RICE UNIVERSITY

Experimental evolution of TetX2: correlating changes in fitness to their structural and functional origins

by

**Katarzyna Walkiewicz**

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
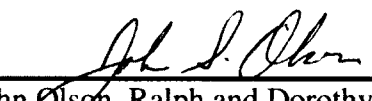
**Doctor of Philosophy**

APPROVED, THESIS COMMITTEE

Yousif Shamoo, Associate Professor of Department
of Biochemistry and Cell Biology, Director

Yizhi Jane Tao, Associate Professor of Department
of Biochemistry and Cell Biology

Edward Nikonowicz, Associate Professor of
Department of Biochemistry and Cell Biology

John Olson, Ralph and Dorothy Looney Professor
of Biochemistry and Cell Biology

Michael R. Diehl, Assistant Professor of
Bioengineering, Assistant Professor of Chemistry

HOUSTON, TEXAS
May 2012

UMI Number: 3521304

# UMI®

Dissertation Publishing

# ProQuest®

# Abstract

## Experimental evolution of TetX2: correlating changes in fitness to their structural and functional origins

by

Katarzyna Walkiewicz

The study of protein evolution and adaptation resides at the junction between the disciplines of biological chemistry and evolutionary biology. We chose the *B. thetaiotaomicron* tetracycline resistant enzyme TetX2, as our model system to study the biophysical basis for adaptation to antibiotics; a phenomenon that continuously poses global health challenges. In the work presented here, experimental evolution and biophysical characterization were used to identify and link the physicochemical properties of TetX2 and its adaptive mutants to increased resistance to minocycline.

*Bacteroides thetaiotaomicron* TetX2 was previously identified as a novel oxidoreductase with broad activity against tetracyclines. Experimental evolution of *E. coli* expressing a chromosomal copy of *tet(X2)* was used to identify an adaptive mutation (TetX2$_{T280A}$) that confers higher resistance to minocycline and tigecycline. In addition to TetX2$_{T280A}$, a family of variants of TetX2 with single amino acid changes in TetX2 sequence that conferred equal or higher resistance towards MCN was identified by error-prone mutagenesis. Changes in fitness of *E. coli* carrying a single chromosomal copy of either wild-type or one of the mutant alleles were assessed by growth rate assays over a

range of minocycline concentrations. Despite similar *in vivo* performances of TetX2$_{T280A}$ and two other variants (TetX2$_{N371I}$ and TetX2$_{N371T}$), TetX2$_{T280A}$ was the only successful mutant in the adaption experiment suggesting that mutational supply may play an important role in evolutionary dynamics of populations undergoing adaptation. The most surprising result is that the differences in growth rates among TetX2 variants arise from small changes in *in vitro* catalytic activity and *in vivo* protein expression.

The steady-state kinetic studies with minocycline and NADPH suggest a binary mechanism for antibiotic inactivation by TetX2 which is supported by the structural characteristics of the enzyme. The atomic structures of the best adaptive mutant TetX2$_{T280A}$ in complex with minocycline and tigecycline reveal the details of substrate recognition and show that the site of the mutation is ~18 Å away from the active site suggesting an indirect mechanism for improved catalysis.

Taken together, our data show that very small changes in the *in vitro* biochemical properties and expression levels can have surprisingly large fitness effects and are important during adaption. In addition, a promising preliminary mathematical model suggests that based on kinetic activity and *in vivo* expression levels the success of bacteria undergoing adaptation to antibiotics can be predicted.

# Acknowledgments

The work presented here would not have been possible without contributions from many people. Foremost, I would like to thank my scientific advisor Dr. Yousif Shamoo for his continuous support during my Ph.D studies, his patience, motivation and encouragement. His wide knowledge and passion for science have helped me continue my work over the years and grow as a scientist that I am today.

I am deeply grateful to my committee members Dr. Tao, Dr. Nikonowicz, Dr. Olson and Dr. Diehl for discussing experiments, making valuable suggestions and for always being there to listen and give advice. Especially, I would like to express my gratitude to Dr. Olson, for his stimulating suggestions and encouragement; the enzyme kinetics work presented here would not have been possible without his expertise.

I am also indebted to the members of Shamoo laboratory, for their practical advice and support. I want thank Matt Pena, who has been a great friend with whom I started this journey. He has always supported me and in many ways inspired me, by sharking his ideas throughout all these years. I want to thank Andres Benitez for taking on the modeling project that added so much value to my research. I want to say thank you to Gerda Saxer, who has shared her knowledge and expertise and supported me during the thesis writing process. Thank you to Christine Sun and Kelsey Lau, two wonderful undergraduate students; their work and commitment have been indispensible for the completion of this project. I also want thank Milya Davlieva and Jay Nix for their help

with the crystallographic studies. I feel very fortunate to have the opportunity to work with Megan Guelker, Corwin Miller, Anisha Perez and all other Shamoo lab members who I did not mention here, but who have in many ways contributed to this work and made the lab a fun place to work at.

In addition, I want to thank Agustina Rodriguez-Granillo, Jayashree Soman, Andria Denmon, Aaron Colier for being great friends and collaborators.

I am incredibly grateful to my family; my parents, Teresa and Stanislaw Walkiewicz and my siblings Magda (and family), Ania (and family) and Kuba whose love and support have given me the strength to continue this journey. Especially, I want to thank Magda for taking such a good care of me and always believing in me especially in the most difficult moments.

Last but not least, I want to express my love and gratitude to my fiancée, Anand Radhakrishnan; he has been a source of courage, inspiration and love for which I am deeply thankful.

# Table of Contents

# List of Figures

# List of Tables

# List of Equations

Eq.4.1
$$v_o = \frac{V_{max}(obs)[A]}{[A] + K_M(obs)}$$

Eq.4.2
$$V_{max}(obs) = \frac{V_{max}[B]}{K_A + [B]}$$

Eq.4.3
$$K_M(obs) = \frac{K_A[B]}{K_B + [B]}$$

# Chapter 1 Introduction

## 1.1. Mutations fuel adaptation

Charles Darwin was the first to recognize that natural selection constitutes the basic mechanism for evolution. In his publication *"On the Origin of Species"* Darwin proposed that new functions arise by the accumulation of adaptive changes resulting in greater fitness over long periods of time (Darwin, 1859). Although controversial at the time, his work became the foundation for evolutionary biology. Spontaneous mutations are one of the most important sources of variation, but the rates at which they occur are very slow (~ $8.9 \times 10^{-11}$ per nucleotide per generation for *E. coli*) (Wielgoss et al., 2011). They arise as a result of aberrations in DNA recombination, replication and/or repair machinery and are constrained by the chemical properties of DNA. For example, transitions (A→G or T→C) are more common than transversions (A→C, A→T, T→G, G→C) , even though there is a higher probability for the latter to occur (Li, 1997; Wielgoss et al., 2011). More recently, it has been suggested that selective pressure increases the rate of mutations and allows observing the differences in fitness resulting from genetic change in a population (Rosenberg and Hastings, 2003).

## 1.2. Adaptive walk to greater fitness along mutational landscapes as a framework for visualizing adaptation

Adaptation happens through accumulation of mutations that lead to higher fitness, as determined by the reproductive success in a selective environment (Elena and Lenski, 2003). Mayanard Smith used an adaptive landscape as a metaphor to describe this process

(Smith, 1970). Smith suggested that mutational pathways exist because changes in fitness occur in a step-like process through functional intermediates that can be imagined as a "walk" across a landscape to increased fitness (**Figure 1.1**). The fate of mutations during adaption depends on their effects on the fitness of the organism. In the simple case of a two-peak fitness landscape, under selection, intermediates (beneficial mutations) must increase in frequency (rise to fixation) in order to have a major influence on a population (Poelwijk, 2007).The adaptive walk continues through subsequent mutations that arise in the background of the previous successful mutant (**Figure 1.1**). The degeneracy of genetic code plays an important role in defining mutational diversity because some amino acid substitutions cannot be achieved through a single nucleotide change (e. g. Val (GUU) cannot be mutated to Ser (AGU) in a single step).

Since Smith's original work, others have further confirmed this theory by studies showing that adaptation occurs over long-periods of time through a punctate process (**Figure 1.1**) (Elena et al., 1996; Gerrish and Lenski, 1998). Consequently, the smaller the fitness benefit conferred by a mutation, the longer it will take for the mutation to fix in a population due to clonal interference (Gerrish and Lenski, 1998). Most beneficial mutations are rare (Orr, 2003) and their success is dependent on several factors such as mutational rate, population size and the distribution of the fitness effects of other beneficial mutations present in a population (Arjan et al., 1999; Lang et al., 2011).
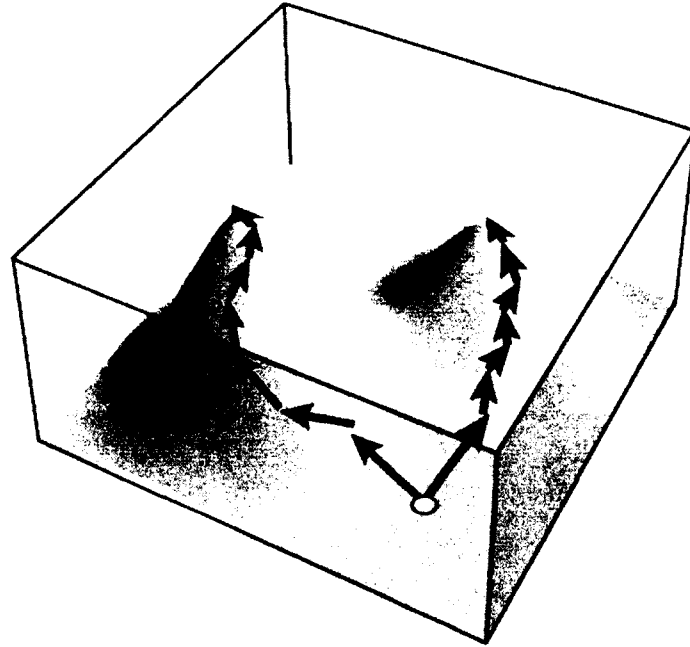
**Figure 1.1 Fitness landscape.**
A progression along the fitness landscape through two independent pathways (red or green) for identical populations adapted independently. Both populations start at low fitness and climbed towards an ultimate fitness peak via a step like process indicated by the dashed arrows where each dash is a step that overall leads to increased fitness. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Genetics (Elena and Lenski, 2003), copyright (2003).

The challenge in relating readily measured physicochemical properties to fitness is difficult because of the complex metabolic context of organisms. In addition, the selection conditions that give rise to mutations are usually not known. In order to overcome these challenges, scientists have turned to experimental evolution that allows a careful examination of the evolutionary process in the laboratory.

## 1.3. Experimental evolution

Stephen J. Gould asked if "replaying life's tape" would always lead to the same result or in other words is evolution repeatable and what role does adaptation, chance and history play in this process (Gould, 1989). Richard Lenski and colleagues addressed the repeatability of evolution in a series of studies using long-term evolution experiments where 12 populations of *Escherichia coli* have been continuously adapted on glucose as limiting source for over 50,000 generations (Barrick et al., 2009; Cooper and Lenski, 2000; Elena et al., 1996; Lenski and Travisano, 1994; Travisano et al., 1995). The results of this experiment have shown that all 12 populations achieved similar fitness, which provided evidence for reproducibility of evolutionary events. As expected, history and chance played an important role on the evolution of traits that only weakly correlated with fitness, such as cell size (Travisano et al., 1995). In addition, Lenski and Travisano (1994) observed that the 12 *E. coli* populations, after 2,000 generations, underwent significant increases in fitness that were drastically decreased between generations 5,000 to 10,000. Together with subsequent studies they showed that initial gains in fitness are large and tend to level off with the increasing number of generations (Barrick et al., 2009; Betancourt and Bollback, 2006; Cooper and Lenski, 2000; Lenski and Travisano, 1994; Novella et al., 1995).

Lenski's long-term evolution experiment has shown that experimental evolution on tractable organisms can be a powerful tool in understanding the principles of adaptation through natural selection. Microbial (Barrick et al., 2009; Cooper and Lenski, 2000; Travisano, 1997) and viral (Bull and Wichman, 1998; Novella et al., 1995) populations are the most common model systems for evolution studies. Among many

advantages of this approach is the fact that experiments can be carried out for many generations under good control of the experimental conditions used during selection (Elena and Lenski, 2003). In addition, the intermediates of adaptation can be retained throughout experimental evolution and their fitness compared to both end point and ancestral strain (Elena and Lenski, 2003). Finally, recent technological advances with next generation DNA sequencing make examination of the entire genome and proteome of these organisms possible.

