998). Preparations of cell wall fractions of mycobacteria. In Mycobacteria Protocols, T. Parish, and N.G. Stoker, eds. (Towata, NJ, Humana Press Inc.), pp. 91-107.

Breitbart, M., and Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? Trends Microbiol. *13*, 278-284.

Brennan, P.J. (2003). Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. Tuberculosis (Edinb) *83*, 91-97.

Brennan, P.J., and Nikaido, H. (1995). The envelope of mycobacteria. Annu. Rev. Biochem. *64*, 29-63.

Bukovska, G., Klucar, L., Vlcek, C., Adamovic, J., Turna, J., and Timko, J. (2006). Complete nucleotide sequence and genome analysis of bacteriophage BFK20: a lytic phage of the industrial producer *Brevibacterium flavum*. Virology *348*, 57-71.

Camps, M., Arrizabalaga, G., and Boothroyd, J. (2002). An rRNA mutation identifies the apicoplast as the target for clindamycin in Toxoplasma gondii. Mol. Microbiol. *43*, 1309-1318.

Canu, A., Malbruny, B., Coquemont, M., Davies, T.A., Appelbaum, P.C., and Leclercq, R. (2002). Diversity of ribosomal mutations conferring resistance to macrolides, clindamycin, streptogramin, and telithromycin in Streptococcus pneumoniae. Antimicrob. Agents Chemother. *46*, 125-131.

Carvalho, C.M., Aires-Barros, M.R., and Cabral, J.M. (1999). Cutinase: from molecular level to bioprocess development. Biotechnol. Bioeng. *66*, 17-34.

Cavanagh, J. (2007). Protein NMR spectroscopy: principles and practice, 2nd edn (Amsterdam ; Boston, Academic Press).

Chen, C.L., Pan, T.Y., Kan, S.C., Kuan, Y.C., Hong, L.Y., Chiu, K.R., Sheu, C.S., Yang, J.S., Hsu, W.H., and Hu, H.Y. (2008). Genome sequence of the lytic bacteriophage P1201 from *Corynebacterium glutamicum* NCHU 87078: evolutionary relationships to phages from Corynebacterineae. Virology *378*, 226-232.

Chittum, H.S., and Champney, W.S. (1994). Ribosomal protein gene sequence changes in erythromycin-resistant mutants of Escherichia coli. J. Bacteriol. *176*, 6192-6198.

Clark, C., Bozdogan, B., Peric, M., Dewasse, B., Jacobs, M.R., and Appelbaum, P.C. (2002). In vitro selection of resistance in Haemophilus influenzae by amoxicillin-

clavulanate, cefpodoxime, cefprozil, azithromycin, and clarithromycin. Antimicrob. Agents Chemother. *46*, 2956-2962.

Cooper, R.A., Ferdig, M.T., Su, X.Z., Ursos, L.M., Mu, J., Nomura, T., Fujioka, H., Fidock, D.A., Roepe, P.D., and Wellems, T.E. (2002). Alternative mutations at position 76 of the vacuolar transmembrane protein PfCRT are associated with chloroquine resistance and unique stereospecific quinine and quinidine responses in *Plasmodium falciparum*. Mol. Pharmacol. *61*, 35-42.

Crabb, B.S., Triglia, T., Waterkeyn, J.G., and Cowman, A.F. (1997). Stable transgene expression in *Plasmodium falciparum*. Mol. Biochem. Parasitol. *90*, 131-144.

Dahl, E.L., Shock, J.L., Shenai, B.R., Gut, J., DeRisi, J.L., and Rosenthal, P.J. (2006). Tetracyclines specifically target the apicoplast of the malaria parasite Plasmodium falciparum. Antimicrob. Agents Chemother. *50*, 3124-3131.

Dauter, Z. (2003). Protein structures at atomic resolution. Methods Enzymol. *368*, 288-337.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. Nmr. *6*, 277-293.

Derewenda, U., Brzozowski, A.M., Lawson, D.M., and Derewenda, Z.S. (1992). Catalysis at the interface: the anatomy of a conformational change in a triglyceride lipase. Biochemistry *31*, 1532-1541.

Desjardins, R.E., Canfield, C.J., Haynes, J.D., and Chulay, J.D. (1979). Quantitative assessment of antimalarial activity in vitro by a semiautomated microdilution technique. Antimicrob. Agents Chemother. *16*, 710-718.

Douthwaite, S., Hansen, L.H., and Mauvais, P. (2000). Macrolide-ketolide inhibition of MLS-resistant ribosomes is improved by alternative drug interaction with domain II of 23S rRNA. Mol. Microbiol. *36*, 183-193.

Drenth, J., and Mesters, J. (2007). Principles of protein x-ray crystallography, 3rd edn (New York, Springer).

Dunne, M.W., Singh, N., Shukla, M., Valecha, N., Bhattacharyya, P.C., Dev, V., Patel, K., Mohapatra, M.K., Lakhani, J., Benner, R.*, et al.* (2005a). A multicenter study of azithromycin, alone and in combination with chloroquine, for the treatment of acute uncomplicated Plasmodium falciparum malaria in India. J. Infect. Dis. *191*, 1582-1588.

Dunne, M.W., Singh, N., Shukla, M., Valecha, N., Bhattacharyya, P.C., Patel, K., Mohapatra, M.K., Lakhani, J., Devi, C.U., Adak, T.*, et al.* (2005b). A double-blind,

randomized study of azithromycin compared to chloroquine for the treatment of *Plasmodium vivax* malaria in India. Am. J. Trop. Med. Hyg. *73*, 1108-1111.

Edelstein, P.H. (2004). Pneumococcal resistance to macrolides, lincosamides, ketolides, and streptogramin B agents: molecular mechanisms and resistance phenotypes. Clin. Infect. Dis. *38 Suppl. 4*, S322-327.

Egebjerg, J., Douthwaite, S., and Garrett, R.A. (1989). Antibiotic interactions at the GTPase-associated centre within Escherichia coli 23S rRNA. Embo. J. *8*, 607-611.

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

Farrar, M.D., Howson, K.M., Bojar, R.A., West, D., Towler, J.C., Parry, J., Pelton, K., and Holland, K.T. (2007). Genome sequence and analysis of a *Propionibacterium acnes* bacteriophage. J. Bacteriol. *189*, 4161-4167.

Farrell, D.J., Morrissey, I., Bakker, S., Buckridge, S., and Felmingham, D. (2004). In vitro activities of telithromycin, linezolid, and quinupristin-dalfopristin against Streptococcus pneumoniae with macrolide resistance due to ribosomal mutations. Antimicrob. Agents Chemother. *48*, 3169-3171.

Feil, D. (2002). The physical origin of anomalous scattering. Crystallography Reviews *8*, 95 - 183.

Feynman, R.P., Hibbs, A.R., and Styer, D.F. (2010). Quantum mechanics and path integrals, Emended edn (Mineola, N.Y., Dover Publications).

Fichera, M.E., and Roos, D.S. (1997). A plastid organelle as a drug target in apicomplexan parasites. Nature *390*, 407-409.

Fichera, M.E., Bhopale, M.K., and Roos, D.S. (1995). In vitro assays elucidate peculiar kinetics of clindamycin action against Toxoplasma gondii. Antimicrob. Agents Chemother. *39*, 1530-1537.

Fidock, D.A., and Wellems, T.E. (1997). Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil. Proc. Natl. Acad. Sci. U. S. A. *94*, 10931-10936.

Fidock, D.A., Nomura, T., and Wellems, T.E. (1998). Cycloguanil and its parent compound proguanil demonstrate distinct activities against Plasmodium falciparum malaria parasites transformed with human dihydrofolate reductase. Mol. Pharmacol. *54*, 1140-1147.

Fidock, D.A., Nomura, T., Talley, A.K., Cooper, R.A., Dzekunov, S.M., Ferdig, M.T., Ursos, L.M., Sidhu, A.B., Naude, B., Deitsch, K.W.*, et al.* (2000). Mutations in the P. falciparum digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. Mol. Cell *6*, 861-871.

Fischetti, V.A. (2008). Bacteriophage lysins as effective antibacterials. Curr. Opin. Microbiol. *11*, 393-400.

Foley, M., and Tilley, L. (1998). Quinoline antimalarials: mechanisms of action and resistance and prospects for new agents. Pharmacol. Ther. *79*, 55-87.

Ford, M.E., Sarkis, G.J., Belanger, A.E., Hendrix, R.W., and Hatfull, G.F. (1998). Genome structure of mycobacteriophage D29: implications for phage evolution. J. Mol. Biol. *279*, 143-164.

Foth, B.J., Ralph, S.A., Tonkin, C.J., Struck, N.S., Fraunholz, M., Roos, D.S., Cowman, A.F., and McFadden, G.I. (2003). Dissecting apicoplast targeting in the malaria parasite Plasmodium falciparum. Science *299*, 705-708.

Franceschi, F., Kanyo, Z., Sherer, E.C., and Sutcliffe, J. (2004). Macrolide resistance from the ribosome perspective. Curr. Drug Targets Infect. Disord. *4*, 177-191.

Gabashvili, I.S., Gregory, S.T., Valle, M., Grassucci, R., Worbs, M., Wahl, M.C., Dahlberg, A.E., and Frank, J. (2001). The polypeptide tunnel system in the ribosome and its gating in erythromycin resistance mutants of L4 and L22. Mol. Cell *8*, 181-188.

Gaffney, K.J., and Chapman, H.N. (2007). Imaging atomic structure and dynamics with ultrafast x-ray scattering. Science *316*, 1444-1448.

Garcia, M., Pimentel, M., and Moniz-Pereira, J. (2002). Expression of Mycobacteriophage Ms6 lysis genes is driven by two sigma(70)-like promoters and is dependent on a transcription termination signal present in the leader RNA. J. Bacteriol. *184*, 3034-3043.

Gardner, M.J., Feagin, J.E., Moore, D.J., Rangachari, K., Williamson, D.H., and Wilson, R.J. (1993). Sequence and organization of large subunit rRNA genes from the extrachromosomal 35 kb circular DNA of the malaria parasite Plasmodium falciparum. Nucleic Acids Res. *21*, 1067-1071.

Gil, F., Catalao, M.J., Moniz-Pereira, J., Leandro, P., McNeil, M., and Pimentel, M. (2008). The lytic cassette of mycobacteriophage Ms6 encodes an enzyme with lipolytic activity. Microbiology *154*, 1364-1371.

Gilham, D., and Lehner, R. (2005). Techniques to measure lipase and esterase activity in vitro. Methods *36*, 139-147.

Gingras, B.A., and Jensen, J.B. (1992). Activity of azithromycin (CP-62,993) and erythromycin against chloroquine-sensitive and chloroquine-resistant strains of Plasmodium falciparum in vitro. Am. J. Trop. Med. Hyg. *47*, 378-382.

Gingras, B.A., and Jensen, J.B. (1993). Antimalarial activity of azithromycin and erythromycin against Plasmodium berghei. Am. J. Trop. Med. Hyg. *49*, 101-105.

Girard, A.E., Girard, D., English, A.R., Gootz, T.D., Cimochowski, C.R., Faiella, J.A., Haskell, S.L., and Retsema, J.A. (1987). Pharmacokinetic and in vivo studies with azithromycin (CP-62,993), a new macrolide with an extended half-life and excellent tissue distribution. Antimicrob. Agents Chemother. *31*, 1948-1954.

Goodyer, I.D., and Taraschi, T.F. (1997). *Plasmodium falciparum*: a simple, rapid method for detecting parasite clones in microtiter plates. Exp. Parasitol. *86*, 158-160.

Gregory, S.T., and Dahlberg, A.E. (1999). Erythromycin resistance mutations in ribosomal proteins L22 and L4 perturb the higher order structure of 23 S ribosomal RNA. J. Mol. Biol. *289*, 827-834.

Gründling, A., Manson, M.D., and Young, R. (2001). Holins kill without warning. Proc. Natl. Acad. Sci. U. S. A. *98*, 9348-9352.

Grzesiek, S., and Bax, A. (1993). Amino-Acid Type Determination in the Sequential Assignment Procedure of Uniformly C-13/N-15-Enriched Proteins. J. Biomol. Nmr. *3*, 185-204.

Gulle, H., Hoppe, E., Osswald, M., Greuer, B., Brimacombe, R., and Stoffler, G. (1988). RNA-protein cross-linking in Escherichia coli 50S ribosomal subunits: determination of sites on 23S RNA that are cross-linked to proteins L2, L4, L24 and L27 by treatment with 2-iminothiolane. Nucleic Acids Res. *16*, 815-832.

Gupta, R., Gupta, N., and Rathi, P. (2004). Bacterial lipases: an overview of production, purification and biochemical properties. Appl. Microbiol. Biotechnol. *64*, 763-781.

Hamid, M.E., Minnikin, D.E., and Goodfellow, M. (1993). A simple chemical test to distinguish mycobacteria from other mycolic-acid-containing actinomycetes. J. Gen. Microbiol. *139*, 2203-2213.

Hansen, J.L., Ippolito, J.A., Ban, N., Nissen, P., Moore, P.B., and Steitz, T.A. (2002). The structures of four macrolide antibiotics bound to the large ribosomal subunit. Mol. Cell *10*, 117-128.

Hardy, L.W., and Poteete, A.R. (1991). Reexamination of the role of Asp20 in catalysis by bacteriophage T4 lysozyme. Biochemistry *30*, 9457-9463.

Hatfull, G.F., Pedulla, M.L., Jacobs-Sera, D., Cichon, P.M., Foley, A., Ford, M.E., Gonda, R.M., Houtz, J.M., Hryckowian, A.J., Kelchner, V.A.*, et al.* (2006). Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. PLoS. Genet. *2*, e92.

Hatfull, G.F., and Sarkis, G.J. (1993). DNA sequence, structure and gene expression of mycobacteriophage L5: a phage system for mycobacterial genetics. Mol. Microbiol. *7*, 395-405.

Hausmann, R. (1988). The T7 group. In The Bacteriophages, R. Calendar, ed. (New York, Plenum), pp. 259-289.

Hoffmann, C., Leis, A., Niederweis, M., Plitzko, J.M., and Engelhardt, H. (2008). Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. Proc. Natl. Acad. Sci. U. S. A. *105*, 3963-3967.

Holm, L., Kaariainen, S., Rosenstrom, P., and Schenkel, A. (2008). Searching protein structure databases with DaliLite v.3. Bioinformatics *24*, 2780.

Imada, M., and Tsugita, A. (1971). Amino-acid sequence of lambda phage endolysin. Nature New Biology *233*, 230-231.

Jerri, A.J. (1977). Shannon sampling theorem: its various extensions and applications - Tutorial Review. P. Ieee. *65*, 1565-1596.

Johnson, D.J., Fidock, D.A., Mungthin, M., Lakshmanan, V., Sidhu, A.B., Bray, P.G., and Ward, S.A. (2004). Evidence for a central role for PfCRT in conferring *Plasmodium falciparum* resistance to diverse antimalarial agents. Mol. Cell *15*, 867-877.

Kain, K.C., Shanks, G.D., and Keystone, J.S. (2001). Malaria chemoprophylaxis in the age of drug resistance. I. Currently recommended drug regimens. Clin. Infect. Dis. *33*, 226-234.

Karakousis, P.C., Bishai, W.R., and Dorman, S.E. (2004). *Mycobacterium tuberculosis* cell envelope lipids and the host immune response. Cell Microbiol. *6*, 105-116.

Karle, J. (1986). Recovering phase information from intensity data. Science *232*, 837-843.

Krupovic, M., Cvirkaite-Krupovic, V., and Bamford, D.H. (2008). Identification and functional analysis of the Rz/Rz1-like accessory lysis genes in the membrane-containing bacteriophage PRD1. Mol. Microbiol. *68*, 492-503.

Kuschner, R.A., Heppner, D.G., Andersen, S.L., Wellde, B.T., Hall, T., Schneider, I., Ballou, W.R., Foulds, G., Sadoff, J.C., Schuster, B.*, et al.* (1994). Azithromycin prophylaxis against a chloroquine-resistant strain of Plasmodium falciparum. Lancet *343*, 1396-1397.

Lakshmanan, V., Bray, P.G., Verdier-Pinard, D., Johnson, D.J., Horrocks, P., Muhle, R.A., Alakpa, G.E., Hughes, R.H., Ward, S.A., Krogstad, D.J.*, et al.* (2005). A critical role for PfCRT K76T in *Plasmodium falciparum* verapamil-reversible chloroquine resistance. EMBO. J. *24*, 2294-2305.

Landau, L.D., and Lifshits, E.M. (1969). Statistical physics, 2d rev. and enl. edn (Oxford, New York,, Pergamon Press).

Leclercq, R. (2002). Mechanisms of resistance to macrolides and lincosamides: nature of the resistance elements and their clinical implications. Clin. Infect. Dis. *34*, 482-492.

Lei, S.P., Lin, H.C., Wang, S.S., Callaway, J., and Wilcox, G. (1987). Characterization of the *Erwinia carotovora* pelB gene and its product pectate lyase. J. Bacteriol. *169*, 4379-4383.

Lell, B., and Kremsner, P.G. (2002). Clindamycin as an antimalarial drug: review of clinical trials. Antimicrob. Agents Chemother. *46*, 2315-2320.

Li, S.C., Goto, N.K., Williams, K.A., and Deber, C.M. (1996). Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment. Proc. Natl. Acad. Sci. U. S. A. *93*, 6676-6681.

Loeffler, J.M., Nelson, D., and Fischetti, V.A. (2001). Rapid killing of *Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase. Science *294*, 2170-2172.

Loessner, M.J., Gaeng, S., and Scherer, S. (1999). Evidence for a holin-like protein gene fully embedded out of frame in the endolysin gene of *Staphylococcus aureus* bacteriophage 187. J. Bacteriol. *181*, 4452-4460.

Loessner, M.J., Gaeng, S., Wendlinger, G., Maier, S.K., and Scherer, S. (1998). The two-component lysis system of *Staphylococcus aureus* bacteriophage Twort: a large TTG-start holin and an associated amidase endolysin. FEMS. Microbiol. Lett. *162*, 265-274.

Longhi, S., and Cambillau, C. (1999). Structure-activity of cutinase, a small lipolytic enzyme. Biochim. Biophys. Acta *1441*, 185-196.

Longhi, S., Nicolas, A., Creveld, L., Egmond, M., Verrips, C.T., de Vlieg, J., Martinez, C., and Cambillau, C. (1996). Dynamics of Fusarium solani cutinase investigated

through structural comparison among different crystal forms of its variants. Proteins *26*, 442-458.

Looareesuwan, S., Viravan, C., Webster, H.K., Kyle, D.E., Hutchinson, D.B., and Canfield, C.J. (1996). Clinical studies of atovaquone, alone or in combination with other antimalarial drugs, for treatment of acute uncomplicated malaria in Thailand. Am. J. Trop. Med. Hyg. *54*, 62-66.

Marinelli, L.J., Piuri, M., Swigonova, Z., Balachandran, A., Oldfield, L.M., van Kessel, J.C., and Hatfull, G.F. (2008). BRED: a simple and powerful tool for constructing mutant and recombinant bacteriophage genomes. PLoS. ONE *3*, e3957.

Masaki, K., Kamini, N.R., Ikeda, H., and Iefuji, H. (2005). Cutinase-like enzyme from the yeast *Cryptococcus* sp. strain S-2 hydrolyzes polylactic acid and other biodegradable plastics. Appl. Environ. Microbiol. *71*, 7548-7550.

McConkey, G.A., Rogers, M.J., and McCutchan, T.F. (1997). Inhibition of *Plasmodium falciparum* protein synthesis: Targeting the plastid-like organelle with thiostrepton. J. Biol. Chem. *272*, 2046-2049.

McLeod, R., Muench, S.P., Rafferty, J.B., Kyle, D.E., Mui, E.J., Kirisits, M.J., Mack, D.G., Roberts, C.W., Samuel, B.U., Lyons, R.E*., et al.* (2001). Triclosan inhibits the growth of Plasmodium falciparum and Toxoplasma gondii by inhibition of apicomplexan Fab I. Int. J. Parasitol. *31*, 109-113.

McNeil, M., Daffe, M., and Brennan, P.J. (1990). Evidence for the nature of the link between the arabinogalactan and peptidoglycan of mycobacterial cell walls. J. Biol. Chem. *265*, 18200-18206.

McNeil, M., Daffe, M., and Brennan, P.J. (1991). Location of the mycolyl ester substituents in the cell walls of mycobacteria. J. Biol. Chem. *266*, 13217-13223.

Mediavilla, J., Jain, S., Kriakov, J., Ford, M.E., Duda, R.L., Jacobs, W.R., Jr., Hendrix, R.W., and Hatfull, G.F. (2000). Genome organization and characterization of mycobacteriophage Bxb1. Mol. Microbiol. *38*, 955-970.

Melo, F., Sanchez, R., and Sali, A. (2002). Statistical potentials for fold assessment. Protein Sci. *11*, 430-448.

Mooers, B.H., and Matthews, B.W. (2006). Extension to 2268 atoms of direct methods in the *ab initio* determination of the unknown structure of bacteriophage P22 lysozyme. Acta Crystallogr. D: Biol. Crystallogr. *62*, 165-176.

Morris, P., Marinelli, L.J., Jacobs-Sera, D., Hendrix, R.W., and Hatfull, G.F. (2008). Genomic characterization of mycobacteriophage Giles: evidence for phage acquisition of host DNA by illegitimate recombination. J. Bacteriol. *190*, 2172-2182.

Nagai, K., Appelbaum, P.C., Davies, T.A., Kelly, L.M., Hoellman, D.B., Andrasevic, A.T., Drukalska, L., Hryniewicz, W., Jacobs, M.R., Kolman, J.*, et al.* (2002). Susceptibility to telithromycin in 1,011 Streptococcus pyogenes isolates from 10 central and Eastern European countries. Antimicrob. Agents Chemother. *46*, 546-549.

Neutze, R., Wouts, R., van der Spoel, D., Weckert, E., and Hajdu, J. (2000). Potential for biomolecular imaging with femtosecond X-ray pulses. Nature *406*, 752-757.

Ng, L.K., Martin, I., Liu, G., and Bryden, L. (2002). Mutation in 23S rRNA associated with macrolide resistance in Neisseria gonorrhoeae. Antimicrob. Agents Chemother. *46*, 3020-3025.

Niederweis, M. (2008). Nutrient acquisition by mycobacteria. Microbiology *154*, 679-692.

Nkrumah, L.N., Muhle, R.A., Moura, P.A., Ghosh, P., Hatfull, G., Jacobs Jr., W.R., and Fidock, D.A. (2006). Efficient site-specific integration in *Plasmodium falciparum* chromosomes mediated by mycobacteriophage Bxb1 integrase. Nature Methods *3*, 615-621.

Noedl, H., Krudsood, S., Chalermratana, K., Silachamroon, U., Leowattana, W., Tangpukdee, N., Looareesuwan, S., Miller, R.S., Fukuda, M., Jongsakul, K.*, et al.* (2006). Azithromycin combination therapy with artesunate or quinine for the treatment of uncomplicated Plasmodium falciparum malaria in adults: a randomized, phase 2 clinical trial in Thailand. Clin. Infect. Dis. *43*, 1264-1271.

O'Brien, A.D., Marques, L.R., Kerry, C.F., Newland, J.W., and Holmes, R.K. (1989). Shiga-like toxin converting phage of enterohemorrhagic *Escherichia coli* strain 933. Microb. Pathog. *6*, 381-390.

O'Neill, P.M., Bray, P.G., Hawley, S.R., Ward, S.A., and Park, B.K. (1998). 4-Aminoquinolines: past, present, and future: a chemical perspective. Pharmacol. Ther. *77*, 29-58.

Ochoa, T.J., Chen, J., Walker, C.M., Gonzales, E., and Cleary, T.G. (2007). Rifaximin does not induce toxin production or phage-mediated lysis of Shiga toxin-producing *Escherichia coli*. Antimicrob. Agents Chemother. *51*, 2837-2841.

Oehler, R., Polacek, N., Steiner, G., and Barta, A. (1997). Interaction of tetracycline with RNA: photoincorporation into ribosomal RNA of Escherichia coli. Nucleic Acids Res. *25*, 1219-1224.

Ohrt, C., Willingmyre, G.D., Lee, P., Knirsch, C., and Milhous, W. (2002). Assessment of azithromycin in combination with other antimalarial drugs against Plasmodium falciparum in vitro. Antimicrob. Agents Chemother. *46*, 2518-2524.

Ohrt, C., Richie, T.L., Widjaja, H., Shanks, G.D., Fitriadi, J., Fryauff, D.J., Handschin, J., Tang, D., Sandjaja, B., Tjitra, E.*, et al.* (1997). Mefloquine compared with doxycycline for the prophylaxis of malaria in Indonesian soldiers: A randomized, double-blind, placebo-controlled trial. Ann. Intern. Med. *126*, 963-972.

Ojha, A.K., Baughn, A.D., Sambandan, D., Hsu, T., Trivelli, X., Guerardel, Y., Alahari, A., Kremer, L., Jacobs, W.R., Jr., and Hatfull, G.F. (2008). Growth of *Mycobacterium tuberculosis* biofilms containing free mycolic acids and harbouring drug-tolerant bacteria. Mol. Microbiol. *69*, 164-174.

Ollis, D.L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J.*, et al.* (1992). The alpha/beta hydrolase fold. Protein Eng. *5*, 197-211.

Pandey, A.K., Raman, S., Proff, R., Joshi, S., Kang, C.M., Rubin, E.J., Husson, R.N., and Sassetti, C.M. (2009). Nitrile-inducible gene expression in mycobacteria. Tuberculosis (Edinb) *89*, 12-16.

Parish, T., Mahenthiralingam, E., Draper, P., Davis, E.O., and Colston, M.J. (1997). Regulation of the inducible acetamidase gene of Mycobacterium smegmatis. Microbiology *143 (Pt 7)*, 2267-2276.

Park, T., Struck, D.K., Dankenbring, C.A., and Young, R. (2007). The pinholin of lambdoid phage 21: control of lysis by membrane depolarization. J. Bacteriol. *189*, 9135-9139.

Park, T., Struck, D.K., Deaton, J.F., and Young, R. (2006). Topological dynamics of holins in programmed bacterial lysis. Proc. Natl. Acad. Sci. U. S. A. *103*, 19713-19718.

Parker, S.K., Curtin, K.M., and Vasil, M.L. (2007). Purification and characterization of mycobacterial phospholipase A: an activity associated with mycobacterial cutinase. J. Bacteriol. *189*, 4153-4160.