In recent years growing advances in DNA technology shifted the focus of studies on experimental adaptation towards whole genome approach and have given rise to functional genomics (Conrad et al., 2011; Hajibabaei et al., 2007). Lessons learnt from these findings provide invaluable insights into the understanding of evolutionary dynamics of microbial populations undergoing adaptation, through mapping the mutational pathways during adaption to antibiotics in *E. coli* (Toprak et al., 2012), identifying the importance of intracellular interactions within a population for the mechanisms for resistance (Lee et al., 2010), and dissecting the complexity in the link between genomic and adaptive evolution (Barrick et al., 2009). In addition, the whole genome sequencing with DNA barcoding allowed investigators to measure the frequencies of specific alleles during adaptation of *E. coli* (Barrick et al., 2009; Lee and Palsson, 2010; Lee et al., 2010). These results were summarized by Conrad et al. and showed that single nucleotide substitutions are the most common mutations and typically constitute about two-thirds of all mutations (Conrad et al., 2011). Deletions, insertions and larger sequence movements are less frequent and make up 29%, 7% and 3% of all mutations observed, respectively (Conrad et al., 2011).

As described previously, the effect of beneficial mutations on fitness of organisms can be assessed through allelic replacement, where a mutation is introduced into an ancestral strain for growth rate determination or a direct competition assay for fitness. However, when multiple mutations are present this approach can be very time consuming and often not feasible. Focusing adaptation on a single gene greatly simplifies the identification of adaptive changes and enables a quantitative approach to study evolution. This approach is termed weak link (Counago et al., 2006; Counago and Shamoo, 2005).

### 1.4. Functional synthesis as a powerful tool in understanding the link between genotype –phenotype- fitness

Dean and Thornton coined the term "functional synthesis" to describe the convergence of molecular evolution, biochemistry and structural biology to deconvolute the complex relationship between biochemical function and evolutionary outcomes (Dean and Thornton, 2007). Molecular biology makes it possible to reconstitute changes in genes and to study their effects on the organism under controlled conditions while in vitro assays allow quantitative evaluation of a protein. Crystallographic studies on isolated proteins provide an additional means to evaluate how genetic changes make alterations in protein structure and function. Together, these methods serve as powerful tools for dissecting the role of individual mutations on the *in vivo* fitness and therefore provide a method to correlate population dynamics and changes in protein function. Most importantly, they serve as means to directly link genotypic changes to phenotypes possibly providing insights into the mechanisms of adaptation. The ultimate goal of functional synthesis would be to successfully predict evolutionary outcomes by

correlating the evolutionary events on a population level to the molecular phenotypes at the protein level.

## 1.5. The link between greater fitness and the physiochemical properties of proteins

Conceptually, understanding the relationship between molecular properties of adaptive mutants and their consequences on cellular function is relatively simple. First, mutations responsible for a particular level of fitness have to be identified along with their functional intermediates. Then, the biophysical properties of proteins and their effects on fitness have to be measured to provide a link between genotype, phenotype and fitness. Finally, a model relating fitness to function that captures mechanism of selection has to be developed. In practice, correlating the changes in genotype to phenotype and its effects on fitness has proven very challenging. However reducing the complexity of the system to a single gene greatly simplifies interpretation of results allows for quantitative analysis and most importantly enables determining the structural and biochemical basis of adaptation. This approach can be particularly important in turning evolutionary biology into a predictive science for the study of antibiotic resistance.

What are the biophysical parameters that correlate with fitness? Natural selection favors mutations that improve cellular function and therefore it was postulated that changes in *in vivo* activities of enzymes essential for survival to specific selection environment closely correlate with organismal fitness (Kacser and Burns, 1981). This hypothesis was tested by Dean and colleagues who showed that growth rates of *E. coli*

were indeed directly correlated with activity of single gene expressing β–galactosidase (Dean et al., 1986).

In enzyme evolution, genetic changes can potentially alter any of the parameters that define their catalytic performance which include: $K_M$, the Michaelis constant; $k_{cat}$, the catalytic rate constant; enzyme concentration; and $k_{cat}/K_M$, enzyme efficiency (Watt and Dean, 2000). $K_M$ and $k_{cat}$ can be directly altered by amino acid substitutions or by changes in stability, which affects the effective enzyme concentrations and therefore alters $V_{max}$ or $k_{cat}[E]_{total}$. Enzyme concentration can also be altered by changes in regulatory genes that affect protein expression patterns and consequently alter the total enzyme concentrations. It is important to note, that although mutation affecting either $k_{cat}$ or total enzyme concentration can have similar effects on the overall catalytic performance, there is a difference in fitness costs associated with these changes. Changes in enzyme concentration impose a significant burden because of costs associated with protein synthesis, whereas changes in $k_{cat}$ are usually traded off for stability (Romero and Arnold, 2009).

### 1.5.1. Trade-off between protein function (activity) and stability dictates the evolutionary dynamics.

The distribution of fitness effects of protein mutations changes exponentially. With the increasing number of mutations to a single protein, the chance for decrease or loss of function also increases (**Figure 1.3**) (Soskine and Tawfik, 2010). This is because proteins are only marginally stable and most mutations have destabilizing effects that can result in increased aggregation and degradation of proteins (DePristo et al., 2005). As

mentioned before, most mutations that affect protein function can have deleterious effects on protein stability and consequently change the *in vivo* expression protein levels. Therefore, the ability to maintain functional and structural integrity of an enzyme is critical during adaption. Proteins with higher stability can tolerate more initial mutations and therefore increase the number of possible subsequent mutations. For example, the number of functional mutations was significantly higher for a more stable variant of TEM-1 β-lactamase (M182T) when compared to of the wild-type enzyme (Bloom et al., 2005).
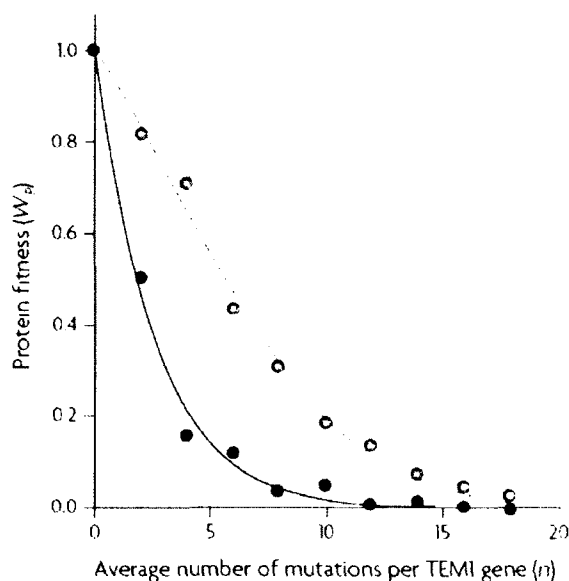


**Figure 1.3. Protein fitness declines as the number of spontaneous mutations begins to accumulate.**
Blue corresponds to fitness of TEM-1 β-lactamase measured at wild-type ampicillin levels, 2500 μg/ml, and red corresponds to fitness determined at lower concentration of 12.5 μg/ml. Both scenarios show a decline in fitness with the higher number of mutations, however at higher ampicillin concentration, the decline is very sharp compared to lower concentrations where more mutations maintain functional genes. Reprinted by permission from Nature Publishing Group: Nature Reviews Genetics (Soskine and Tawfik, 2010), copyright (2010).

Tokuriki et al. have argued that protein dynamism also plays a critical role in evolvability (Tokuriki and Tawfik, 2009b). The conformational variability due to side chain fluctuations or more radical domain movements can serve as basis of protein evolvability (Tokuriki and Tawfik, 2009b). This can be seen in enzymes which catalyze more than one reaction and often require flexibility for catalysis (Skopalik et al., 2008). For example, the most rigid variant of mammalian cytochromes P450, CYP2A6 exhibits narrow substrate specificity whereas another variant, CYP3A4, recognizes a broad range of substrates and is the most flexible (Skopalik et al., 2008). Directed evolution studies of TEM-1 β-lactamase also show that mutations responsible for wider specificity were mostly located in flexible loops in the active site (Tokuriki et al., 2008).

In general, proteins with marginal stabilities can tolerate fewer mutations because most will be on average, highly destabilizing. On the other hand, proteins with higher stability are likely to accept more mutations and consequently adapt further to new functions (Bloom et al., 2006). Francis Arnold referred to this phenomenon as the protein stability threshold (Romero and Arnold, 2009). The distance from the minimal stability threshold of a parent protein determines the number of tolerable changes. Proteins with higher a minimal stability margin can explore a wider mutation range and thus have a better opportunity to evolve new functions (**Figure 1.4**). Stabilizing mutations often occur in the background of adaptive mutations that improve function but decrease stability.
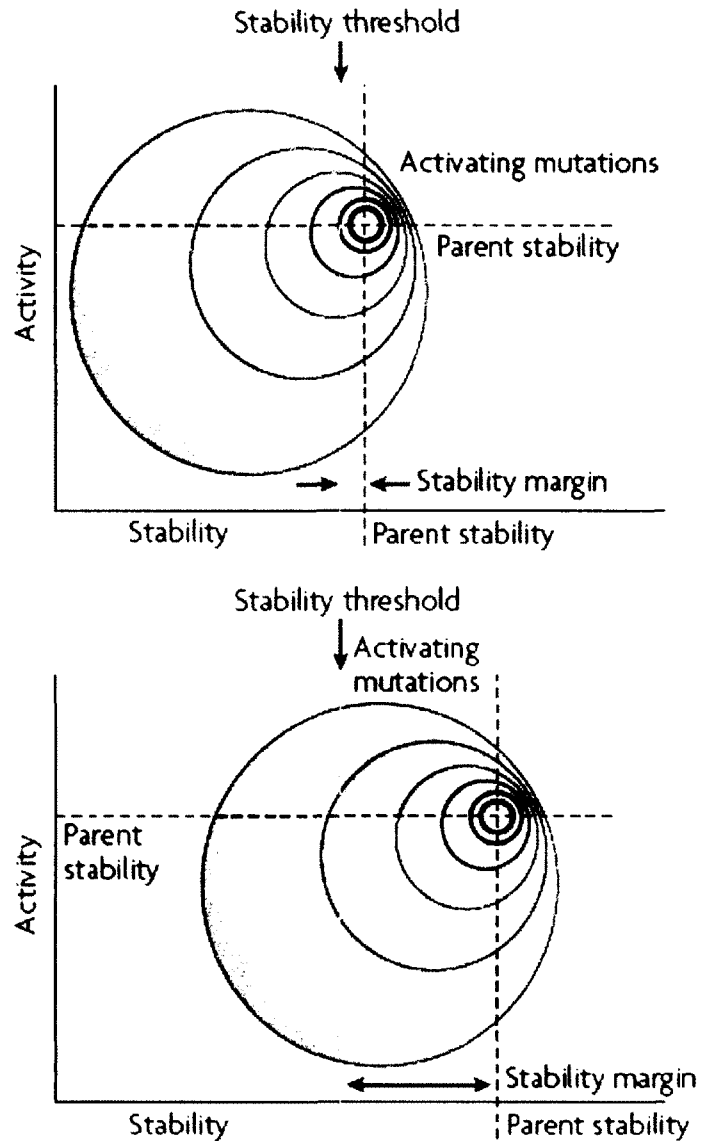
**Figure 1.4. Stability threshold in protein evolution.**
The distance from the minimal stability threshold (dashed line) of a parent protein determines the number of tolerable changes. Proteins with marginal stabilities (top panel) can tolerate fewer mutations (green area) because most mutations are highly destabilizing. On the other hand, proteins with higher minimal stability margin (bottom panel) can explore wider mutation range and thus can tolerate more mutations (green area). Reprinted by permission from Nature Publishing Group: Nature Reviews Molecular Cell Biology (Romero and Arnold, 2009), copyright (2009).

Using *in vitro* directed evolution, Weinreich et al. have shown that the effect of single amino acids substitution in β-lactamase on the resistance of an organism was highly dependent on the order and the genetic background in which the mutations arose (Weinreich et al., 2006). The benefit of each individual mutation on the resistance of the organism was greatly decreased in the absence of the other mutations, which is due to the fact that some mutations, e.g. G238S, increase the activity of the enzyme while others, e.g. E104K improved protein's stability, showing that only when both mutations were present simultaneously, the optimal organismal fitness was reached (Orencia et al., 2001). This work highlights the importance of epistasis during adaptation. Epistasis defines the dependence of the fitness effect of a particular mutation on the genetic background that it appears in a selective environment (Weinreich et al., 2005) Consequently, epistasis determines the interaction of accessible mutations and produces constrained pathways that are able to maintain both protein activity and stability (Salverda et al., 2011; Weinreich et al., 2005).

### 1.5.2. Non-adaptive mutations play an important role in evolutionary dynamics

During adaptation, not all selected mutations are adaptive. Mutations that do not directly improve protein function can be selected for in the presence of compensatory mechanisms. The mutational effects can be buffered intrinsically by compensatory mutations or through extrinsic mechanisms, e.g. gene duplication events, regulatory mechanisms and chaperones (Camps et al., 2007; Soskine and Tawfik, 2010). Gene duplication helps to buffer otherwise deleterious mutations by relaxing the selective pressure. A duplication event often leads to increased protein level but also provides new

material for selection which consequently opens up new adaptive pathways. In addition, duplication can produce genes under control of different regulatory elements, which may also lead to new functions. Upregulation of protein expression has also been recognized as a compensatory mechanism in the adaptation of TEM-1 to higher antibiotic concentrations where a single mutation to a promoter was responsible for 10-fold increases in protein levels (Goldsmith and Tawfik, 2009). Chaperones or heat shock proteins also play important roles in buffering the effects of mutations that lead to misfolding or aggregation. For example, the number of tolerable mutations was increased by two fold when chaperones GroEL/GroES were overexpressed in *Escherichia coli* (Tokuriki and Tawfik, 2009a). Finally, compensatory mutations are often found to accompany adaptive mutations that improve function but also decrease stability. These mutations play an important role in evolutionary dynamics because they allow access to otherwise inaccessibly to adaptive pathways (DePristo et al., 2005; Tokuriki and Tawfik, 2009b). In resistance of β-lactamese towards cefataxime, compensatory mutations were necessary to restore organismal fitness (Orencia et al., 2001). In this system, mutation G238S that improved enzymatic catalysis was accompanied by the stabilization mutation M182T that always appeared in its background (Orencia, 2001).

### 1.5.3. Experimental tests linking biophysical properties of enzymes to organismal fitness

To date, several studies on molecular evolution of enzymes have shown that based on their biophysical properties a quantitative link between enzyme function and fitness can be derived (for review see (Dean and Thornton, 2007). One of the most characterized model systems to study adaptation of antibiotic resistance is β-lactamase,

which is known to confer resistance to penicillins. Multiple variants of TEM-1 β-lactamase with increased resistance towards cefataxime (a third generation cephalosporin) have been identified in natural and in laboratory settings. Directed evolution of TEM-1 β-lactamase, was successful in recapitulating clinically relevant mutations responsible for increased resistance towards cefataxime (Orencia, 2001). Analysis of the crystal structure of TEM-52 (a triple mutant of the TEM-1 β-lactamase - clinical isolate) revealed that mutation G238S, which increases the Minimal Inhibitory Concentration (MIC) towards cefataxime about eight fold was located in the active site of the protein and resulted in a widening of the catalytic site but simultaneously lowered stability of the enzyme. When G238S is combined with a second mutation, E104K, a significant increase in the MIC for cefataxime (267 fold) was observed (Orencia et al., 2001). The role of a third mutation, M182T was to compensate the destabilized effects of the other variants as M182T by itself does not increase the MIC (Orencia, 2001). Individual mutations alone did not have great fitness effects when compared to combinations of mutations, supporting important role of epistasis (discussed in section 1.5.1). A recent directed evolution study of a metallo-β-lactamase identified mutants that altered catalysis of the enzyme by altering flexibility (Tomatis et al., 2008). Flexibility was directly linked to higher fitness (resistance) and showed that protein dynamics can be a target for natural selection to act upon particularly in enzymes.

Counago et al. (2006) established a direct link between changes in fitness and an increase of enzymatic stability and activity (Counago et al., 2006). In this study, adaptation of a single maladapted essential gene encoding adenylate kinase (adk) was investigated. The biophysical analysis of mutants that arose during adaptation of the adk

gene to temperature revealed a shift in the activity maxima, stability and kinetic properties of all mutant enzymes. Further characterization of the mutant enzymes was linked to protein activity and folding dynamics (Counago et al., 2008; Pena et al., 2010). In addition, based on the *in vitro* properties of the mutant enzymes, a fitness function was built to capture the evolutionary trajectories in the population and to predict the success of adaptive mutants during adaptation of adenylate kinase to temperature (Counago et al., 2006; Pena et al., 2010). This study has shown that molecular properties that result from adaptation can be used to build a model that predicts evolutionary population dynamics. Ultimately, this work became the inspiration and provided basis for research presented in this thesis.

### 1.6. Focus of this thesis

The combination of experimental evolution to identify adaptive changes in proteins undergoing selection, and their subsequent physicochemical characterization, provides a physical basis for linking changes in protein structure and function to organismal fitness. This is particularly relevant for the study of antibiotic resistance where the evolution of drug resistant bacteria is clinically important. Therefore, we set to discover the principles that govern the adaptation of naturally evolved *Escherichia coli* strains to tetracycline antibiotics and develop a quantitative model predicting the outcomes in evolution of antibiotic resistance to this class of drugs.

## 1.7. Antibiotic resistance: origins and mechanisms of drug resistance

Natural selection of antibiotic resistant pathogens is a continuously growing global challenge (Andersson and Hughes, 2010). Introduction of new antibiotics into clinics is almost immediately succeeded by incredibly fast emergence and proliferation of antibiotic resistant bacteria (Alekshun and Levy, 2007). Today, resistance to all clinically available antibiotics has been observed worldwide (Cohen, 2000). Although it might be tempting to view the rise of resistant pathogens as solely linked to the extensive use of antibiotics in clinical and agricultural environments, recent data suggest that spread of antibiotic resistance genes is an ancient and naturally occurring phenomena (D'Costa et al., 2011). This finding poses new challenges in the development of novel therapeutics and strategies used for their implementation, leading to the fundamental question; what is the biophysical basis for resistance? Most clinically available antibiotics originate from naturally occurring compounds produced by microorganisms such as fungi or soil-dwelling bacteria (Walsh, 2000). They are mainly targeted at cell wall biosynthesis, DNA replication and repair, or protein synthesis.

The determinants of resistance are often located on different mobile genetic elements such as plasmids, transposons and bacteriophages that can be transferred between the same and different bacterial taxonomic groups (Levy and Marshall, 2004). In addition, drug resistance can be altered intrinsically through a step-wise mechanism which occurs by the sequential accumulation of chromosomal mutations that result in the upregulation of the gene expression responsible for resistance or the modification of a pre-existing protein for new specificity (Alekshun and Levy, 2007). Despite different molecular mechanisms, natural selection is the common driving force that leads microbial

populations to adapt to certain drugs by improving their fitness in the environment (Barlow and Hall, 2003; Elena and Lenski, 2003; Hall, 2004).

## 1.8. Tetracycline resistance as a model for experimental evolution

Tetracyclines (**Figure 1.5**) are broad-spectrum antibiotics that inhibit protein synthesis by non-covalently binding to the ribosome and blocking the binding of aminoacyl-tRNA to the ribosomal acceptor site (Speer et al., 1992). A high-affinity binding site for tetracycline is located on the 30S ribosomal subunit (Connell et al., 2003). Two independent crystal structures of the tetracycline-bound 30S subunit showed either two (Brodersen et al., 2000) or six (Pioletti et al., 2001) binding sites and it remains unclear as to which sites are inhibitory (**Figure 1.6**).

Since their discovery in the 1940's, tetracyclines have been commonly used in medicine against both Gram-positive and Gram-negative bacteria, atypical organisms and protozoan parasites, as well as in agriculture as animal growth promoters (Chopra and Roberts, 2001). Their broad use in medicine, agriculture and aquaculture has resulted in an increase of tetracycline resistance (Thaker et al., 2010).

To date, 40 proteins that confer tetracycline resistance have been identified (Thaker et al., 2010). These proteins are classified into three groups based on their mechanism of resistance: efflux, ribosomal protection, and enzymatic (**Figure 1.7**). The efflux class of tetracycline resistance proteins is common in Gram negative bacteria (*Shigella* and *Salmonella*) (Chopra and Roberts, 2001). All efflux tetracycline resistance proteins export tetracycline from the cell, which reduces the intracellular drug

concentration (Chopra and Roberts, 2001). TetA is one of the most common members of

this class and often complements the ribosomal protection proteins
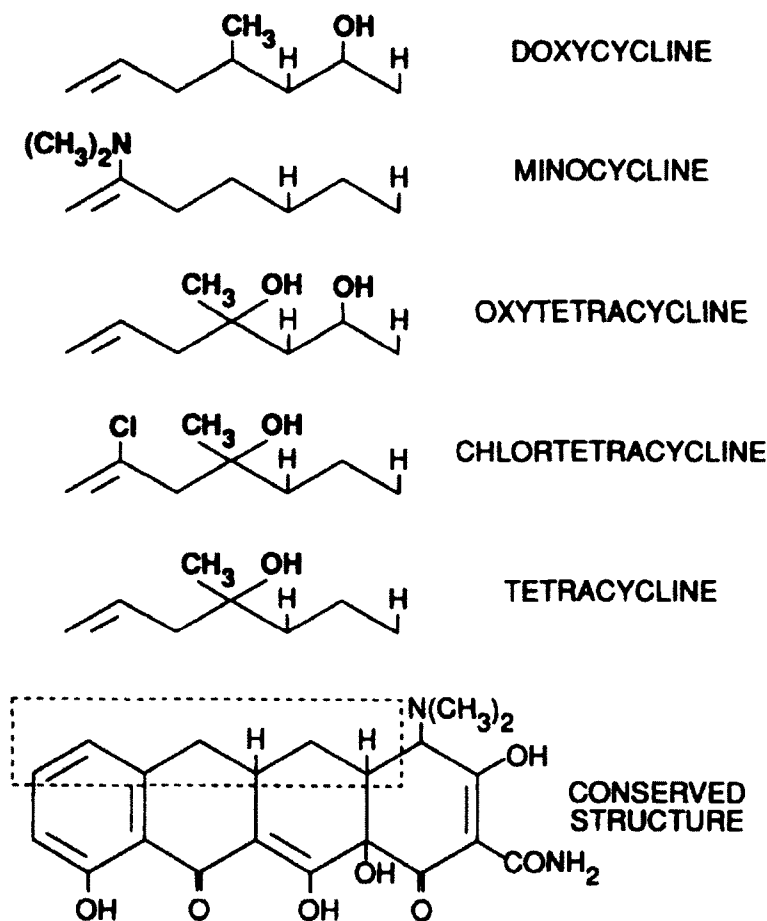


**Figure 1.5. Chemical structures of tetracycline derivatives.**
Tetracyclines share a common conserved structure shown on the bottom of the figure that consists of four (tetra) hydrocarbon rings. Tetracycline derivatives differ amongst each other by chemical modification of positions shown in a box on the conserved structure. Modifications to the hydrophobic portion of the molecule are known to inhibit its ribosome binding. Reprinted by permission from American Society of Microbiology, Clinical Microbiology Review:(Speer et al., 1992), copyright (1992).