Patterson, A.L. (1934). A Fourier series method for the determination of the components of interatomic distances in crystals. Phys. Rev. *46*, 0372-0376.

Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N.R.*, et al.* (2003). Origins of highly mosaic mycobacteriophage genomes. Cell *113*, 171-182.

Perozzo, R., Kuo, M., bir Singh Sidhu, A., Valiyaveettil, J.T., Bittman, R., Jacobs, W.R., Jr., Fidock, D.A., and Sacchettini, J.C. (2002). Structural elucidation of the specificity of the antibacterial agent triclosan for malarial enoyl ACP reductase. J. Biol. Chem. *277*, 13106-13114.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF chimera: A visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605-1612.

Pfefferkorn, E.R., and Borotz, S.E. (1994). Comparison of mutants of Toxoplasma gondii selected for resistance to azithromycin, spiramycin, or clindamycin. Antimicrob. Agents Chemother. *38*, 31-37.

Pham, T.T., Jacobs-Sera, D., Pedulla, M.L., Hendrix, R.W., and Hatfull, G.F. (2007). Comparative genomic analysis of mycobacteriophage Tweety: evolutionary insights and construction of compatible site-specific integration vectors for mycobacteria. Microbiology *153*, 2711-2723.

Pihlajamaki, M., Kataja, J., Seppala, H., Elliot, J., Leinonen, M., Huovinen, P., and Jalava, J. (2002). Ribosomal mutations in Streptococcus pneumoniae clinical isolates. Antimicrob. Agents Chemother. *46*, 654-658.

Piuri, M., and Hatfull, G.F. (2006). A peptidoglycan hydrolase motif within the mycobacteriophage TM4 tape measure protein promotes efficient infection of stationary phase cells. Mol. Microbiol. *62*, 1569-1585.

Plunkett, G., III, Rose, D.J., Durfee, T.J., and Blattner, F.R. (1999). Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. J. Bacteriology. *181*, 1767-1778.

Poehlsgaard, J., and Douthwaite, S. (2005). The bacterial ribosome as a target for antibiotics. Nat. Rev. Microbiol. *3*, 870-881.

Portevin, D., De Sousa-D'Auria, C., Houssin, C., Grimaldi, C., Chami, M., Daffe, M., and Guilhot, C. (2004). A polyketide synthase catalyzes the last condensation step of mycolic acid biosynthesis in mycobacteria and related organisms. Proc. Natl. Acad. Sci. U. S. A. *101*, 314-319.

Prunier, A.L., Malbruny, B., Laurans, M., Brouard, J., Duhamel, J.F., and Leclercq, R. (2003). High rate of macrolide resistance in Staphylococcus aureus strains from patients with cystic fibrosis reveals high proportions of hypermutable strains. J. Infect. Dis. *187*, 1709-1716.

Ralph, S.A., van Dooren, G.G., Waller, R.F., Crawford, M.J., Fraunholz, M.J., Foth, B.J., Tonkin, C.J., Roos, D.S., and McFadden, G.I. (2004). Tropical infectious diseases:

metabolic maps and functions of the Plasmodium falciparum apicoplast. Nat. Rev. Microbiol. *2*, 203-216.

Rasmussen, B., Noller, H.F., Daubresse, G., Oliva, B., Misulovin, Z., Rothstein, D.M., Ellestad, G.A., Gluzman, Y., Tally, F.P., and Chopra, I. (1991). Molecular basis of tetracycline action: identification of analogs whose primary target is not the bacterial ribosome. Antimicrob. Agents Chemother. *35*, 2306-2311.

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

Reed, M.B., Saliba, K.J., Caruana, S.R., Kirk, K., and Cowman, A.F. (2000). Pgh1 modulates sensitivity and resistance to multiple antimalarials in *Plasmodium falciparum*. Nature *403*, 906-909.

Reinert, R.R., Reinert, S., van der Linden, M., Cil, M.Y., Al-Lahham, A., and Appelbaum, P. (2005). Antimicrobial susceptibility of Streptococcus pneumoniae in eight European countries from 2001 to 2003. Antimicrob. Agents Chemother. *49*, 2903-2913.

Reinert, R.R., Wild, A., Appelbaum, P., Lutticken, R., Cil, M.Y., and Al-Lahham, A. (2003). Ribosomal mutations conferring resistance to macrolides in Streptococcus pneumoniae clinical strains isolated in Germany. Antimicrob. Agents Chemother. *47*, 2319-2322.

Rennell, D., Bouvier, S.E., Hardy, L.W., and Poteete, A.R. (1991). Systematic mutation of bacteriophage T4 lysozyme. J. Mol. Biol. *222*, 67-88.

Retsema, J., and Fu, W. (2001). Macrolides: structures and microbial targets. Int. J. Antimicrob. Agents *18 Suppl 1*, S3-10.

Retsema, J., Girard, A., Schelkly, W., Manousos, M., Anderson, M., Bright, G., Borovoy, R., Brennan, L., and Mason, R. (1987). Spectrum and mode of action of azithromycin (CP-62,993), a new 15-membered-ring macrolide with improved potency against gram-negative organisms. Antimicrob. Agents Chemother. *31*, 1939-1947.

Rogers, M.J., Bukhman, Y.V., McCutchan, T.F., and Draper, D.E. (1997). Interaction of thiostrepton with an RNA fragment derived from the plastid-encoded ribosomal RNA of the malaria parasite. Rna *3*, 815-820.

Rosendahl, G., and Douthwaite, S. (1994). The antibiotics micrococcin and thiostrepton interact directly with 23S rRNA nucleotides 1067A and 1095A. Nucleic Acids Res. *22*, 357-363.

Rupp, B. (2010). Biomolecular crystallography (New York, NY, Garland Science).

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. *234*, 779-815.

Saliba, K.J., Folb, P.I., and Smith, P.J. (1998). Role for the plasmodium falciparum digestive vacuole in chloroquine resistance. Biochem. Pharmacol. *56*, 313-320.

Schlunzen, F., Harms, J.M., Franceschi, F., Hansen, H.A., Bartels, H., Zarivach, R., and Yonath, A. (2003). Structural basis for the antibiotic activity of ketolides and azalides. Structure *11*, 329-338.

Schrödinger, E. (1992). What is life? : the physical aspect of the living cell ; with, Mind and matter ; & Autobiographical sketches (Cambridge ; New York, Cambridge University Press).

Schuch, R., Nelson, D., and Fischetti, V.A. (2002). A bacteriolytic agent that detects and kills *Bacillus anthracis*. Nature *418*, 884-889.

Shen, M.Y., and Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. Protein Sci. *15*, 2507-2524.

Sidhu, A.B., Valderramos, S.G., and Fidock, D.A. (2005). pfmdr1 mutations contribute to quinine resistance and enhance mefloquine and artemisinin sensitivity in Plasmodium falciparum. Mol. Microbiol. *57*, 913-926.

Sidhu, A.B., Verdier-Pinard, D., and Fidock, D.A. (2002). Chloroquine resistance in Plasmodium falciparum malaria parasites conferred by pfcrt mutations. Science *298*, 210-213.

Sidhu, A.B.S., Uhlemann, A.-C., Valderramos, S.G., Valderramos, J.-C., Krishna, S., and Fidock, D.A. (2006). Decreasing *pfmdr1* copy number in *Plasmodium falciparum* malaria heightens susceptibility to mefloquine, lumefantrine, halofantrine, quinine, and artemisinin. J. Infect. Dis. *194*, 528-535.

Snapper, S.B., Melton, R.E., Mustafa, S., Kieser, T., and Jacobs, W.R., Jr. (1990). Isolation and characterization of efficient plasmid transformation mutants of Mycobacterium smegmatis. Mol. Microbiol. *4*, 1911-1919.

Steiner, G., Kuechler, E., and Barta, A. (1988). Photo-affinity labelling at the peptidyl transferase centre reveals two different positions for the A- and P-sites in domain V of 23S rRNA. Embo. J. *7*, 3949-3955.

Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. (2006). Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U. S. A. *103*, 8060-8065.

Strynadka, N.C., and James, M.N. (1996). Lysozyme: a model enzyme in protein crystallography. E. X. S. *75*, 185-222.

Summer, E.J., Berry, J., Tran, T.A., Niu, L., Struck, D.K., and Young, R. (2007). Rz/Rz1 lysis gene equivalents in phages of Gram-negative hosts. J. Mol. Biol. *373*, 1098-1112.

Sun, Q., Kuty, G.F., Arockiasamy, A., Xu, M., Young, R., and Sacchettini, J.C. (2009). Regulation of a muralytic enzyme by dynamic membrane topology. Nat. Struct. Mol. Biol. *16*, 1192-1194.

Suttle, C.A. (2007). Marine viruses: major players in the global ecosystem. Nat. Rev. Microbiol. *5*, 801-812.

Tait-Kamradt, A., Davies, T., Cronan, M., Jacobs, M.R., Appelbaum, P.C., and Sutcliffe, J. (2000a). Mutations in 23S rRNA and ribosomal protein L4 account for resistance in pneumococcal strains selected in vitro by macrolide passage. Antimicrob. Agents Chemother. *44*, 2118-2125.

Tait-Kamradt, A., Davies, T., Appelbaum, P.C., Depardieu, F., Courvalin, P., Petitpas, J., Wondrack, L., Walker, A., Jacobs, M.R., and Sutcliffe, J. (2000b). Two new mechanisms of macrolide resistance in clinical strains of Streptococcus pneumoniae from Eastern Europe and North America. Antimicrob. Agents Chemother. *44*, 3395-3401.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673-4680.

Tu, D., Blaha, G., Moore, P.B., and Steitz, T.A. (2005). Structures of MLSBK antibiotics bound to mutated large ribosomal subunits provide a structural explanation for resistance. Cell *121*, 257-270.

Tuinstra, R.L., Peterson, F.C., Kutlesa, S., Elgin, E.S., Kron, M.A., and Volkman, B.F. (2008). Interconversion between two unrelated protein folds in the lymphotactin native state. Proc. Natl. Acad. Sci. U. S. A. *105*, 5057-5062.

Uhlemann, A.-C., Yuthavong, Y., and Fidock, D.A. (2005). Mechanisms of antimalarial drug action and resistance. In Molecular approaches to malaria, I. Sherman, ed. (Washington, DC, ASM Press), pp. 229-261.

Uson, I., and Sheldrick, G.M. (1999). Advances in direct methods for protein crystallography. Curr. Opin. Struct. Biol. *9*, 643-648.

van Kessel, J.C., and Hatfull, G.F. (2007). Recombineering in Mycobacterium tuberculosis. Nat. Methods *4*, 147-152.

Vester, B., and Douthwaite, S. (2001). Macrolide resistance conferred by base substitutions in 23S rRNA. Antimicrob. Agents Chemother. *45*, 1-12.

Vilcheze, C., and Jacobs, W.R., Jr. (2007). The mechanism of isoniazid killing: clarity through the scope of genetics. Annu. Rev. Microbiol. *61*, 35-50.

Wallace, B.A., and Mao, D. (1984). Circular dichroism analyses of membrane proteins: an examination of differential light scattering and absorption flattening effects in large membrane vesicles and membrane sheets. Anal. Biochem. *142*, 317-328.

Waller, R.F., Reed, M.B., Cowman, A.F., and McFadden, G.I. (2000). Protein trafficking to the plastid of Plasmodium falciparum is via the secretory pathway. Embo. J. *19*, 1794-1802.

Waller, R.F., Keeling, P.J., Donald, R.G., Striepen, B., Handman, E., Lang-Unnasch, N., Cowman, A.F., Besra, G.S., Roos, D.S., and McFadden, G.I. (1998). Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. Proc. Natl. Acad. Sci. U. S. A. *95*, 12352-12357.

Wang, I.N., Deaton, J., and Young, R. (2003). Sizing the holin lesion with an endolysin-beta-galactosidase fusion. J. Bacteriol. *185*, 779-787.

Wang, I.N., Dykhuizen, D.E., and Slobodkin, L.B. (1996). The evolution of phage lysis timing. Evol. Ecol. *10*, 545-558.

Wang, I.N., Smith, D.L., and Young, R. (2000). Holins: the protein clocks of bacteriophage infections. Annu. Rev. Microbiol. *54*, 799-825.

Watanabe, M., Aoyagi, Y., Ridell, M., and Minnikin, D.E. (2001). Separation and characterization of individual mycolic acids in representative mycobacteria. Microbiology *147*, 1825-1837.

Weisblum, B. (1995). Erythromycin resistance by ribosome modification. Antimicrob. Agents Chemother. *39*, 577-585.

Wellems, T.E., Panton, L.J., Gluzman, I.Y., do Rosario, V.E., Gwadz, R.W., Walker-Jonah, A., and Krogstad, D.J. (1990). Chloroquine resistance not linked to *mdr*-like genes in a *Plasmodium falciparum* cross. Nature *345*, 253-255.

West, N.P., Chow, F.M., Randall, E.J., Wu, J., Chen, J., Ribeiro, J.M., and Britton, W.J. (2009). Cutinase-like proteins of *Mycobacterium tuberculosis*: characterization of their variable enzymatic functions and active site identification. Faseb. J. *23*, 1694-1704.

White, N.J. (1996). The treatment of malaria. N. Engl. J. Med. *335*, 800-806.