A                                      B
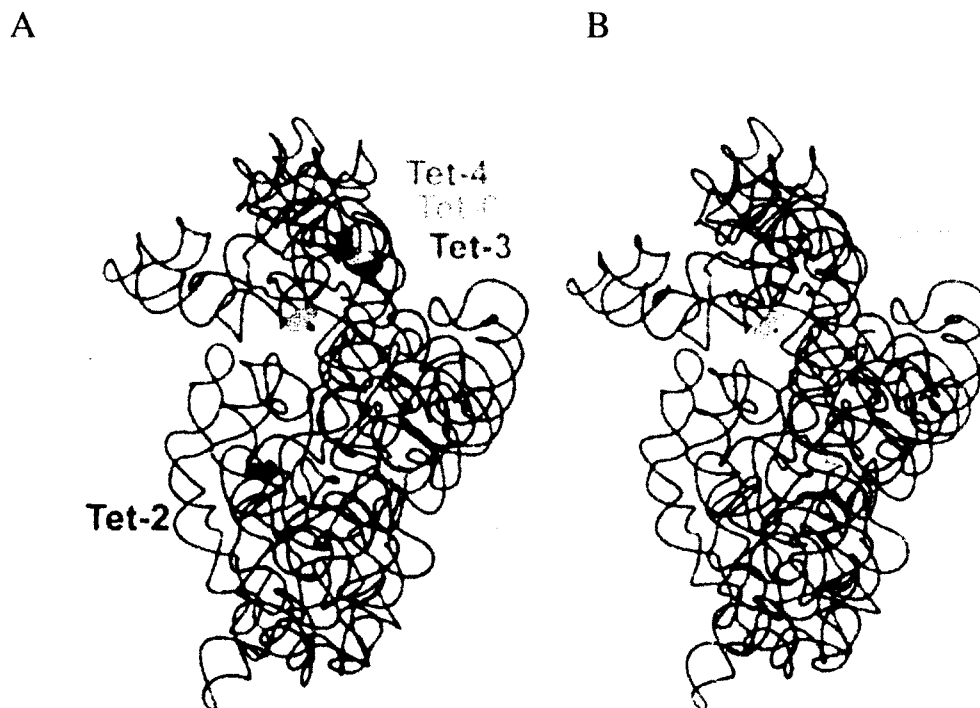


**Figure 1.6. Tetracycline binding sites on the 30S ribosomal subunit.**
Tetracycline binding sites determined by Pioletti et al. with tetracycline bound at several sites. The increasing numbers Tet-1 to Tet-6 indicates the relative occupancy in the electron density map. B) Primary tetracycline binding site is shown in red and secondary site in orange as determined by Brodensen et al. Red binding sites on both structures overlap suggesting that this could be a high-affinity binding site. Reprinted by permission from American Society of Microbiology Antimicrobial Agents and Chemotherapy: (Connell et al., 2003), copyright (2003).
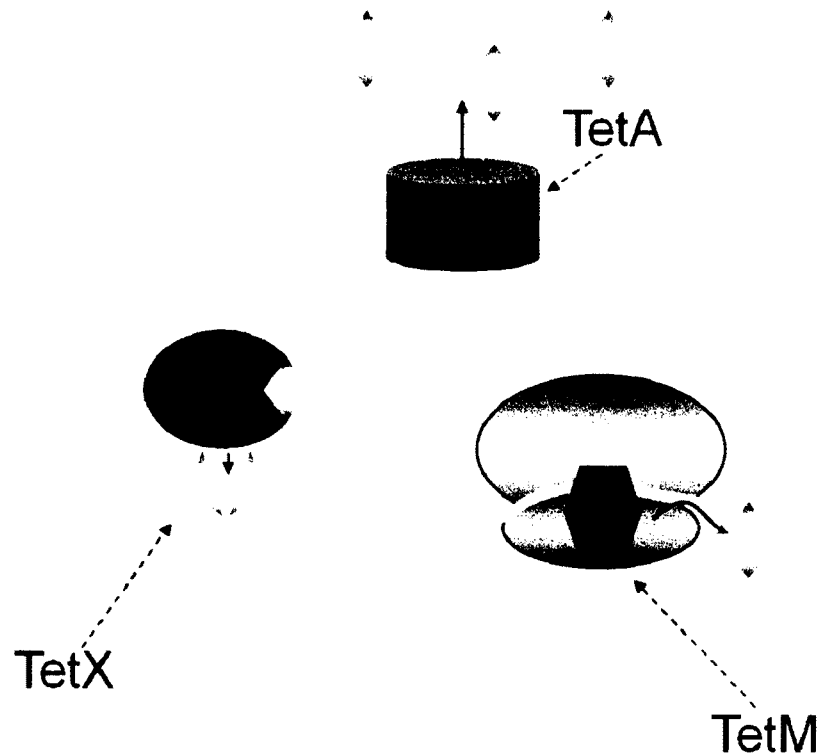
**Figure 1.7. Different mechanisms of tetracycline resistance.**

Tetracyclines accumulate inside the cell at high enough concentrations to inhibit protein translation. Bacteria can upregulate efflux pumps (TetA) that translocate antibiotics out of the cell, lowering their internal concentrations. In a different scenario, ribosomal protection proteins are expressed (TetM –green) that bind to the ribosome and displace tetracycline (yellow). A less common mechanism for tetracycline resistance is through expression of soluble enzymes (TetX -blue) that bind and enzymatically deactivate tetracyclines.

The ribosomal protection class of proteins (RPPs) achieve tetracycline resistance by releasing bound tetracycline from the ribosome (Connell et al., 2003). This is the most common mechanism among all tetracycline resistance strains. An example of a RPP that has been well characterized is TetM. This 73 kDa protein was initially isolated from *Streptococcus* and has been found in a wide range of pathogens including mycoplasmas, *Neisseria gonorrhoeae*, *Clostridia difficile* and *Staphylococci* (Burdett, 1991). It shows 75% sequence identity to TetO, a RPP that has been studied extensively. In addition, both

TetM and TetO share a 25% sequence identity in the N-terminal region to *E. coli* EF-G, thus suggesting how they might bind to the ribosome. When TetM binds to the ribosome, the tetracycline dissociation constant increases from 5 to 30 µM causing its displacement (Connell et al., 2003). The cryo-EM studies of TetO indicated that RPPs bind to the ribosome in a manner similar to EF-G (Spahn et al., 2001).

Enzymatic modification was not associated with the inactivation of tetracyclines until the discovery of TetX isolated from *Bacteroides* transposons Tn4351/Tn4400 (Speer et al., 1991). TetX is a 44 kDa flavin adenine dinucleotide (FAD) containing monooxygenase that requires NADPH and oxygen for its enzymatic activity (Wangrong Yang, 2004). Predictions based on primary sequence analysis reveal the presence of a flavin-binding domain at the N-terminus end followed by a Pfam monooxygenase fold at the C-terminus end (Wangrong Yang, 2004). Two orthologs of *tetX* have been identified from another transposons, CTnDOT, in *Bacteroides* (Whittle et al., 2001). TetX1 shows 66% identity to TetX, but it is lacking the FAD binding domain. TetX2 shows 99% identity to TetX (**Figure 1.8**).
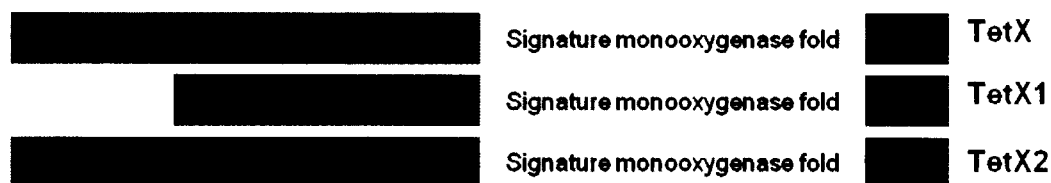


**Figure 1.8. Architecture of TetX from *Bacteroides fragilis* Tn4351/Tn4400 and its orthologs TetX1 and TetX2 from *Bacteroides* trasposon CTnDOT.**

Recently, the molecular mechanism by which TetX2 deactivates tetracycline has been elucidated (Wangrong Yang, 2004). TetX2 hydroxylates tetracycline at position

C11a in the presence of NADPH, O₂ and MgCl₂. The product of this reaction is unstable

at physiological pH and causes a change in color of the reaction mixture from yellow to

reddish brown. In addition, TetX can inactivate Tygacil®, a tetracycline antibiotic that

was recently approved by the FDA (**Figure 1.9**), thus suggesting a possible emergent

threat in the clinical environment (Moore et al., 2005).
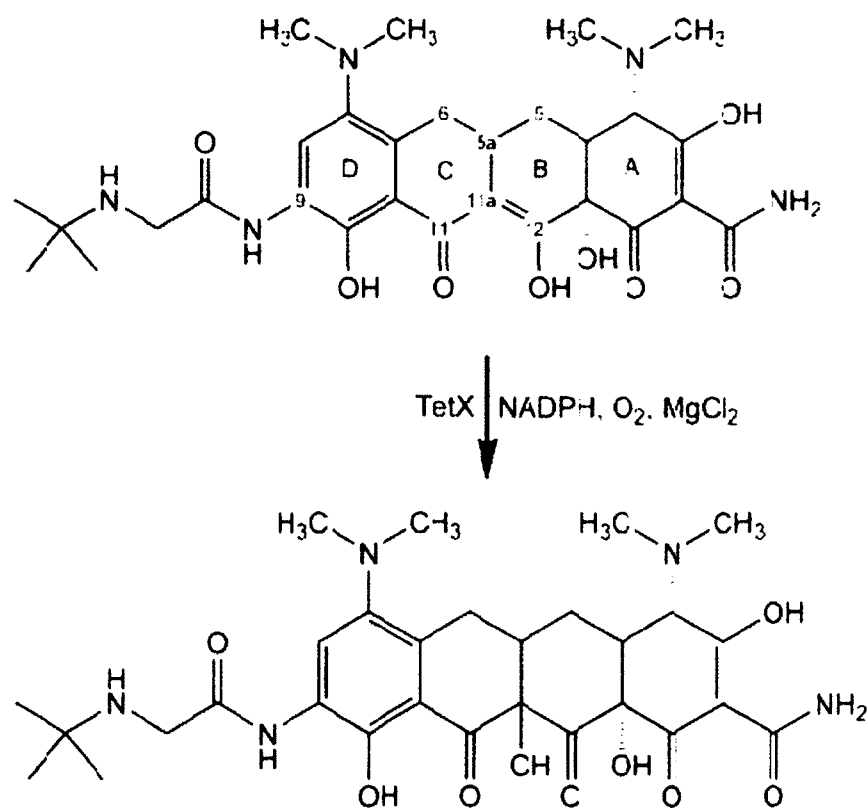


**Figure 1.9. Regiospecific hydroxylation of tigecycline catalyzed by TetX2 in the presence of NADPH, O₂ and MgCl₂.**
The inactivation of tigecycline results in the formation of 11a-hydroxytigecycline by hydroxylation at position 11a (Moore et al., 2005).

## 1.9. Objectives of this research

Experimental evolution studies on microbial population have been focused on establishing a link between genetic variation and fitness and successfully shown that adaptation, history and chance play an important role on evolutionary outcomes. On the other hand, the biophysical studies of adaptive mutations have only recently started to identify biophysical properties that correlate protein function with organismal fitness. The thesis work presented here aims to bridge experimental evolution with the biophysical analysis of a specific class of adaptive mutations to investigate the biochemical and structural basis for adaptation in a quantitative manner. The specific aim of this work is to test the hypothesis that physicochemical characterization of all potential adaptive mutants provides basis for predicting evolutionary dynamics of a microbial population undergoing adaptation to antibiotics. As our model system we chose *Bacteroides thetaiotaomicron* TetX2 that was previously identified as a novel oxidoreductase with broad activity against tetracyclines (Wangrong Yang, 2004).

First, we adapted *E. coli* expressing a chromosomal copy of *tetX2* to increasing minocycline concentrations where we identified a single adaptive mutant with increased resistance towards MCN. We then used *in vitro* error-prone mutagenesis to identify all possible adaptive mutants of TetX2 and tested their *in vivo* fitness using a growth rate assay at clinically relevant drug concentrations. The outcomes of these experiments are presented in Chapter 3. Second, we characterized the biophysical and structural properties of mutant enzymes using X-ray crystallography and kinetics to characterize the enzymes and link the physicochemical properties of adaptive mutants to increased organismal fitness under specific selection conditions (i.e. relevant MCN concentration

and temperature). These results are shown in Chapter 4 (kinetics) and Chapter 5 (crystallography). Finally, in Chapter 6, I discuss the overall finding from this thesis work and our current and future efforts in building an *in silico* model to test whether using experimental data we can predict the *in vivo* growth rates of adaptive mutants and their evolutionary dynamics in a population undergoing selection.

# Chapter 2 Materials and Methods

## 2.1. Strains and constructs

### 2.1.1. Construction of strains used for natural evolution experiments and growth studies

Integration of *tet(X2)* into the chromosome of *E. coli* strain BW25113 [F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), lambda- ,rph-1, Δ(rhaD-rhaB)568, hsdR514, pKD46] was performed using a short homology recombineering approach (Datsenko and Wanner, 2000). This method was introduced by Datsenko and Wanner (2000) and relies on phage λ Red recombinase expressed from a low copy plasmid pKD46. This plasmid is a temperate sensitive replicon, so it is easily cured by shifting the growth temperature from 30 to 37°C. *Tet(X2)* was recombined into the well regulated *spc* operon of *E. coli* between *prlA* (SecY) and *rmpJ* (L36) (**Figure 2.1**). Linear DNA containing a ribosome binding site (rbs) and *tet(X2)* sequence flanked by short homologous sequences to *prlA* and *rpmJ* was amplified by Polymerase Chain Reaction (PCR) using pUC19-*tet(X2)* construct as a template. Vector pUC19-*tet(X2)* was constructed by subcloning *tet(X2)* into a pUC19 vector behind a constitutively active lacZ promoter (Invitrogen, Carlsbad, CA). Forward primer used for amplification of *tet(X2)* included 50 nts homologous to a sequence on *prlA* and 20 nt long sequences corresponding to the ribosome binding site (*rbs*) located upstream of *tet(X2)* in pUC19 vector. Reverse primer included a 50 nt homologous to a sequence on *rpmJ* and the last 24 nt of *tet(X2)*. Primer sequences are shown below: The sequences corresponding to the chromosomal insertions sites are underlined.

5'<u>GAAGAAGGCGAACCTGAAAGGCTACGGCCGATAATTGGTCGCCCGAGAAG</u>

TAGAGGACCTAGAAGGAGAC (forward)

5'<u>CGGCATAATTTCTTGACGGAAGCACGAACTTTCATTTTTACTCTCCGTAA</u>

<u>C</u>TTATTATACATTTAACAATTGCTG (reverse)



*E. coli* chromosome
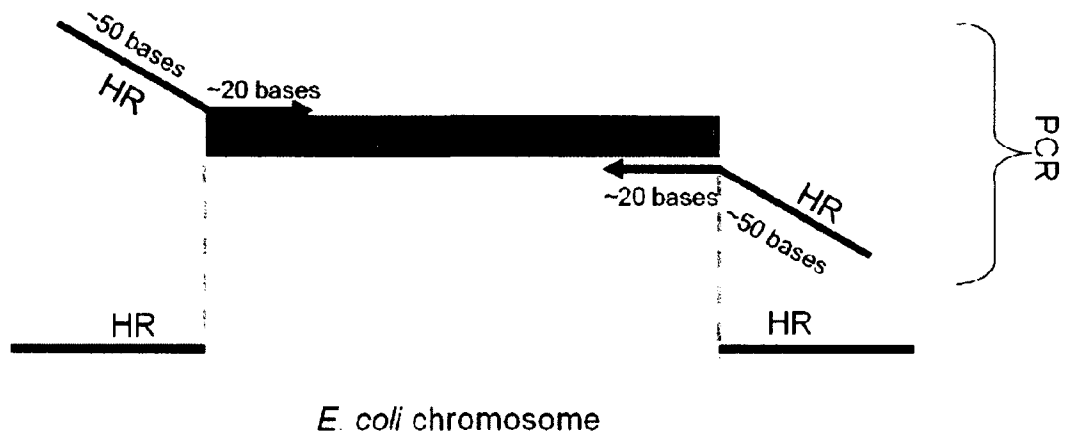
**Figure 2.1. Schematic representation of short homology recombineering strategy used for integration of *tet(X2)* into the *spc* operon of *E. coli*.**

Recombinants were screened on LB agar plates containing 10 μg/ml of tetracycline. Conformation PCR was performed using genomic DNA isolated from colonies grown at 10 μg/ml of tetracycline using UltraClean® Microbial DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA) with primers aligning to the chromosome outside of the integration site (5'ATGGCTAAACAACCGGGATTAGATTTTCAAAGT GC and 5' GCGATTACGGCATGCTTATGATCAGGAATG). In the case of successful recombination, the PCR product amplified with these primers was expected to consist of 3kb whereas, if the gene of interest was not integrated, then the product was only 1.8kB.

The 3 kb PCR products were sent for sequencing for final confirmation of successful integration (all sequencing was done by SeqWright Inc, Houston, TX). Furthermore, the temperature sensitive helper vector pKD46 was cured by growth at 37°C overnight. Cells grown at temperature not permissive for the replication of the vector pKD46 lost their ability to grow on ampicillin (pKD46 harbors *bla*). Glycerol stocks of BW25113 cells carrying chromosomal copy of gene of interest were prepared by mixing 50% sterile glycerol and overnight liquid cultures in 1:2 ratios and were stored at -80°C. All mutants examined in this study were integrated into the chromosome using the same primers and procedures as used for the wild-type *tet(X)* gene.

### 2.1.2. Generation of constructs for expression of mutant enzymes

The original vector harboring wild-type *tet(X2)* was a generous gift from Dr. G. D. Wright (McMaster University, Canada). All mutant enzymes examined in this study were introduced into *tet(X2)* by site-directed mutagenesis using QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Incorporated, Santa Clara, California) with an expression vector, pET-28b(+) containing *Bacteroides thetaiotaomicron tet(X2)* (residues 11-388) as a template. Primers for the mutants are listed below:

Primer pair for T280S:

5' CTACAAAGAATTGATTCATTCGACGTTGTCATTTGTAGG 3'

5' CCTACAAATGACAACGTCGAATGAATCAATTCTTTGTAG 3'

Primer pair for F235Y:

5'CCAATAATAATGGTGCATTGCATTATGGAATAAGTTTTAAAACACCTG 3'

5' CAGGTGTTTTAAAACTTATTCCATAATGCAATGCACCATTATTATTGG 3'

Primer pair for N371I:

5' CACAAGAAGAATCAACTCAAATCGAAATTGAAATGTTTAAACC 3'

5' GGTTTAAACATTTCAATTTCGATTTGAGTTGATTCTTCTTGTG 3'

Primer pair for N371T:

5' CACAAGAAGAATCAACTCAAACCGAAATTGAAATGTTTAAAC 3'

5' GTTTAAACATTTCAATTTCGGTTTGAGTTGATTCTTCTTGTG 3'

Primer pair for K64R:

5' CCTTGACCTACACAGAGGTTCAGGTCAG 3'

5' CTGACCTGAACCTCTGTGTAGGTCAAGG 3'

Primer pair for S326I:

5' GCAGGGAGTAAATATTGGGTTGGTGGATGC 3'

5' GCATCCACCAACCCAATATTTACTCCCTGC 3'

Mutagenesis reactions were performed according to the protocols supplemented with the kit. The reaction products were transformed into chemically competent XL-1 Blue cells (Agilent Technologies, Incorporated, Santa Clara, California) and were plated on LB-agar plates containing 50 µg/ml kanamycin. Plasmid DNA was isolated from single-colonies using Wizard® *Plus* SV Minipreps DNA purification system (Promega, Madison, WI) and the DNA encoding *tet(X2)* was sequenced.

## 2.2. Serial-passage experiment

Adaptation of *E. coli* BW25113-*tet(X2)* to increasing minocycline (MCN) concentrations was performed by serial transfer experiment carried out in 250 mL Erlenmeyer flasks (**Figure 3.1**). MCN concentrations below the Minimal Inhibitory Concentration (MIC) values were used as starting dosage for the serial-passage evolution experiments. The experiments were performed in 50 mL of sterilized LB media supplemented with MCN concentrations ranging between 2 to 5 μg/mL over 3 days for *E. coli* BW25113 and from 10 to 320 μg/mL over 10 days for *E. coli* BW25113-*tet(X2)*. Each day a 50 μl of overnight culture was transferred to fresh LB broth supplemented with higher antibiotic concentration and was incubated in an orbital shaker shaking at 250 revolutions per minute (RPM) overnight at 37°C. The drug dosages were increased daily according to the following schedule 10, 16, 24, 36, 52, 80, 120, 180, 300, 320 μg/ml of final MCN concentration from day 1 to day 10. Samples from each day of the maintained culture were 1) plated on LB-agar plates and genomic DNA originating from single colonies was purified and the *tet(X2)* gene sequenced 2) stored as glycerol stock at -80°C. The serial passage experiment was repeated twice under the same conditions.

## 2.3. Determination of Minimal Inhibitory Concentrations (MICs)

Minimum inhibitory concentrations (MIC) were determined by agar-dilution method (**Figure 3.2**) according to the National Council of Clinical Laboratory Service (NCCLS) guidelines where, the MIC is defined as minimum concentration of antibiotics that prevents visible growth. Briefly, *E. coli* BW25113-*tet(X2)* and *E. coli* BW25113-*tet(X2)*$_{T280A}$ cells from glycerol stocks were plated on cation-adjusted Mueller-Hinton agar (MHA) plates. MHA plates containing serially diluted antibiotics (minocycline, tetracycline or tigecycline) were prepared at concentrations ranging between 2-128 μg/ml. Sterile loops were used to prepare inocula from cells grown on non-selective MHA plates that were diluted in MH broth to OD = 0.08-0.1 and were set aside. A subset of these cells were further diluted by 10-fold right before spotting 10 μl on MHA plates.

Cells diluted to OD= 0.08-0.1 were used in an E-test (Figure 3.3). For this test, commercially available E-strips (AB Biodisk, bioMérieux, Incorporated, Durham, NC) containing a gradation of antibiotics were used to determine the MIC values for minocycline also on cation-adjusted MHA. Cells were evenly distributed across MHA plate using a cotton swab and E-strip was plated in the center. As soon as the strip is laid on the plates, the antibiotics diffuse into the agar. After 24 hour incubation period at 37°C, zone of inhibition which formed an ellipsoid was examined. The values at which the ellipsoid touched the strip were considered the MICs.

The drug dosages for evolution experiments were designed based on the MIC tests performed in liquid media. Single colonies were used to inoculate 10 mL LB media grown for eight hours. 10 μl of each of these cultures were used to inoculate 10 ml LB

liquid media containing MCN concentrations ranging from 0 to 100 µg/mL for the

BW25113-*tet(X2)*. The cells were then incubated at 37°C overnight and MIC values

were determined to be the lowest concentrations at which the cells failed to grow within

24 hours. MICs for MCN of evolved strain (day 9) were also determined by MIC test in

liquid media.

## 2.4. Determination of Growth Rates at Various Minocycline Concentrations

Growth rates of *E. coli* expressing chromosomal copies of variant

enzymes were measured and compared at various MCN concentrations ranging from 0-

50 µg/ml. Cells BW25113 preserved in glycerol stocks were re-streaked on fresh LB-agar

plates containing 10 µg/ml of tetracycline. Single colonies were picked and used to

inoculate 1 mL of sterile LB in a deep well 96-well block and were incubated in an

orbital shaker at 37°C for 24 hours. The stationary cultures were diluted 100 times in LB

containing appropriate MCN concentrations (990 µL of LB was mixed with 10 µL of

overnight culture) and 150 µL of the mixture was transferred to a 96-well flat bottom

plate (BD Falcon, BD, NJ). Growth of bacteria was monitored by measuring the

absorbance at 600 nm recorded every 5 minutes over 24 hours in a Synergy 2 micoplate

reader (BioteK®, Winooski, VA, USA).The fastest growth was defined as a slope of a

tangent parallel to linear log phase growth. The experiments were done in triplicate. The

data was analyzed in RStudio (Free Software Foundation, Inc.) (see **Figure 3.6**).