Wilson, R.J., Denny, P.W., Preiser, P.R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D.J., Moore, P.W*., et al.* (1996). Complete gene map of the plastid-like DNA of the malaria parasite Plasmodium falciparum. J. Mol. Biol. *261*, 155-172.

Wolter, N., Smith, A.M., Farrell, D.J., Schaffner, W., Moore, M., Whitney, C.G., Jorgensen, J.H., and Klugman, K.P. (2005). Novel mechanism of resistance to oxazolidinones, macrolides, and chloramphenicol in ribosomal protein L4 of the pneumococcus. Antimicrob. Agents Chemother. *49*, 3554-3557.

Worbs, M., Huber, R., and Wahl, M.C. (2000). Crystal structure of ribosomal protein L4 shows RNA-binding sites for ribosome incorporation and feedback control of the S10 operon. Embo. J. *19*, 807-818.

Xu, M., Struck, D.K., Deaton, J., Wang, I.N., and Young, R. (2004). The signal arrest-release (SAR) sequence mediates export and control of the phage P1 endolysin. Proc. Natl. Acad. Sci. U. S. A. *101*, 6415-6420.

Xu, M., Arulandu, A., Struck, D.K., Swanson, S., Sacchettini, J.C., and Young, R. (2005). Disulfide isomerization after membrane release of its SAR domain activates P1 lysozyme. Science *307*, 113-117.

Yeo, A.E., and Rieckmann, K.H. (1995). Increased antimalarial activity of azithromycin during prolonged exposure of Plasmodium falciparum in vitro. Int. J. Parasitol. *25*, 531-532.

Yonath, A. (2005). Antibiotics targeting ribosomes: resistance, selectivity, synergism and cellular regulation. Annu. Rev. Biochem. *74*, 649-679.

Young, R. (1992). Bacteriophage lysis: mechanism and regulation. Microbiol. Rev. *56*, 430-481.

Young, R., Wang, I., and Roof, W.D. (2000). Phages will out: strategies of host cell lysis. Trends Microbiol. *8*, 120-128.

Zuber, B., Chami, M., Houssin, C., Dubochet, J., Griffiths, G., and Daffe, M. (2008). Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state. J. Bacteriol. *190*, 5672-5680.

APPENDIX A

NMR ASSIGNMENT OF $^{i}R^{21}$

Table A-1. Chemical shifts of backbone atom in $^{i}R^{21}$

Residue	number	Atom	nuclear	shift
Ser	29	N	15N	116.056
Ser	29	H	1H	8.523
Ser	29	CB	13C	63.622
Ser	29	CA	13C	58.055
Gly	30	N	15N	110.536
Gly	30	H	1H	8.402
Gly	30	CA	13C	45.182
Asn	31	N	15N	118.868
Asn	31	H	1H	8.402
Asn	31	CB	13C	38.687
Asn	31	CA	13C	52.859
Asp	32	N	15N	120.264
Asp	32	H	1H	8.397
Asp	32	CB	13C	40.813
Asp	32	CA	13C	54.192
Gly	33	N	15N	108.554
Gly	33	H	1H	8.329
Gly	33	CA	13C	45.193
Leu	34	N	15N	121.459
Leu	34	H	1H	8.101
Leu	34	CB	13C	42.673
Leu	34	CA	13C	54.484
Glu	35	N	15N	122.584
Glu	35	H	1H	8.717
Glu	35	CB	13C	29.845
Glu	35	CA	13C	56.581
Gly	36	N	15N	110.786
Gly	36	H	1H	8.528
Gly	36	CA	13C	45.083
Val	37	N	15N	121.299
Val	37	H	1H	7.871
Val	37	CB	13C	32.812
Val	37	CA	13C	61.074

Table A-1 (continued)

Residue	number	Atom	nuclear	shift
Ser	38	N	15N	119.553
Ser	38	H	1H	7.999
Ser	38	CB	13C	64.598
Ser	38	CA	13C	56.321
Tyr	39	N	15N	125.367
Tyr	39	H	1H	8.958
Tyr	39	CB	13C	37.727
Tyr	39	CA	13C	58.759
Ile	40	N	15N	117.421
Ile	40	H	1H	7.6
Ile	40	CB	13C	41.181
Ile	40	CA	13C	56.809
Pro	41	CB	13C	32.341
Pro	41	CA	13C	62.945
Tyr	42	N	15N	121.842
Tyr	42	H	1H	9.48
Tyr	42	CB	13C	41.021
Tyr	42	CA	13C	55.325
Lys	43	N	15N	121.326
Lys	43	H	1H	8.31
Lys	43	CB	13C	32.988
Asp	44	N	15N	125.375
Asp	44	H	1H	8.096
Asp	44	CB	13C	40.365
Asp	44	CA	13C	52.253
Ile	45	N	15N	114.256
Ile	45	H	1H	7.715
Ile	45	CB	13C	38.332
Ile	45	CA	13C	63.115
Val	46	N	15N	115.265
Val	46	H	1H	7.814
Val	46	CB	13C	30.738
Val	46	CA	13C	60.326
Gly	47	N	15N	109.772
Gly	47	H	1H	8.097
Gly	47	CA	13C	45.311
Val	48	N	15N	123.779
Val	48	H	1H	8.409

Table A-1 (continued)

Residue	number	Atom	nuclear	shift
Val	48	CB	13C	32.18
Val	48	CA	13C	61.566
Trp	49	N	15N	130.064
Trp	49	H	1H	8.442
Trp	49	CB	13C	29.19
Trp	49	CA	13C	58.223
Thr	50	N	15N	119.423
Thr	50	H	1H	9.266
Thr	50	CB	13C	71.054
Thr	50	CA	13C	60.15
Val	51	N	15N	121.674
Val	51	H	1H	8.79
Val	51	CB	13C	35.78
Cys	52	N	15N	120.034
Cys	52	H	1H	8.759
Cys	52	CB	13C	31.21
Cys	52	CA	13C	54.676
His	53	N	15N	126.098
His	53	H	1H	9.013
His	53	CB	13C	27.09
Gly	54	N	15N	107.493
Gly	54	H	1H	7.765
Gly	54	CA	13C	45.262
His	55	N	15N	118.761
His	55	H	1H	7.965
His	55	CB	13C	27.873
His	55	CA	13C	57.167
Thr	56	N	15N	116.099
Thr	56	H	1H	7.57
Thr	56	CB	13C	71.141
Thr	56	CA	13C	59.027
Gly	57	N	15N	106.878
Gly	57	H	1H	8.512
Gly	57	CA	13C	44.221
Lys	58	N	15N	118.502
Lys	58	H	1H	8.511
Lys	58	CB	13C	31.781
Lys	58	CA	13C	57.065

Table A-1 (continued)

Residue	number	Atom	nuclear	shift
Asp	59	N	15N	116.63
Asp	59	H	1H	8.63
Asp	59	CB	13C	39.345
Asp	59	CA	13C	52.88
Ile	60	N	15N	121.598
Ile	60	H	1H	7.126
Ile	60	CB	13C	37.859
Ile	60	CA	13C	63.101
Met	61	N	15N	129.37
Met	61	H	1H	8.037
Met	61	CB	13C	32.257
Met	61	CA	13C	53.209
Leu	62	N	15N	124.422
Leu	62	H	1H	8.152
Leu	62	CB	13C	39.38
Leu	62	CA	13C	56.237
Gly	63	N	15N	108.877
Gly	63	H	1H	8.597
Gly	63	CA	13C	44.971
Lys	64	N	15N	123.781
Lys	64	H	1H	7.415
Lys	64	CB	13C	34.588
Lys	64	CA	13C	55.117
Thr	65	N	15N	119.672
Thr	65	H	1H	8.194
Thr	65	CB	13C	68.263
Thr	65	CA	13C	61.636
Tyr	66	N	15N	102.645
Tyr	66	H	1H	10.099
Tyr	66	CB	13C	40.073
Tyr	66	CA	13C	57.892
Thr	67	N	15N	110.711
Thr	67	H	1H	9.043
Thr	67	CB	13C	71.302
Thr	67	CA	13C	59.363
Lys	68	N	15N	120.512
Lys	68	H	1H	8.861
Lys	68	CB	13C	32.171

Table A-1 (continued)

Residue	number	Atom	nuclear	shift
Lys	68	CA	13C	60.882
Ala	69	N	15N	121.683
Ala	69	H	1H	8.603
Ala	69	CB	13C	17.992
Ala	69	CA	13C	54.9
Glu	70	N	15N	121.29
Glu	70	H	1H	8.263
Glu	70	CB	13C	30.472
Glu	70	CA	13C	59.024
Cys	71	N	15N	116.401
Cys	71	H	1H	9.02
Cys	71	CA	13C	53.195
Lys	72	N	15N	124.957
Lys	72	H	1H	8.498
Lys	72	CB	13C	31.736
Lys	72	CA	13C	59.185
Ala	73	N	15N	122.338
Ala	73	H	1H	8.087
Ala	73	CB	13C	17.445
Ala	73	CA	13C	54.968
Leu	74	N	15N	119.169
Leu	74	H	1H	8.127
Leu	74	CB	13C	42.628
Leu	74	CA	13C	57.237
Leu	75	N	15N	119.765
Leu	75	H	1H	7.543
Leu	75	CB	13C	40.778
Asn	76	N	15N	117.461
Asn	76	H	1H	8.45
Asn	76	CB	13C	37.158
Asn	76	CA	13C	55.73
Lys	77	N	15N	122.178
Lys	77	H	1H	7.985
Lys	77	CB	13C	31.89
Lys	77	CA	13C	59.246
Asp	78	N	15N	122.755
Asp	78	H	1H	8.411
Asp	78	CB	13C	39.304

Table A-1 (continued)

Residue	number	Atom	nuclear	shift
Asp	78	CA	13C	57.293
Leu	79	N	15N	120.714
Leu	79	H	1H	9.004
Leu	79	CB	13C	40.272
Ala	80	N	15N	123.272
Ala	80	H	1H	7.906
Ala	80	CB	13C	17.27
Ala	80	CA	13C	54.291
Thr	81	N	15N	118.284
Thr	81	H	1H	7.973
Thr	81	CB	13C	67.969
Thr	81	CA	13C	66.153
Val	82	N	15N	122.186
Val	82	H	1H	7.59
Val	82	CB	13C	31.032
Val	82	CA	13C	65.958
Ala	83	N	15N	121.807
Ala	83	H	1H	7.886
Ala	83	CB	13C	16.91
Ala	83	CA	13C	55.338
Arg	84	N	15N	115.612
Arg	84	H	1H	7.54
Arg	84	CB	13C	29.976
Arg	84	CA	13C	58.563
Gln	85	N	15N	115.601
Gln	85	H	1H	7.742
Ile	86	N	15N	109.529
Ile	86	H	1H	8.144
Ile	86	CB	13C	40.285
Ile	86	CA	13C	60.601
Asn	87	N	15N	124.611
Asn	87	H	1H	8.879
Asn	87	CB	13C	35.366
Asn	87	CA	13C	56.734
Pro	88	CB	13C	30.852
Pro	88	CA	13C	64.741
Tyr	89	N	15N	112.761
Tyr	89	H	1H	7.942

Table A-1 (continued)

Residue	number	Atom	nuclear	shift
Tyr	89	CB	13C	38.234
Tyr	89	CA	13C	57.057
Ile	90	N	15N	122.655
Ile	90	H	1H	7.77
Ile	90	CA	13C	61.496
Lys	91	N	15N	126.203
Lys	91	H	1H	9.22
Lys	91	CB	13C	32.709
Lys	91	CA	13C	55.069
Val	92	N	15N	111.662
Val	92	H	1H	7.024
Val	92	CB	13C	34.775
Val	92	CA	13C	58.563
Asp	93	N	15N	121.898
Asp	93	H	1H	8.522
Asp	93	CB	13C	40.59
Asp	93	CA	13C	54.131
Ile	94	N	15N	120.592
Ile	94	H	1H	7.878
Ile	94	CB	13C	38.397
Ile	94	CA	13C	56.63
Pro	95	CB	13C	32.168
Pro	95	CA	13C	62.321
Glu	96	N	15N	125.137
Glu	96	H	1H	9.027
Glu	96	CB	13C	29.233
Glu	96	CA	13C	59.642
Thr	97	N	15N	108.236
Thr	97	H	1H	8.111
Thr	97	CB	13C	67.164
Thr	97	CA	13C	63.565
Met	98	N	15N	120.778
Met	98	H	1H	7.084
Met	98	CB	13C	33.582
Met	98	CA	13C	56.502
Arg	99	N	15N	120.298
Arg	99	H	1H	7.607
Arg	99	CB	13C	29.352

Table A-1 (continued)

Residue	number	Atom	nuclear	shift
Arg	99	CA	13C	60.242
Gly	100	N	15N	103.435
Gly	100	H	1H	8.172
Gly	100	CA	13C	47.117
Ala	101	N	15N	124.737
Ala	101	H	1H	7.6
Ala	101	CB	13C	18.409
Ala	101	CA	13C	54.476
Leu	102	N	15N	118.178
Leu	102	H	1H	7.879
Leu	102	CB	13C	41.102
Leu	102	CA	13C	57.266
Tyr	103	N	15N	118.126
Tyr	103	H	1H	8.606
Tyr	103	CB	13C	35.731
Tyr	103	CA	13C	57.646
Ser	104	N	15N	116.541
Ser	104	H	1H	8.086
Ser	104	CB	13C	39.764
Ser	104	CA	13C	63.535
Phe	105	N	15N	123.296
Phe	105	H	1H	8.574
Phe	105	CB	13C	39.142
Phe	105	CA	13C	61.949
Val	106	N	15N	119.991
Val	106	H	1H	8.676
Val	106	CB	13C	29.853
Val	106	CA	13C	66.025
Tyr	107	N	15N	122.173
Tyr	107	H	1H	8.76
Tyr	107	CB	13C	38.245
Tyr	107	CA	13C	58.68
Asn	108	N	15N	112.864
Asn	108	H	1H	6.798
Asn	108	CB	13C	40.774
Asn	108	CA	13C	56.61
Val	109	N	15N	111.309
Val	109	H	1H	7.87