## 2.5. Expression and purification of recombinant proteins

### 2.5.1. Expression of recombinant proteins

An expression vector, pET-28b(+) containing *Bacteroides thetaiotaomicron tet(X2)* (residues 11-388) was a generous gift from Dr. G. D. Wright (McMaster University, Canada). TetX2 was expressed and purified according to protocol modified from Yang et al. (2004) (Wangrong Yang, 2004) . *Tet(X2)* containing vector was transformed into *E. coli* BL21 (DE3) Star strain (Agilent Technologies, Incorporated, Santa Clara, California). Bacteria were cultured in an orbital shaking incubator at 37°C in 2 x YT media (16g/L tryptone, 10g/L yeast extract and 5g/L NaCl) supplemented with 50 µg/ml kanamycin in 5L flasks with 2L of media at 225 Revolutions Per Minute (RPM). Cultures were grown until cells reached optical density of ~0.8 at 600 nm when the temperature was reduced to 16°C and expression was induced with isopropyl β-D-1-thiogalactopyranoside to 0.5 mM final concentration. Bacteria were further incubated at 16°C overnight at 225 RPM. Cells were harvested by centrifugation at 8000 x g for 10 min, and the pellets were stored at -80°C. Frozen cells were thawed on ice and resuspended in approximately 150 mL of lysis buffer consisting of 20 mM Tris pH 8, 500 mM NaCl, 1 mM phenylmethanesulfonyl fluoride and 0.2 mM β–mercaptoethanol. Cells were lysed by sonication and the lysate was centrifuged at 15,000 × g for 30 min to remove the cell debris.

## 2.5.2. **Affinity chromatography**

The soluble fraction containing TetX2 was applied to a nickel chromatography column. The His•Bind® Resin (EMD Chemicals, San Diego, CA) was manually packed into a glass column and charged with 100 mM Nickel sulfate. Once the nickel charged resin settled, the column was washed with 5 column volumes (CV) of MiliQ water and equilibrated with at least 5 CVs of Buffer A (20mM Tris pH 8, 500mM NaCl, 1 mM phenylmethanesulfonyl fluoride and 0.2 mM $\beta$-mercaptoethanol). The protein was expressed with an N-terminal $His_6$-tag. The polyhistidine has an affinity for divalent metals such as Nickel so it enabled immobilization of a protein onto a column. $His_6$-tagged-TetX2 was captured on charged resin that was washed with 5 CV of Buffer A. The protein of interest was eluted with imidazole containing Buffer B (20mM Tris pH 8, 1M NaCl, 500mM imidazol, 1 mM phenylmethanesulfonyl fluoride and 0.2 mM $\mu$-mercaptoethanol). Fractions yellow in color were analyzed on a 15% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel (**Figure 2.2.A**). An equimolar amount of exogenous FAD was added to the fractions containing TetX2 after each purification step. Pooled fractions were dialysed against 20 mM Tris pH 8, 100 mM NaCl, 1 mM phenylmethanesulfonyl fluoride and 0.2 mM $\beta$-mercaptoethanol, overnight.

The N-terminal $His_6$-tag was removed by digestion of TetX2 protein sample with Thrombin for 24 h at 4°C (EMD Chemicals, San Diego, CA). The TetX2 digest was analyzed on a 15% SDS-PAGE gel and applied on a Chromatography HisTrap™ HP column (GE Healthcare). The flow through and wash fractions were analyzed on a 15% SDS-PAGE gel (**Figure 2.2.B**). Fractions with cleaved $His_6$-tag were pooled and dialyzed

against 20 mM Tris pH 8, 100 mM NaCl, 0.2 mM phenylmethanesulfonyl fluoride, 5 mM

mM β-mercaptoethanol, and 1 mM EDTA, overnight.
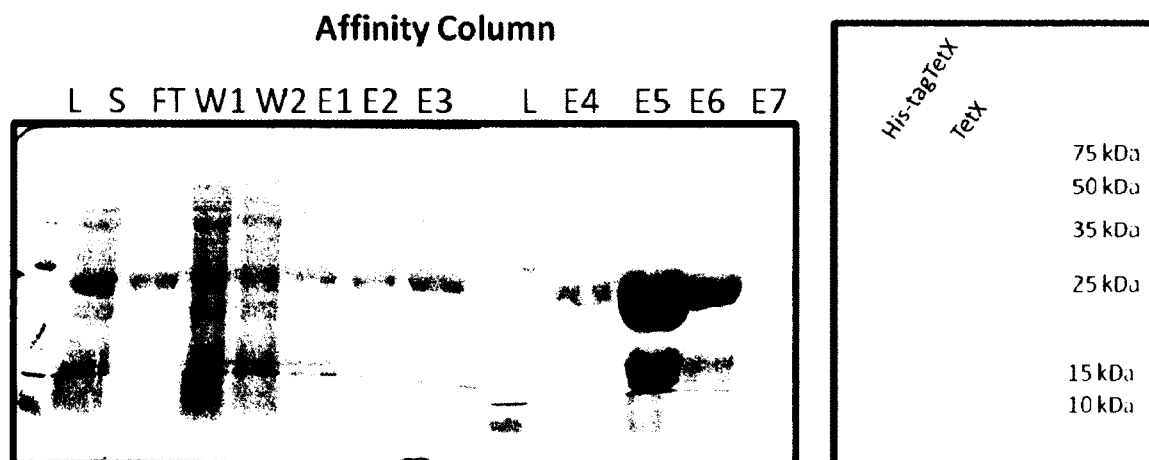


**Figure 2.2. Anion Exchange Chromatography of TetX2.**
A) Soluble fraction after cell lysis was collected and applied onto the Ni-NTA column, where His-tag-TetX2 was captured. B) The elute from this purification step (E4-E6) was then incubated overnight with thrombin to cleave the His-tag. Fractions were resolved on 15% SDS-PAGE gel.

### 2.5.3. Anion-exchange chromatography

Ion exchange chromatography is a common method used to separate molecules based on their net charge. The theoretical isoelectric point of TetX2 was calculated using EXPASY proteomics server and was about 5.4 suggesting that at neutral pH the protein is negatively charged. Anion exchange chromatography was performed on ÄKTA fast protein liquid chromatography (FPLC) unit and Unicorn™ software. The protein sample was removed from dialysis tubing and filtered through 0.22 μm syringe filter. The sample was then loaded onto a HiTrap Q Sepharose XL anion exchange chromatography

column. The column was pre-equilibrated with 5 CV of Buffer A (20 mM Tris, 100 mM NaCl, 0.2 mM phenylmethanesulfonyl fluoride, 5 mM $\beta$-mercaptoethanol, and 1 mM EDTA, pH 8). The negatively charged protein bound to the positively charged resin and was washed with 2 CV of low salt Buffer A. A linear gradient of 0-1 M NaCl over 20 CV was applied onto the column to elute TetX2. Using high salt containing negatively charged chloride outcompetes positively charged protein from the column. Fractions containing yellow protein were analyzed on 15% SDS-PAGE gel (**Figure 2.3.A**).

### 2.5.4. Size-exclusion chromatography

Size exclusion chromatography is often used in protein purification to separate molecules based on their mass and hydrodynamic shape. The limitations of this approach are that the shape of the molecule affects how fast it travels in the column as well as the fact that molecules to be separated have to be at least twice the size from each other to be effectively separated. Following anion exchange chromatography, fractions containing TetX2 were pooled, concentrated to 2 mL total volume and loaded onto HiLoad16/60 Superdex-200 column (GE Healthcare) on the AKTA FPLC system. TetX2 was eluted as a single peak. Finally, yellow fractions were analyzed on 15% SDS-PAGE gel, were pooled and concentrated to final concentration of 15 mg/ml (**Figure 2.3.B**). Protein samples were flash frozen in liquid nitrogen and were stored at -80°C. The other TetX2 mutants was expressed and purified as described above.
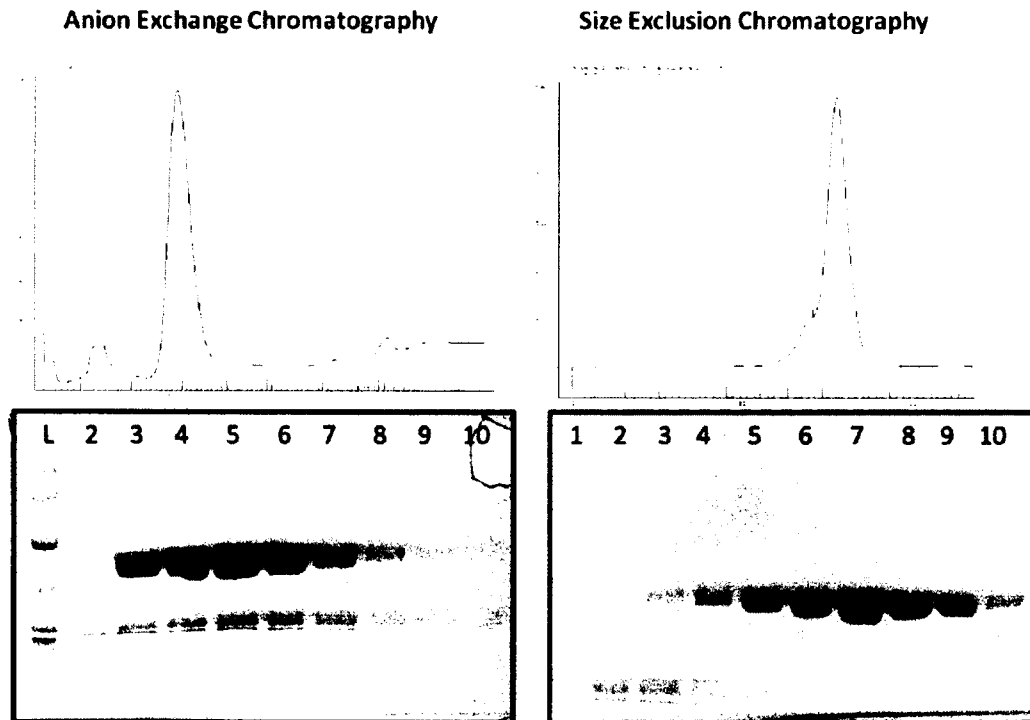
**Figure 2.3. Anion Exchange and Size Exclusion Chromatography of TetX2.**
A) After overnight dialysis His-tag cleaved protein sample was applied onto Q-XL Sepharose column. Fraction 2-10 corresponding to the largest peak on the chromatogram were resolved on 15% SDS-PAGE gel. B) Fractions 3-7 from anion-exchange chromatography were pooled, concentrated and applied onto Superdex-200 size exclusion column. Fractions 5-10 were pooled, concentrated to 15 mg/ml and flash frozen in liquid nitrogen for storage at -80°C.

### 2.5.5. Expression and purification of Seleno-methionine labeled TetX2

Seleno-methionine labeled TetX2 (SeMet-TetX2) was expressed according to the methionine inhibition pathway method (Doublie, 1997). The cell culture was grown in M9 Minimal media supplemented with 50 µg/ml of kanamycin. Prior to induction of protein expression 50 mg/L of L-selenomethionine, 100 mg/L of leucine, isoleucine, phenylalanine, and 50 mg/L of threonine, lysine, valine were added to the media. The expression was carried out in the shaking incubator in the dark, overnight at 16°C. The SeMet-TetX2 was purified following the same steps as described above for the native protein.