Table A-1 (continued)

Residue	number	Atom	nuclear	shift
Val	109	CB	13C	32.152
Val	109	CA	13C	63.149
Gly	110	N	15N	112.217
Gly	110	H	1H	8.687
Ala	111	N	15N	128.841
Ala	111	H	1H	9.099
Ala	111	CB	13C	17.985
Ala	111	CA	13C	55.677
Gly	112	N	15N	107.315
Gly	112	H	1H	8.884
Gly	112	CA	13C	46.868
Phe	114	N	15N	118.672
Phe	114	H	1H	8.349
Phe	114	CA	13C	61.211
Arg	115	N	15N	119.483
Arg	115	H	1H	8.901
Arg	115	CB	13C	29.883
Arg	115	CA	13C	59.187
Thr	116	N	15N	106.168
Thr	116	H	1H	7.307
Thr	116	CB	13C	68.194
Thr	116	CA	13C	60.357
Ser	117	N	15N	117.323
Ser	117	H	1H	7.556
Ser	117	CB	13C	65.421
Ser	117	CA	13C	58.56
Thr	118	N	15N	120.975
Thr	118	H	1H	8.999
Thr	118	CB	13C	67.374
Leu	119	N	15N	122.877
Leu	119	H	1H	8.535
Leu	119	CB	13C	41.649
Leu	119	CA	13C	59.081
Leu	120	N	15N	118.497
Leu	120	H	1H	7.78
Leu	120	CB	13C	39.286
Leu	120	CA	13C	57.234
Arg	121	N	15N	118.177

Table A-1 (continued)

Residue	number	Atom	nuclear	shift
Arg	121	H	1H	7.102
Arg	121	CB	13C	29.424
Arg	121	CA	13C	59.651
Lys	122	N	15N	116.055
Lys	122	H	1H	8.276
Lys	122	CB	13C	31.176
Lys	122	CA	13C	58.09
Ile	123	N	15N	120.453
Ile	123	H	1H	8.222
Ile	123	CB	13C	37.816
Ile	123	CA	13C	65.758
Asn	124	N	15N	116.613
Asn	124	H	1H	8.473
Asn	124	CB	13C	37.657
Asn	124	CA	13C	54.223
Gln	125	N	15N	118.349
Gln	125	H	1H	7.724
Gln	125	CB	13C	29.288
Gln	125	CA	13C	55.075
Gly	126	N	15N	110.058
Gly	126	H	1H	8.413
Gly	126	CA	13C	45.199
Asp	127	N	15N	119.569
Asp	127	H	1H	7.965
Asp	127	CB	13C	39.193
Asp	127	CA	13C	51.215
Ile	128	N	15N	124.332
Ile	128	H	1H	7.438
Ile	128	CB	13C	36.967
Ile	128	CA	13C	63.934
Lys	129	N	15N	120.966
Lys	129	H	1H	8.404
Lys	129	CB	13C	31.678
Lys	129	CA	13C	59.113
Gly	130	N	15N	109.234
Gly	130	H	1H	8.38
Gly	130	CA	13C	46.865
Ala	131	N	15N	125.903

Table A-1 (continued)

Residue	number	Atom	nuclear	shift
Ala	131	H	1H	8.684
Ala	131	CB	13C	17.745
Ala	131	CA	13C	55.214
Cys	132	N	15N	116.355
Cys	132	H	1H	8.066
Cys	132	CB	13C	39.808
Cys	132	CA	13C	58.375
Asp	133	N	15N	118.279
Asp	133	H	1H	8.021
Asp	133	CB	13C	40.994
Asp	133	CA	13C	57.394
Gln	134	N	15N	115.345
Gln	134	H	1H	8.152
Gln	134	CB	13C	27.913
Leu	135	N	15N	121.517
Leu	135	H	1H	8.428
Leu	135	CB	13C	41.307
Leu	135	CA	13C	57.589
Arg	136	N	15N	118.706
Arg	136	H	1H	8.576
Arg	136	CB	13C	29.874
Arg	136	CA	13C	59.41
Arg	137	N	15N	115.5
Arg	137	H	1H	7.541
Arg	137	CA	13C	55.545
Trp	138	N	15N	124.287
Trp	138	H	1H	7.591
Trp	138	CB	13C	28.579
Trp	138	CA	13C	58.644
Thr	139	N	15N	122.725
Thr	139	H	1H	7.586
Thr	139	CB	13C	68.579
Thr	139	CA	13C	61.16
Tyr	140	N	15N	127.392
Tyr	140	H	1H	8.757
Tyr	140	CA	13C	58.176
Ala	141	N	15N	127.3
Ala	141	H	1H	8.782

Table A-1 (continued)

Residue	number	Atom	nuclear	shift
Ala	141	CB	13C	18.989
Ala	141	CA	13C	52.211
Gly	142	N	15N	109.265
Gly	142	H	1H	8.506
Gly	142	CA	13C	45.349
Gly	143	N	15N	109.299
Gly	143	H	1H	8.521
Lys	144	N	15N	119.72
Lys	144	H	1H	7.788
Lys	144	CB	13C	33.253
Lys	144	CA	13C	55.172
Gln	145	N	15N	122.903
Gln	145	H	1H	8.584
Gln	145	CB	13C	29.325
Gln	145	CA	13C	59.081
Trp	146	N	15N	127.776
Trp	146	H	1H	8.914
Trp	146	CA	13C	57.492
Lys	147	N	15N	127.523
Lys	147	H	1H	8.982
Lys	147	CB	13C	35.567
Lys	147	CA	13C	54.683
Gly	148	N	15N	107.196
Gly	148	H	1H	8.702
Gly	148	CA	13C	43.81
Leu	149	N	15N	119.327
Leu	149	H	1H	8.861
Leu	149	CA	13C	54.612
Met	150	N	15N	119.89
Met	150	H	1H	8.296
Met	150	CB	13C	34.727
Met	150	CA	13C	55.102
Thr	151	N	15N	117.094
Thr	151	H	1H	9.066
Thr	151	CB	13C	70.254
Thr	151	CA	13C	60.818
Arg	152	N	15N	121.016
Arg	152	H	1H	8.277

Table A-1 (continued)

Residue	number	Atom	nuclear	shift
Arg	152	CB	13C	28.859
Arg	152	CA	13C	59.673
Arg	153	N	15N	117.009
Arg	153	H	1H	8.046
Arg	153	CB	13C	30.767
Arg	153	CA	13C	58.634
Glu	154	N	15N	121.038
Glu	154	H	1H	7.76
Ile	155	N	15N	121.103
Ile	155	H	1H	8.12
Ile	155	CB	13C	38.397
Ile	155	CA	13C	65.592
Glu	156	N	15N	118.069
Glu	156	H	1H	8.319
Glu	156	CB	13C	29.382
Glu	156	CA	13C	58.981
Arg	157	N	15N	118.087
Arg	157	H	1H	7.959
Arg	157	CB	13C	29.737
Ile	159	N	15N	118.131
Ile	159	H	1H	8.105
Ile	159	CB	13C	36.212
Ile	159	CA	13C	64.11
Cys	160	N	15N	115.696
Cys	160	H	1H	8.464
Cys	160	CB	13C	47.2
Cys	160	CA	13C	62.097
Leu	161	N	15N	118.585
Leu	161	H	1H	7.693
Leu	161	CB	13C	41.34
Trp	162	N	15N	125.184
Trp	162	H	1H	7.916
Trp	162	CB	13C	29.287
Trp	162	CA	13C	57.147
Gly	163	N	15N	106.744
Gly	163	H	1H	8.694
Gly	163	CA	13C	45.188
Gln	164	N	15N	119.379

Table A-1 (continued)

Residue	number	Atom	nuclear	shift
Gln	164	H	1H	7.839
Gln	164	CB	13C	29.233
Gln	164	CA	13C	55.111
Gln	165	N	15N	121.921
Gln	165	H	1H	8.444
Gln	165	CB	13C	28.627
Gln	165	CA	13C	53.194

APPENDIX B

SUPPLEMENTAL MATERIAL FOR CHAPTER III

Table B-1. Oligonucleotides

Name	Sequence	Use
LJM05	TTTTTTCATATGAGCAAGCCCTGGCTGTTCAC	D29*gene12*[CP] pLAM3
LJM06	TTTTTTAAGCTTGATCTGTCGTAGGAACTCGACCGC	D29*gene12*[CP] pLAM3
D29gp12F	TGGTTTTTCATATGAGCAAGCCCTGGCTG	D29*gene12*[CP] pKMC2
D29gp12R	GTTGGGTGTGCTAGCTCAGATCTGTCGTA	D29*gene12*[CP] pKMC2
D29B-S82AF	GATGGCGGGTTACGCGCAGGGAGCCATCG	D29BS82A site-dir. mutagenesis pKC20
D29B-S82AR	CGATGGCTCCCTGCGCGTAACCCGCCATC	D29BS82A site-dir. mutagenesis pKC20
D29B-NdeIF	TGTTTTGTCATATGAGCAAGCCCTGGCTGTTCAC	D29*gene12*[CP] pKMC9
D29B-HindIIIR	TGGTGTAAGCTTTCAGATCTGTCGTAGGAACTCGAC	D29*gene12*[CP] pKMC9
GilesB-DO2	CTCTGATGGCGTGGAAACCTACCGAATACCAAATCGGAGATC GGCACCCCCACGTCAGAGACTGGGCCGGGCGCACCCCGGCCG TCACCGCATAACACAC	Giles*gene*32 1146 bpΔ
GilesB-EP1.2	ATGTCGACGGCATCGTGGGTCCGCTCACGGCCGCGAAGATTG GGTTGGTGCTCTGATGGCGTGGAAACCTACCGA	Extend GilesB-DO2
GilesB-EP2.2	TGGCCGGGTTCGGATTGGGCGGATTCGTGCTCACTGTTGCGAT CCCTTCGGTGTGTTATGCGGTGACGGCCGGGG	Extend GilesB-DO2
GilesB-DiagF	GAACGTCGTGAAGCTGCAAACGACGTTG	Giles *gene*32 flanking primer
GilesB-DiagR3	CGTCACCTCGTCCTCACGTTTCCGTTTG	Giles *gene*32 flanking primer
GilesB-MAMAPCR3	CGAATACCAAATCGGAGATCGGCACCCAC	Giles *gene*32 DADA-PCR primer

Note: [CP]Cloning Primer

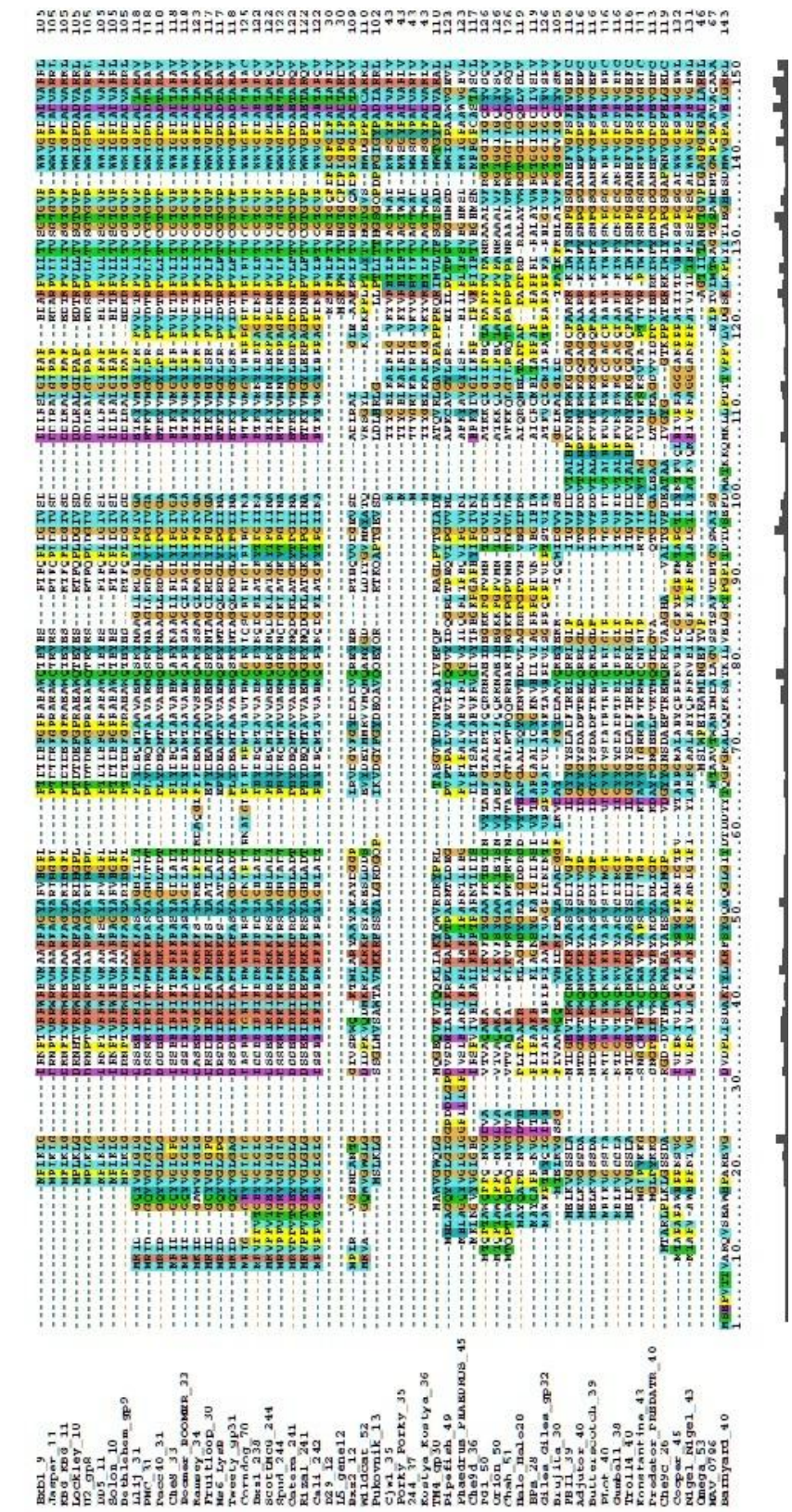


Figure B-1. Alignment of LysB protein using Clustal X.

The LysB proteins from 60 sequenced mycobacteriophage genomes, mycobacteriophage Ms6, and a putative prophage of *M. avium* were aligned using ClustalX. Absolutely conserved residues are indicated by asterisks above the alignment. Periods and colons above the alignment indicate partly conserved residues.

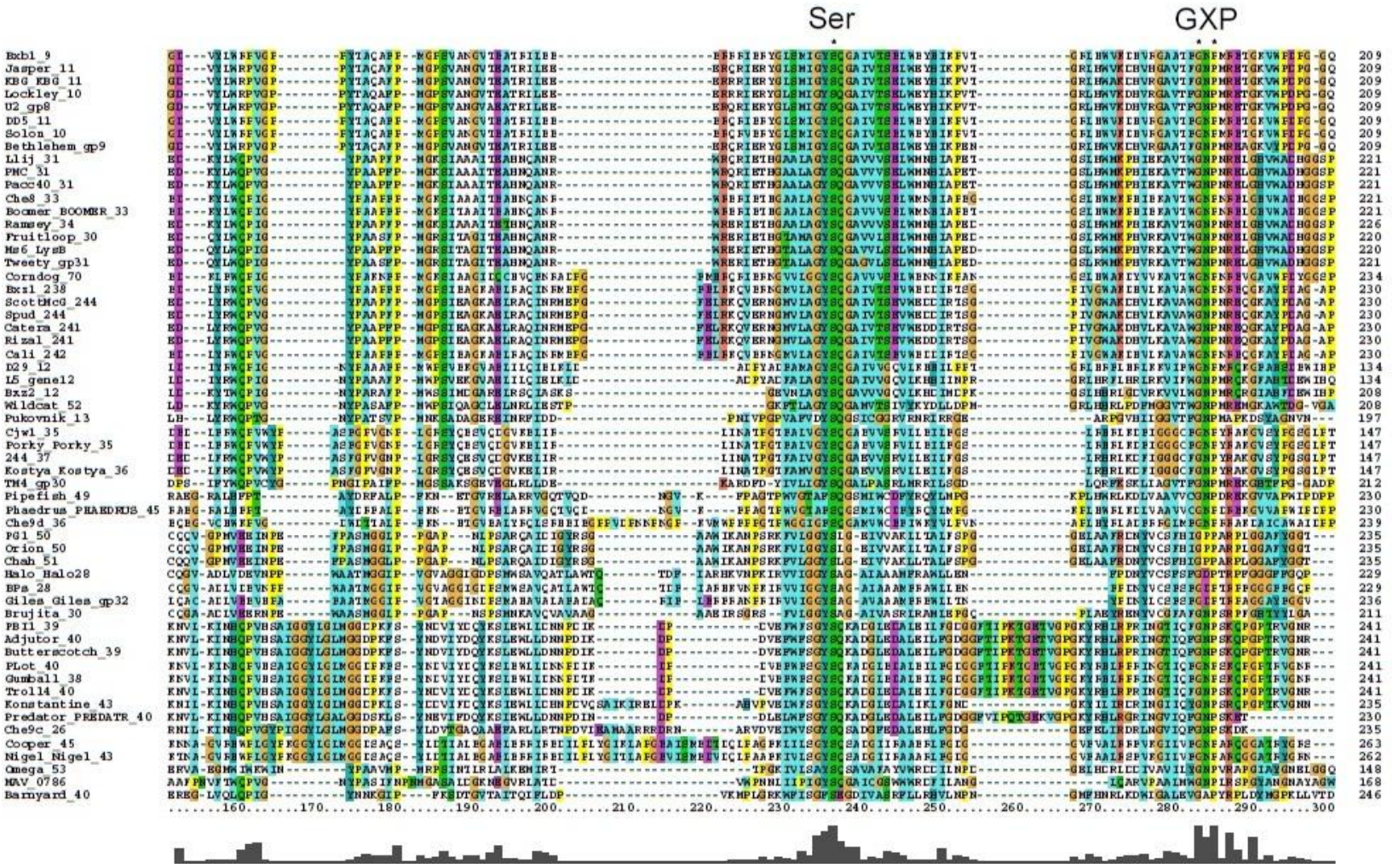


Figure B-1 (Continued)

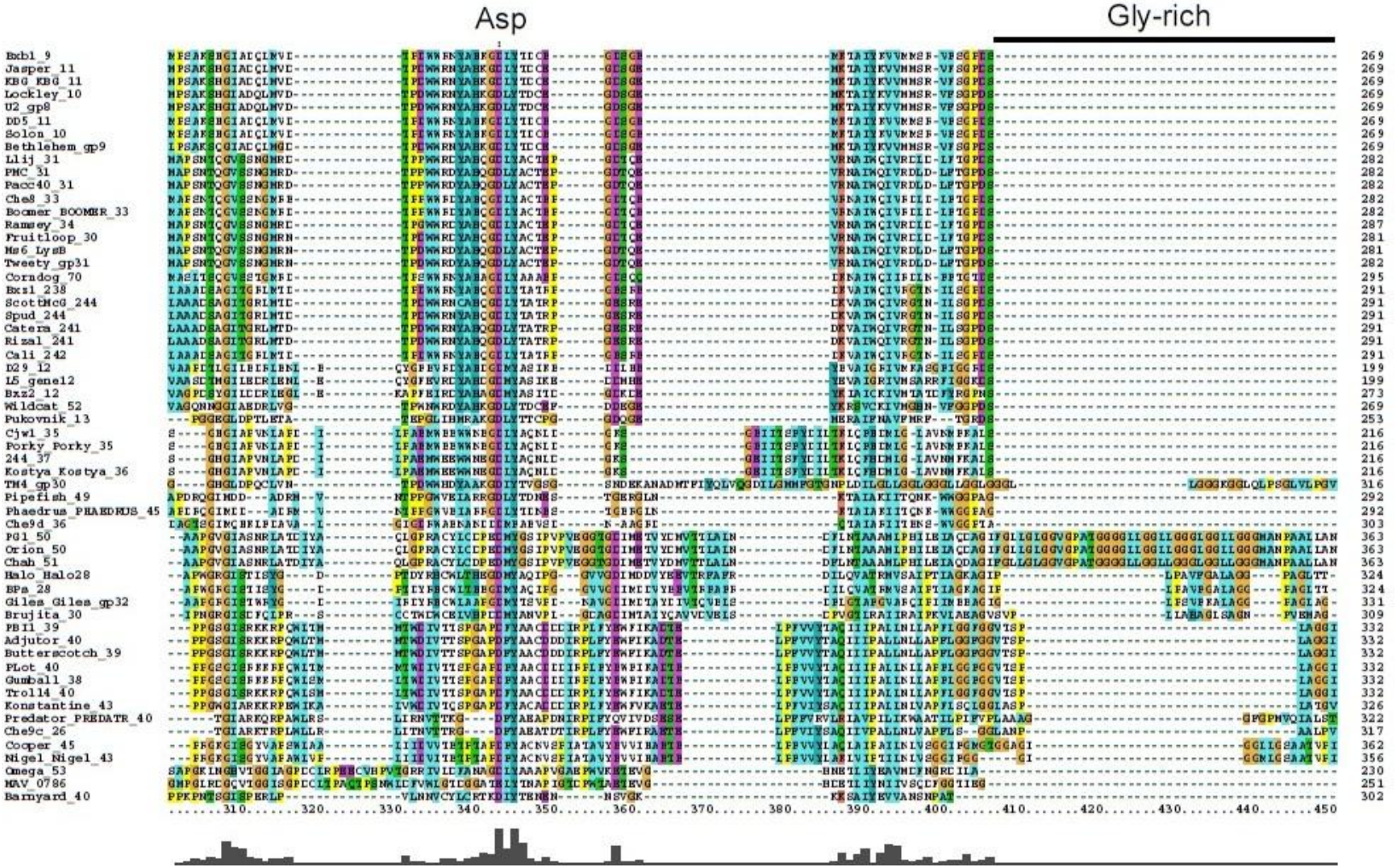


Figure B-1 (Continued)

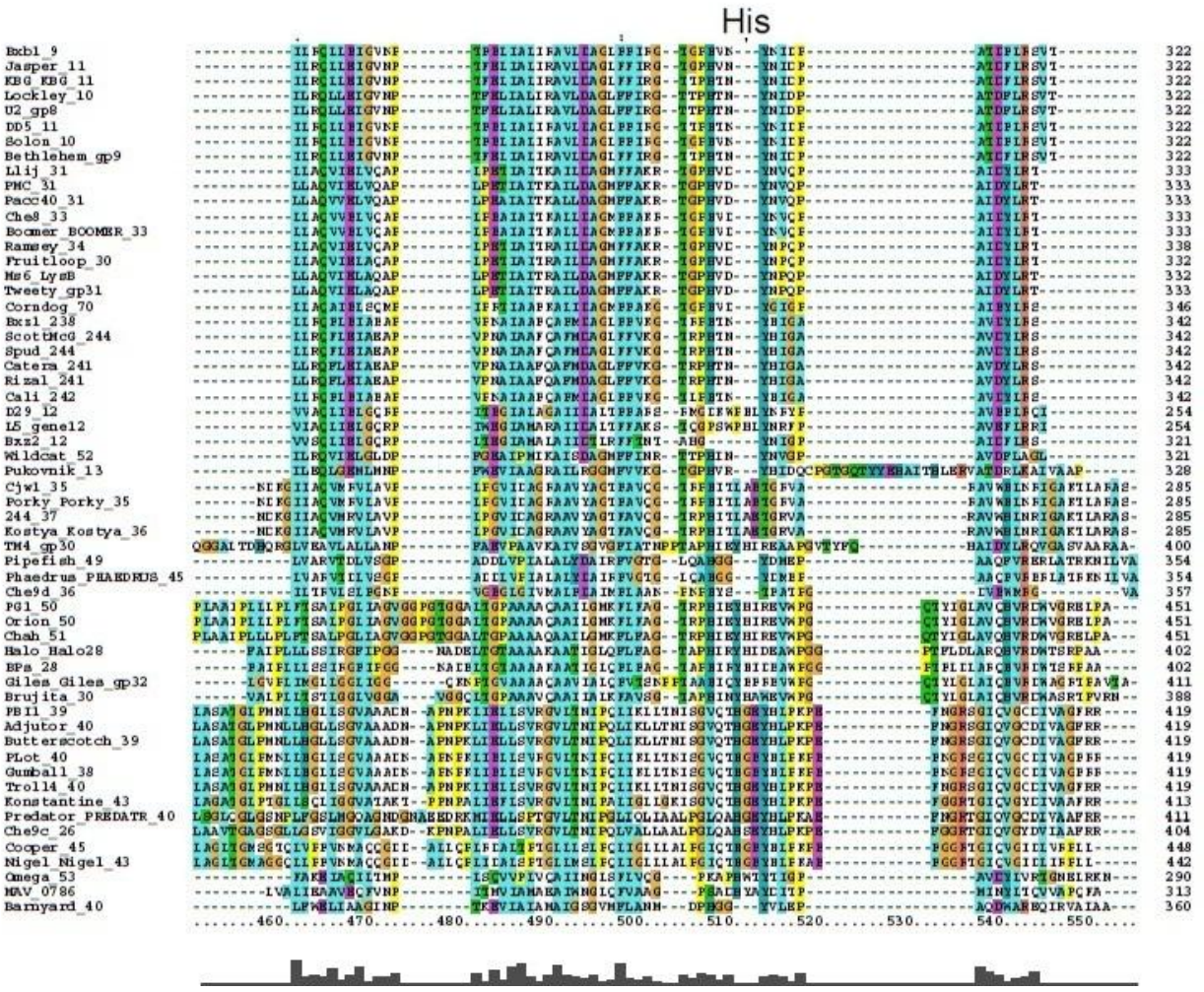


Figure B-1 (Continued)