### 2.6. Measurement of kinetic parameters of TetX2 and its mutants at various minocycline concentrations

Steady-state kinetic parameters were measured using a previously described, but modified TetX2 activity assay (Wangrong Yang, 2004). Changes in absorbance corresponding to MCN hydroxylation by TetX2 was monitored at 400 nm ($\epsilon_{400}$ = 12000 $M^{-1} cm^{-1}$) over 5 minutes in a Infinite® M1000 plate reader (Tecan Group Ltd.) (**Figure 2.4**). In order to determine the kinetic mechanism of TetX2, the steady–state kinetics parameters were measured at various MCN (0-400 µM) and NADPH (12.5-200 µM) concentrations at 37°C. Each reaction included 5 mM $MgCl_2$, about 0.2 µM protein, 0-400 µM MCN and 12.5-200 µM NADPH in 20 mM Tris, pH 8.0. The assay was performed in triplicate. Steady-state kinetic parameters were determined by fitting initial reaction rates ($\upsilon_0$) to binary complex mechanism equation 1 (Eq 4.3) (**Figure 4.2**) where [A] = MCN concentration, and [B] is NADPH concentration, using Prism (GraphPad).
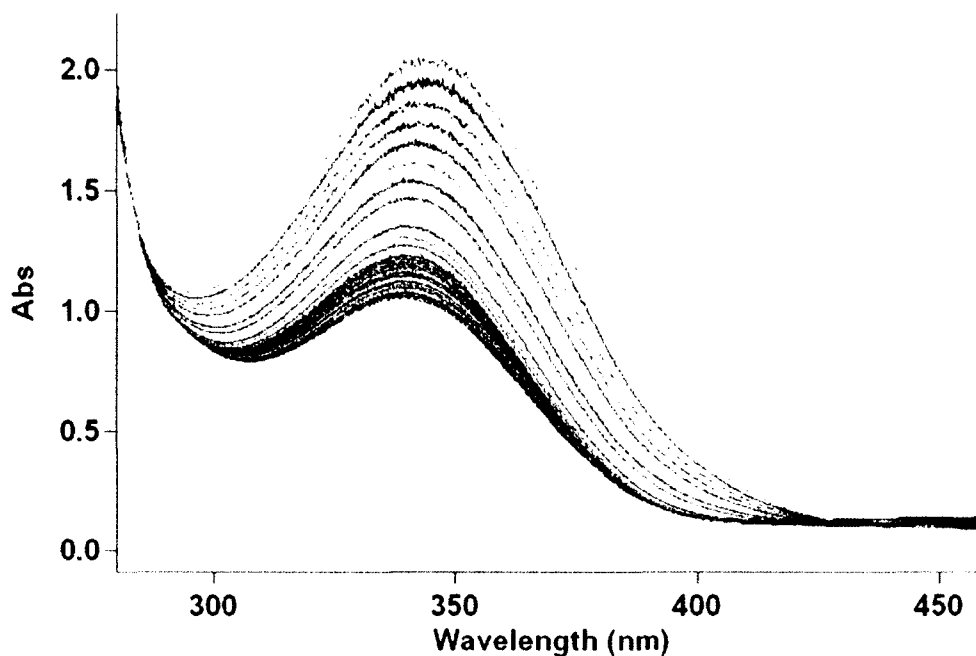
**Figure 2.4 Change in absorbance corresponding to minocycline degradation as a function of time.**
TetX2 actively degrades minocycline in the presence of $Mg^{2+}$ oxygen and NADPH over 20 minutes..

## 2.7. TetX2 activity from cell extracts

*In vivo* TetX2 mediated inactivation of MCN was estimated from cell extracts prepared from each of the *E. coli* BW25113 cell lines expressing chromosomal copies of wild-type and mutant *tetX2*. Background levels of MCN inactivation activity were determined from extracts of BW25113 without *tet(X2)*. Cells were grown in LB without MCN at 37°C and harvested by centrifugation at an OD ~0.6-0.8 prior to lysis by sonication. The steady-state TetX2 activity for inactivation of MCN by the soluble crude extracts was assayed using the previously described *in vitro* assay. TetX2 activity was

measured at 25, 50 200 μM MCN at 200 μM of NAPDH. All measurements were done

in triplicate. The total amount of protein in each cell lysate was estimated by Bradford

assay to account for variability in the efficiency of cell lysis. Relative enzyme

concentrations were calculated by fitting initial velocity measured at 200 μM MCN using

binary complex equation (to incorporate $K_{M(MCN)}$ and $K_{M(NADPH)}$, since the measurements

were done under non-saturating conditions for some enzymes) (**Eq. 4.1**) (**Figure 4.4**),

where $V_{max}=k_{cat}[E_0]$, $E_0$=total enzyme concentration.

## 2.8. Crystallization, Structure Determination and Refinement

### 2.8.1. Crystallization and Data Collection

The initial crystallization conditions for TetX2 were determined using the Hydra

II Plus One robot and commercially available screens (Qiagen, Hampton) by Dr. Gang

Wu. Further optimization of the crystallization conditions was carried by hand using a

hanging and sitting drop vapor diffusion methods starting with conditions determined by

the automated screening, which included 2.1 M Ammonium Sulfate, 2% (v/v)

methylpentanediol, 0.1 M CHES pH 9.0 at 20 °C. Initially, only small crystals of TetX2

were obtained, but by varying the precipitant concentrations, pH, and lowering the

incubation temperature to 10 °C bigger crystal were grown within 3-7 days. These

crystals had a tendency to form thin rods clustered in the center, however, through

addition of additives such as ethanol, di-oxane, and glycerol, single long rod-like crystals

were obtained under 1.8 M Ammonium Sulfate, 0.5% (v/v) ethanol or glycerol, 0.1 M

CHES pH 9.2 at 10 °C (**Figure 2.5**). SeMet-TetX2 crystals were obtained under similar crystallization conditions using the sitting-drop vapor diffusion method by Dr. Milya Davlieva.



**Figure 2.5. Optimization of TetX2 crystallization conditions.**
A) Initially, TetX2 crystals grew as tiny clusters. B) Lowering temperature resulted in larger rod-like crystal C) Additives, such as glycerol produced single, well shaped TetX2 crystals.

Diffraction data for the native TetX2 crystals were collected at our home source using a Rigaku R-AXIS IV$^{++}$ detector (**Figure 2.6**). Multiple-wavelength anomalous dispersion (MAD) data were collected by Jay Nix at Advanced Light Source (ALS) beamline 4.2.2 using a NOIR-1 MBC detector. The native data were processed using d*TREK (Pflugrath, 1999) and the data collected on the SeMet-TetX2 crystals were processed using HKL2000 (Otwinowski, 1997) (Table 2.1). Native and SeMet-TetX2 crystals diffracted to 2.5 Å and 2.8 Å, respectively. Both proteins crystallized in space group $P2_1$ with cell dimensions a = 87.65 Å, b = 67.41 Å, c = 152.35 Å and angles $\alpha = \beta$ = 90.0° and $\gamma$ = 101.68°.
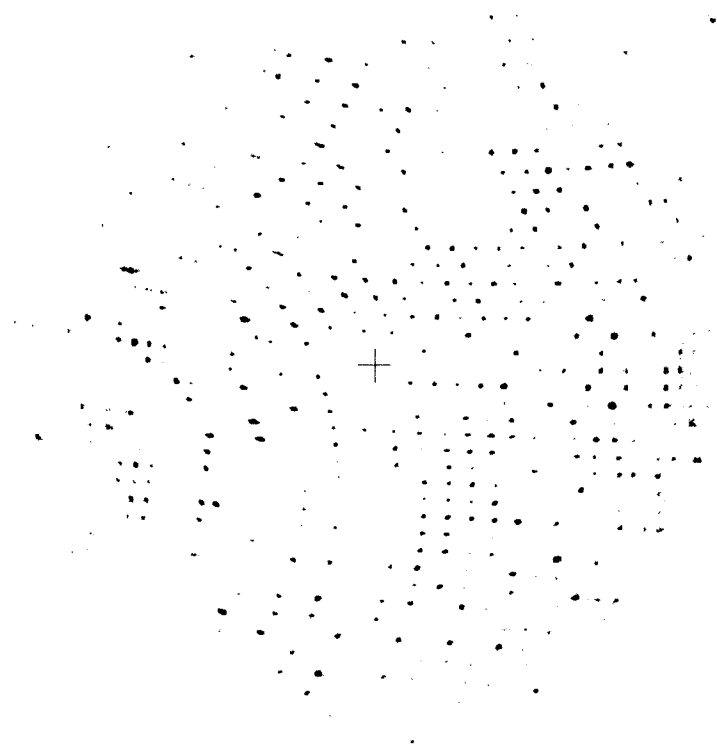
**Figure 2.6. Diffraction image of TetX2 native crystal collected on a home source X-ray generator at Rice University with high resolution spots diffracting to 2.8 Å.**

Structure factor data quality was analyzed using Phenix.Xtriage which indicated presence of an additional high intensity peak away from the origin (Adams et al., 2010). The summary of output is shown in **Figure 2.7**. A Native Patterson map was generated using xxft program from XtalView suite (McRee, 1999) . A significant off origin peak with intensity of 51.56% of the original peak was detected at position 0.5, 0.05 0.5 **(Figure 2.8)**. The significance of the peak is determined by the p-value which determines the probability of a peak of this intensity to occur by chance. The p-value for the native data set was very low (5.357E-05), indicating that the chance for the peak to occur randomly was very low. In addition, the intensity of the off-origin peak being half of the origin peak strongly implied that the non-crystallographic symmetry was pseudo

translational. Translational and pseudo-translational symmetry can often complicate the structure solution using either molecular replacement or multi-wavelength anomalous dispersion. This may explain why initial attempts in solving the structure by MR failed using Phaser. BALBES, which is known to handle pseudo-translational symmetry better than Phaser was used for structure solution (Long et al., 2008).

```
Twinning and intensity statistics summary (acentric data):

Statistics independent of twin laws
  - <I^2>/<I>^2 : 2.862
  - <F>^2/<F^2> : 0.688
  - <|E^2-1|>   : 0.895
  - <|L|>, <L^2>: 0.524, 0.359
      Multivariate Z score L-test: 3.405
      The multivariate Z score is a quality measure of the given
      spread in intensities. Good to reasonable data are expected
      to have a Z score lower than 3.5.
      Large values can indicate twinning, but small values do not
      necessarily exclude it.

No (pseudo)merohedral twin laws were found.

Patterson analyses
  - Largest peak height  : 51.761
   (corresponding p value : 5.357e-05)

The analyses of the Patterson function reveals a significant off-origin
peak that is 51.76 % of the origin peak, indicating pseudo translational symmetry.
The chance of finding a peak of this or larger height by random in a
structure without pseudo translational symmetry is equal to the 5.3569e-05.
The detected tranlational NCS is most likely also responsible for the elevated intensity ratio.
See the relevant section of the logfile for more details.
The results of the L-test indicate that the intensity statistics
behave as expected. No twinning is suspected.
----------------------------------------------------------------------
```

**Figure 2.7. Summary of data quality analysis for native P21 crystals by Phenix.Xtrage.**
Data analysis using Phenix.Xtriage showed that there is no pseudo merohedral twinning. A significant off-origin peak was detected at height corresponding to 51.76% of the origin peak indicating presence of pseudo translational symmetry.
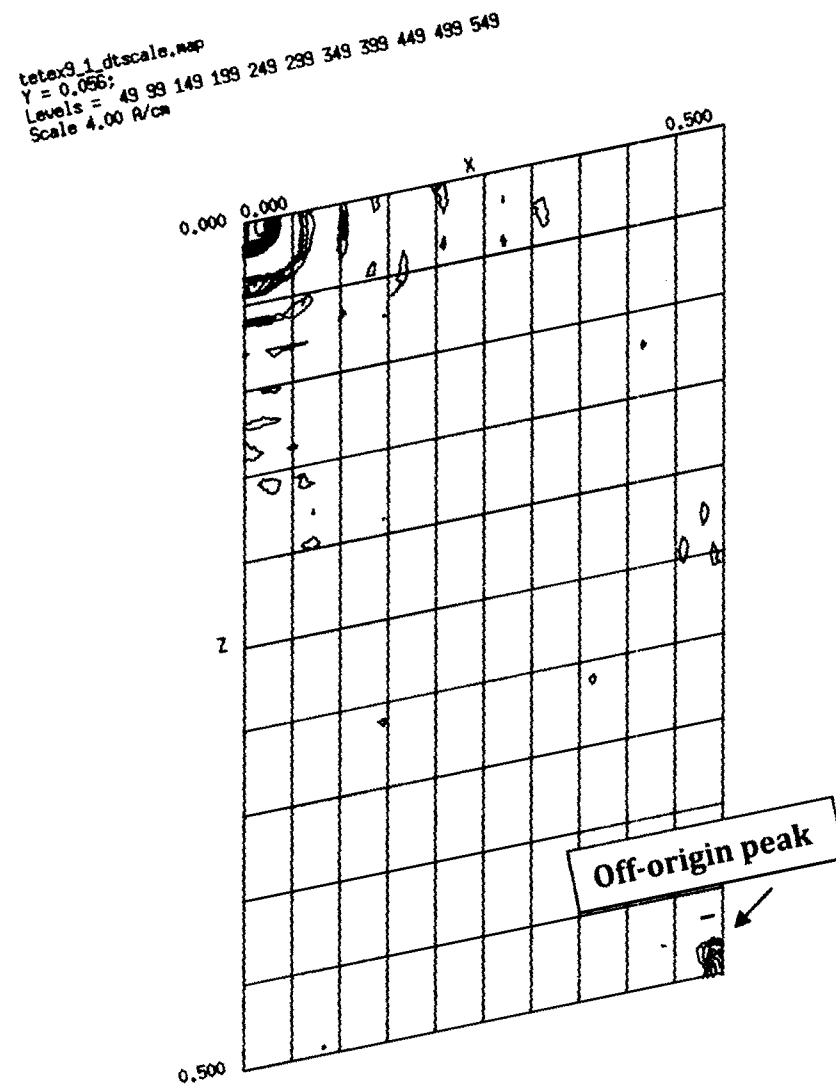
**Figure 2.8. Native Patterson map generated by in CCP4 suite (Winn et al. 2011).** Native Patterson map was generated to investigate presence of non-crystallographic symmetry which was indicated by Phenix.Xtriage. An off-origin peak was present at position 0.5, 0.05 and 0.5, which was half intensity of the origin peak, indicating presence of pseudo-translational symmetry.