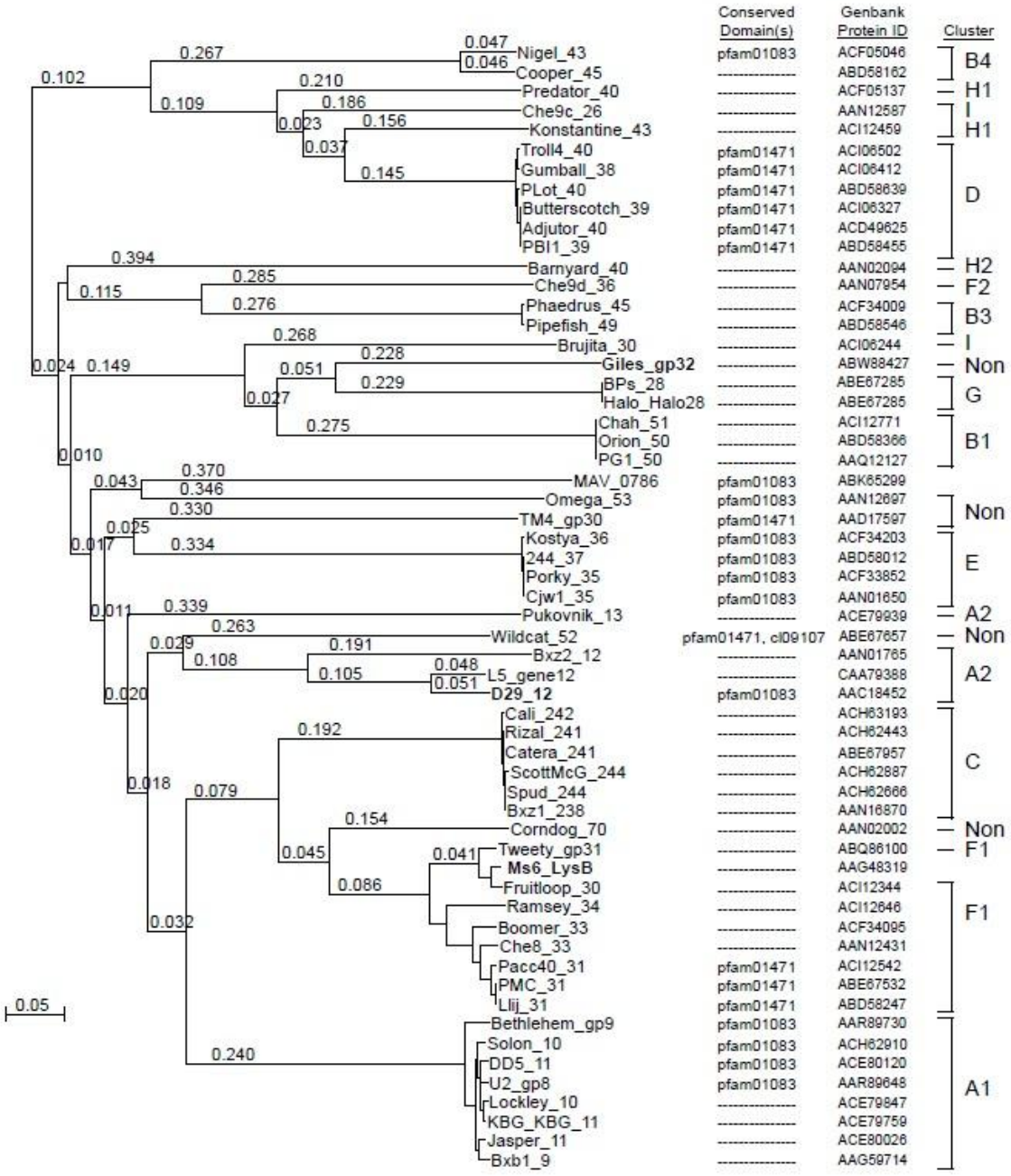


Figure B-2. Phylogenetic relationships of LysB proteins. The LysB proteins of sixty completely sequenced mycobacteriophages, mycobacteriophage Ms6, and of a putative *M. avium* prophage (MAV-0786) were aligned in CLustal X and displayed using Njplot. Conserved domains identifiable in each LysB protein are shown to the right: pfam01083, Cutinase superfamily; pfam01471, putative peptidoglycan binding domain; cl09107, Esterase_lipase superfamily. The Genbank protein ID numbers are shown on the right. The three LysB proteins that have been examined experimentally are shown in bold.

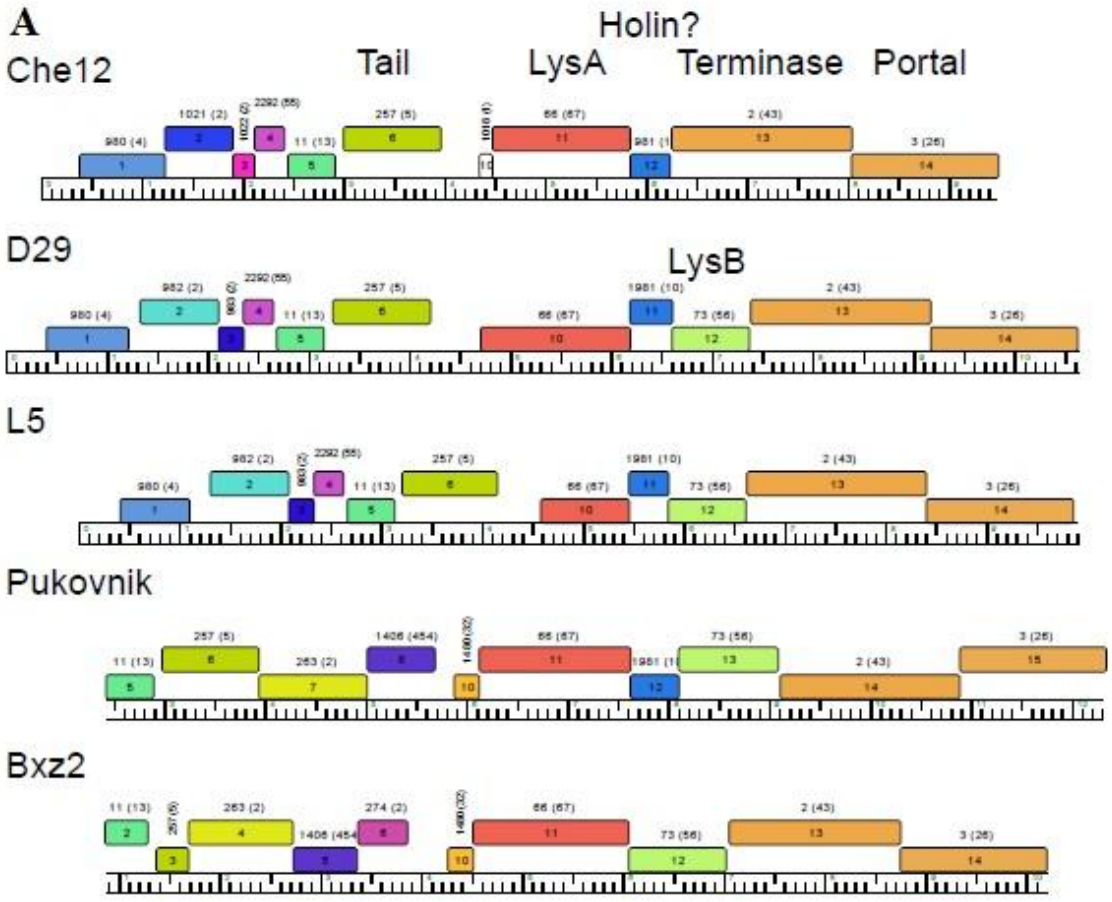


Figure B-3. Genetic organization of mycobacteriophages lacking *lysB*.

A. The lysis cassettes and surrounding genes of the five cluster A2 genomes are shown. All contain a *lysA* gene, but Che12 lacks *lysB*, seemingly through simple loss of the genes. Note that the putative holin is also apparently missing in Bxz2. The pham number with the number of phamily members in parentheses is shown above each gene.

B. The lysis cassettes of five cluster B genomes are shown; PG1 belongs to cluster B1, Qyrzula and Rosebush to cluster B2, Phaedrus to cluster B3 and Cooper to cluster B4. Each genome contains a *lysA* gene, but both Qyrzula and Rosebush lack *lysB*. While both these genomes contain a different gene in the same location (Qyrzula gene *45* and Rosebush gene *47*), these genes constitute Pham1989, which PSI-Blast analysis suggests is distantly related to the Pham 10 group of holins shown in **Fig. A-3A**. The location of putative holin genes in PG1, Phaedrus and Cooper is not known.

C. The lysis cassettes of four phages of cluster C are shown; Rizal, ScottMcG and Spud all belong to subcluster C1, while Myrna belongs to subcluster C2. Each genome contains a *lysA* gene, but Myrna lacks *lysB*. The adjacent gene (*244*) is a candidate for involvement in lysis, but it is not related to the large family of cutinase-like proteins or other esterases. Myrna gene *245* is of unknown function.

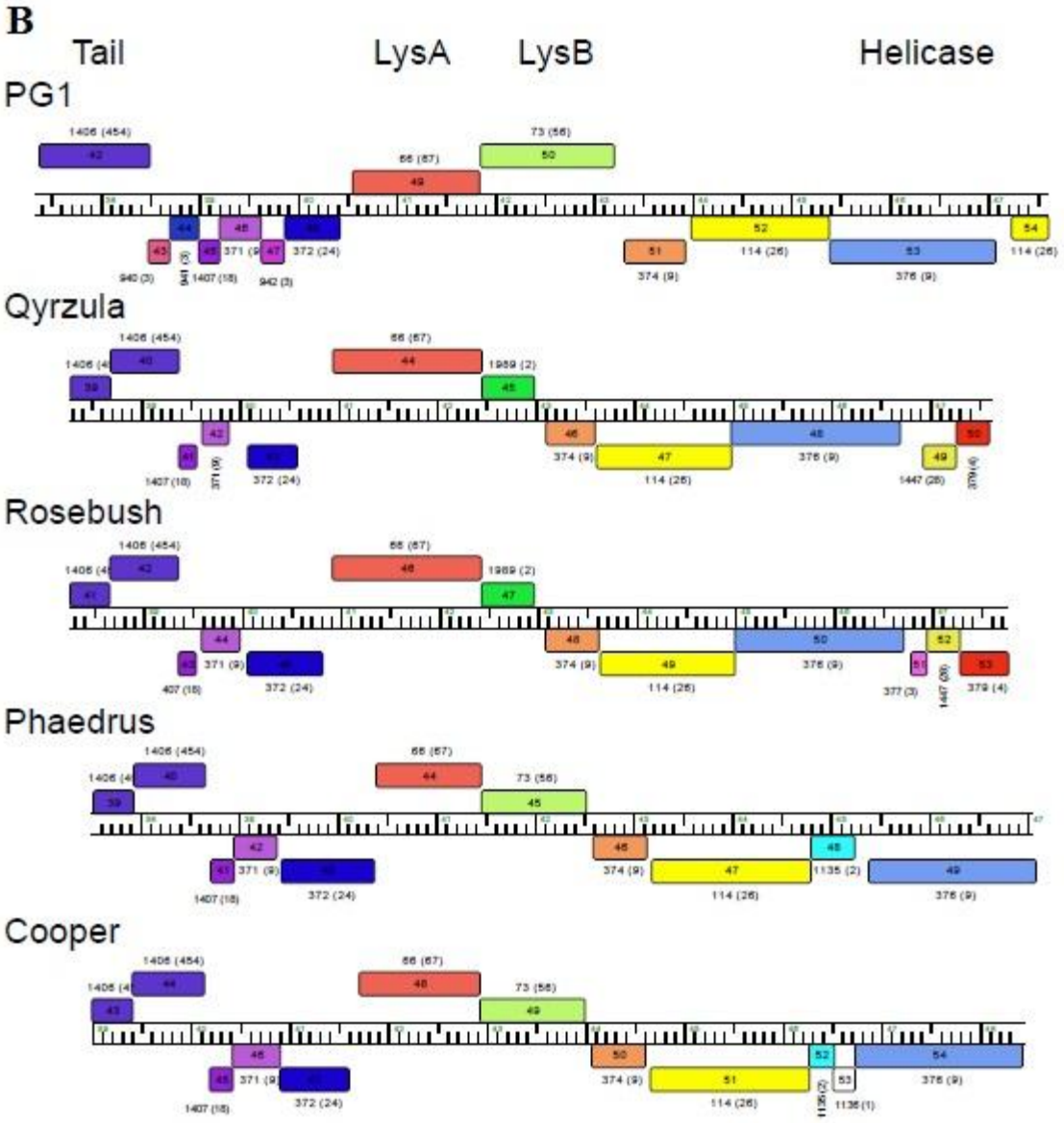


Figure B-3 Continued

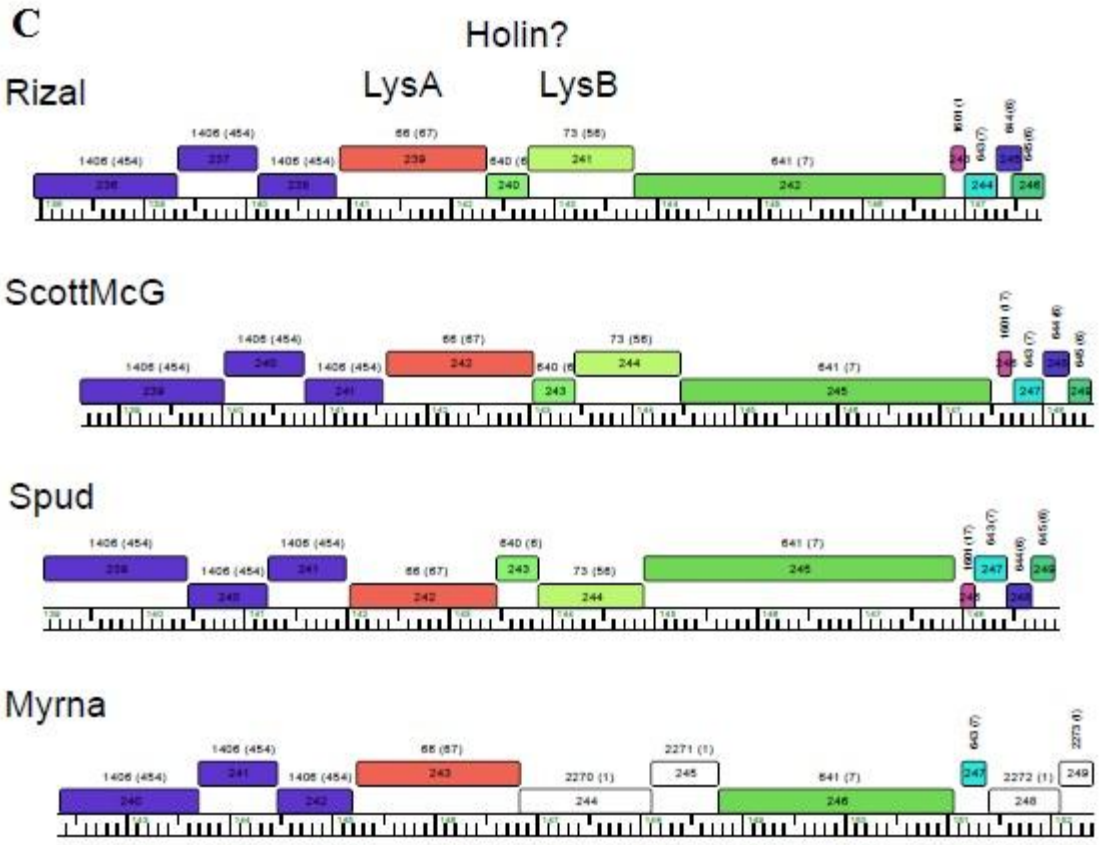


Figure B-3 Continued

APPENDIX C

SUPPLEMENTAL MATERIAL FOR CHAPTER IV

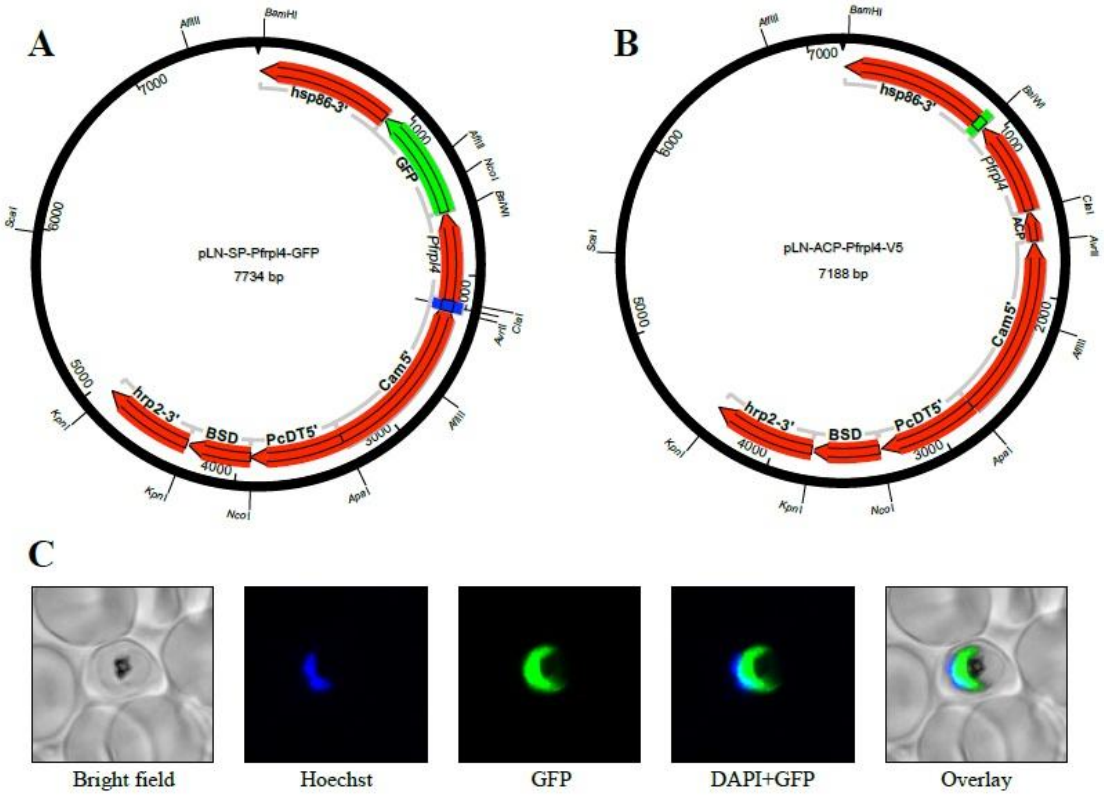


Figure C-1. Plasmid maps and live imaging of Pfrpl4-V5 transfected parasites. A. map of pLN-SP-Pfrpl4-GFP plasmid. Transfection of this plasmid into *P.falciparum* (Dd2 strain) resulted in cytosolic expression of the PfRPL4-GFP protein (see panel C). **B**. map of pLN-ACP-Pfrpl4-V5 plasmid. Parasites transfected with either this plasmid or pLN-ACP-Pfrpl4-GFP (not shown) revealed no protein expression by Western blot (data not shown). **C**. Live imagingof Dd2 parasites transfected with pLN-SP-Pfrpl4-GFP, showing cytosolic expression of the PfRPL4-GFP fusion with a N-terminal signal peptide sequence from the ACP protein. DNA was stained using DAPI. Images were taken on an Olympus IX81 inverted microscope at 90× magnification, as described (Nkrumah *et al.* (2006) *Nature Methods* 3, 615-621).

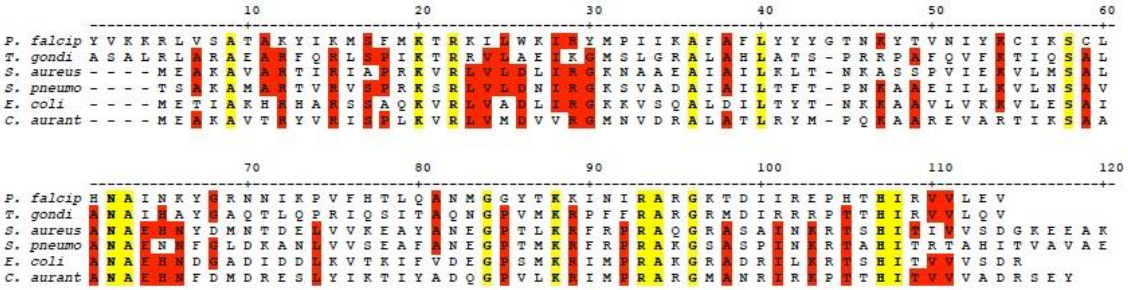


Figure C-2. Amino acid sequence alignment of the ribosomal protein L22. The aligned sequences are from *Plasmodium falciparum* (P. falcip), *Taxoplasma gondii* (T. gondi), *Staphylococcus aureus* (S. aureus), *Streptococcus pneumoniae* (S. pneumo), *Escherichia coli* (E. coli) and *Chloroflexus aurantiacus* (C. aurant, an aerobic facultative bacterium that can photosynthesize in anaerobic situations). The long amino terminal extension of this protein in *P. falciparum* and *T. gondii* is not shown. Conserved residues are indicated in yellow shade, whereas the red shade shows the homology between these protein sequences.

Table C-1. Antimalarial IC_{50} values in 48 h and 96 h drug assays*

	Dd2		7G8		GCO3	
	48 h	96 h	48 h	96 h	48 h	96 h
Azithromycin (nM)						
Mean	3464	103.4	15742	561.2	22585	43.3
SEM	1603	18.0	1527	203.6	2116	1.0
Thiostrepton (nM)						
Mean	2107.0	487.5	688.5	924.6	2277.0	916.5
SEM	1206.0	76.7	160.2	148.1	293.8	63.9
Triclosan (nM)						
Mean	6258.0	4186.0	773.9	424.4	2708.0	1526.0
SEM	270.4	408.7	129.6	24.5	21.6	56.9
Chloroquine (nM)						
Mean	96.8	64.2	92.3	41.6	4.9	4.8
SEM	7.4	14.5	7.2	10.7	0.1	0.0

*Assays were performed in duplicate on six to eight separate occasions

Table C-2. Functional PfCRT and PfMDR1 haplotypes of recombinant lines and parents*

	PfCRT								PfMDR1					Ref.	AZ IC_{50}
Clones	72	74	75	76	220	271	326	356	86	184	1034	1042	1246		
Dd2	C	I	E	T	S	E	S	T	N	F	S	D	D	1	103.4±18.5
C-1^{Dd2}	C	I	E	T	S	E	S	T	N	F	S	D	D	1	95.7±14.7
T76K-1^{Dd2}	C	I	E	K	S	E	S	T	N	F	S	D	D	1	121.8±12.5
7G8	S	M	N	T	S	Q	D	L	N	F	C	D	Y	1	190.2±20.6
C-1^{7G8}	S	M	N	T	S	Q	D	L	N	F	C	D	Y	1	141.8±10.8
T76K-1^{7G8}	S	M	N	K	S	Q	D	L	N	F	C	D	Y	1	114.8±24.5
GC03	C	M	N	K	A	Q	N	I	N	F	S	D	D	2	77.1±11.5
$C2^{GC03}$	C	M	N	K	A	Q	N	I	N	F	S	D	D	2	48.0±6.5
$C4^{Dd2}$	C	I	E	T	S	E	S	T	N	F	S	D	D	2	66.7±5.4
$C6^{7G8}$	S	M	N	T	S	Q	D	L	N	F	S	D	D	2	76.7±9.1
SDD^{GC03}	C	M	N	K	A	Q	N	I	N	F	S	D	D	3	63.3±7.3
SND^{GC03}	C	M	N	K	A	Q	N	I	N	F	S	N	D	3	47.8±4.6
CDY^{GC03}	C	M	N	K	A	Q	N	I	N	F	C	D	Y	3	70.2±11.8
3BA6	C	I	E	T	S	E	S	T	N	F	S	D	D	3	180.3±42.5
SDD^{3BA6}	C	I	E	T	S	E	S	T	N	F	S	D	D	3	154.7±52.9
SND^{3BA6}	C	I	E	T	S	E	S	T	N	F	S	N	D	3	157.9±37.5
CDY^{3BA6}	C	I	E	T	S	E	S	T	N	F	C	D	Y	3	194.2±51.9

*IC_{50} values (Mean±SEM, nM) of these parasite lines for AZ are also indicated. C-1^{Dd2}, C-1^{7G8}, $C2^{GCO3}$, SDD^{GCO3} and SDD^{3BA6} are control lines. 1. Lakshmanan *et al*. (2005) EMBO J. 24, 2294-305; 2. Sidhu *et al*. (2002) Science 298, 210-13;3. Sidhu *et al.* (2005) Mol. Microbiol. 57, 913-26.

Table C-3. Antimalarial response of AZ-resistant lines*

	Dd2	AZI-R^{Dd2}	7G8	AZI-R^{7G8}
Azithromycin (nM)				
Mean	124.1	1938.7	228.1	3971.7
SEM	22.2	454.0	24.7	500.1
Erythromycin (μM)				
Mean	9.4	34.7	20.3	39.5
SEM	2.0	5.9	2.9	4.1
Tetracycline (nM)				
Mean	397.1	323.9	696.7	364.7
SEM	50.3	53.2	108.5	67.9
Doxycycline (nM)				
Mean	432.2	302.5	1444.1	603.1
SEM	80.4	43.6	267.2	220.8
Thiostrepton (nM)				
Mean	292.5	174.7	554.8	217.4
SEM	46.0	39.5	88.9	67.4
Chloroquine (nM)				
Mean	138.6	108.5	132.2	89.6
SEM	7.8	8.4	21.9	8.3

*IC_{50} values (mean±SEM) were calculated from 96 hr assays performed on 4 to 6 separate occasions in duplicate.

VITA

Name:	Qingan Sun
Address:	Department of Biochemistry and Biophysics Texas A&M University 2128 TAMU College Station, TX 77843-2128 c/o James C. Sacchettini
Email Address:	quinsun@gmail.com

Education:	B.S., Applied Physics, Fudan University, Shanghai, 1997
	M.S., Biophysics, Fudan University, 2000
	Ph.D., Biochemistry, Texas A&M University, 2011
Employment:	Shanghai Scientific and Technical Publishers, Shanghai, 2000-2003
Award:	2011 John Mack Prescott Award