## 2.8.2. Native TetX2 Structure Determination and Refinement

The initial structure determination by molecular replacement (MR) was performed using BALBES (Long et al., 2008). The best solution was obtained when PhzS from *Psedomonas auerugosa* (PDBID:3c96) was used as a search model (21% sequence identity). The solution from the MR model suggested four molecules in the asymmetric unit. The arrangement of the molecules in the asymmetric unit is shown in **Figure 2.9.** The asymmetric unit consists of two pairs of dimers related by 2-fold non-crystallographic symmetry (AC:BD). The matrices describing how to go from one dimer to the other is shown below:

Rotation:
r={-0.99986, 0.01632, -0.00239}
{ 0.01630, 0.99985, 0.00607}
{ 0.00248, 0.00603, -0.99998}
Translation:
t={31.00092}, {28.49513}, {75.90825}

The initial model was submitted to phenix.autobuild and phenix.refine for automatic building and structure refinement (Adams et al. 2010). Diffraction data collected at the peak wavelength 0.97889 Å were used for structure determination by Single-wavelength anomalous dispersion (SAD) using phenix.autosol (Adams et al. 2010). The partial model obtained by MR was used to search for Se sites. Automated model building by phenix.autobuild (Adams et al. 2010) resulted in successful placement of ~55-60% of the protein structure with the initial R-factor = 45% and R-free = 47.5%. The model was further built manually in COOT (Emsley and Cowtan, 2004) and refined by phenix.refine. The initial refinement strategy included rigid-body, positional and group ADP refinement with non-crystallographic symmetry (NCS) restraints. FAD was

fit manually into an unoccupied density found in all 4 molecules. Additionally, TLS refinement was carried out in phenix.refine (Adams et al. 2010). An anisotropy correction applied to SAD data set resulted in a 2% drop in R-factor and R-free after refinement (Strong et al., 2006). Data collection summary and refinement statistics are summarized in **Table 2.1**. Structure coordinates were deposited in the Protein Data Bank under accession number 3P9U.
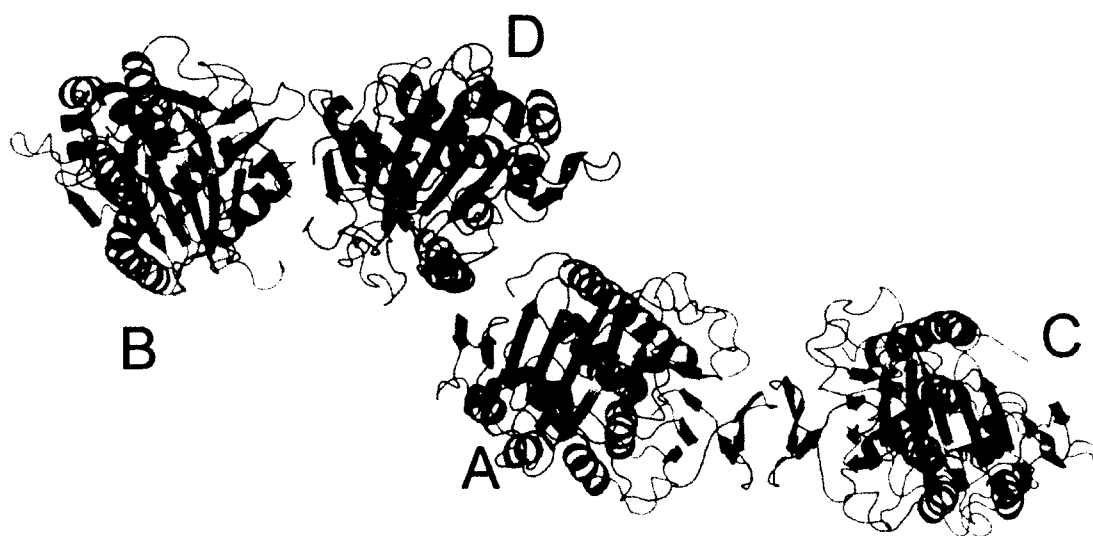


**Figure 2.9. Arrangement of TetX2 molecules in the asymmetric unit.**
The asymmetric unit is composed of two pairs of dimers (AC-BD) related by 2-fold non-crystallographic symmetry. The rotation and translation matrices to go from one dimer to the other are shown above.

**Table 2.1. Summary of Data Collection and Refinement Statistics for TetX2**

| Data Collection | SeMet TetX2 | |
|---|---|---|
| Wavelength (Å) | 0.97889 | |
| Resolution (Å)[a] | 50 – 2.80 (2.85 - 2.80) | |
| Space group | $P2_1$ | |
| Unit Cell (Å) | a = 87.70, b = 67.33, c = 153.79 | |
| | $\alpha = \gamma = 90.0°$, $\beta = 100.31°$ | |
| Total number of reflections[a] | 132098 | |
| Unique reflections | 39866 | |
| Average redundancy[a] | 3.3 | (3.2) |
| Completeness (%)[a] | 92.0 | (86.7) |
| $R_{merge}$ (%)[a,b] | 11.8 | (28.6) |
| Output <I/sigI>[a] | 16 | (3.9) |

| Refinement | | |
|---|---|---|
| $R_{work}$ (%)[c] | 24.00 | |
| $R_{free}$ (%)[d] | 28.24 | |
| r.m.s.d.[e] from ideality | | |
| Bonds (Å) | 0.009 | |
| Angles (°) | 1.255 | |
| Average B-factor (Å2) | 25.05 | |
| Ramachandran[f] | | |
| favored (%) | 92.1 | |
| additional allowed (%) | 7.9 | |
| disallowed (%) | 0 | |
| PDB accession number | 3P9U | |

[a]Values for the last shell are in parentheses.

[b]$R_{merge}$ $\sum |I - <I>| / \sum I$, where I is measured intensity for reflections with indices of hkl.

[c]$R_{work}$ $\sum |F_o - F_c| / \sum |F_o|$ for all data with $F_o > 2 \sigma (F_o)$ excluding data to calculate $R_{free}$.

[d]$R_{free}$ $\sum |F_o - F_c| / \sum |F_o|$ for all data with $F_o > 2 \sigma (F_o)$ excluded from refinement.

[e]Root mean square deviation.

[f]Calculated by using MolProbity

### 2.8.3. Structure Determination and Refinement of TetX2 in Complex with Minocycline and Tigecycline

Crystallization of TetX2$_{T280A}$ was performed under similar conditions to the wild-type TetX2 described before using vapor diffusion sitting-drop method (Walkiewicz et al., 2011). Native crystals were soaked in various MCN and Tigecycline concentrations for 16 hours prior to cryoprotection. Crystals were cryoprotected in 25% (v/v) glycerol and immediately flash frozen in liquid nitrogen. Single wavelength data were collected by Jay Nixat the Advanced Light Source beamline 4.2.2 using a NOIR 1 MBC detector.

#### 2.8.3.1. TetX2:MCN complex:

The diffraction data for TetX2:MCN complex crystals were processed using HKL2000 software in space group P1 to 2.7Å resolution (Otwinowski, 1997). The initial structure determination by molecular replacement (MR) was performed using phenix.autosol (Adams et al., 2010) with wild-type TetX2 structure, omitting FAD cofactor as a search model [Protein Data Bank (PDB) ID: 3P9U]. Based on the initial MR solution, four molecules were positioned in the asymmetric unit. The initial model was submitted for automated building and refinement to phenix.autobuild and phenix.refine (Adams et al., 2010), respectively. Automated model building and refinement resulted in $R_{work}$=24.1% and $R_{free}$=30.1%. The electron density for flavin adenine dinucleotide (FAD) was visible in all four molecules in the asymmetric unit. FAD was built manually in COOT (Emsley and Cowtan, 2004) and was used in the refinement of the structure. In addition, unoccupied electron density corresponding to MCN molecules was identified near the FAD isoalloxazine ring in the composite omit map generated in PHENIX

(Adams et al., 2010). MCN molecules were fitted manually in COOT and were included in final stages of refinement. Data collection and refinement statistics are shown in **Table 2.2.** Structure coordinates were deposited in the PDB under accession number 3V3N.

### 2.8.3.2. TetX2:tigecycline complex

The strategy used for structure solution of TetX2:tigecycline complex closely resembled the one followed for TetX2:MCN complex. The diffraction data were processed using HKL2000 software in space group P1 to 2.9Å resolution (Otwinowski, 1997). The initial structure determination by MR was performed using phenix.autosol (Adams et al., 2010) with the same initial search model [Protein Data Bank (PDB) ID: 3P9U]. Similarly, four molecules were positioned in the asymmetric unit and the initial model was submitted for automated building and refinement. The monomers in the asymmetric unit are related by 2-fold non-crystallographic symmetry axis. Automated model building and refinement resulted in $R_{work}=26.3\%$ and $R_{free}=33.7\%$. The electron density for flavin adenine dinucleotide (FAD) was visible in all four molecules in the asymmetric unit. FAD was built manually in COOT (Emsley and Cowtan, 2004) and was used in the refinement of the structure. The electron density for tigecycline was visible in copies A and C and was weaker in copies D and B based on the (2Fo-Fc) SIGMAA weighted composite omit map. The electron density for the aliphatic chain of tigecycline was visible only in copy A. Tigecycline molecules were fitted manually in COOT and were included in final stages of refinement. Data collection and refinement statistics are shown in **Table 2.2.** Structure coordinates were deposited in the PDB under accession number 3V3O.

**Table 2.2. Summary of Data Collection and Refinement Statistics**

| Data Collection | TetX2$_{T280A}$ : MCN | | TetX2$_{T280A}$ : TIG | |
|---|---|---|---|---|
| Wavelength (Å) | 1.00004 | | 1.00004 | |
| Resolution (Å)[a] | 49.3 – 2.70 (2.70 - 2.60) | | 48.94 – 2.90 (2.98 - 2.90) | |
| Space group | P1 | | P1 | |
| Unit Cell (Å) | a = 68.43, b = 80.10, c = 87.36 α = 111.11° β =90.10° γ = 93.01° | | a = 68.13, b = 80.08, c = 87.24 α = 111.02° β =89.97° γ = 92.97° | |
| Total number of reflections[a] | 72314 | | 72992 | |
| Unique reflections | 41120 | | 40388 | |
| Average redundancy[a] | 1.7 | (1.5) | 1.8 | (1.7) |
| R$_{merge}$ (%)[a,b] | 7.4 | (31.7) | 10.2 | (35.8) |
| Output <I/sigI>[a] | 16 .1 | (2.0) | 14. 6 | (1.96) |
| **Refinement** | | | | |
| R$_{work}$ (%)[c] | 22.08 | | 23.73 | |
| R$_{free}$ (%)[d] | 27.57 | | 29.00 | |
| r.m.s.d.[e] from ideality | | | | |
| Bonds (Å) | 0.009 | | 0.01 | |
| Angles (°) | 1.235 | | 1.46 | |
| Average B-factor (Å2) | 50.60 | | 68.78 | |
| Ramachandran[f] | | | | |
| favored (%) | 93.47 | | 95.5 | |
| outliers (%) | 0.28 | | 0 | |
| PDB accession number | 3V3N | | 3V30 | |

[a]Values for the last shell are in parentheses.
[b]R$_{merge}$ $\sum$ |I – <I>| / $\sum$ I, where I is measured intensity for reflections with indices of hkl.
[c]R$_{work}$ $\sum$ |F$_o$ - F$_c$| / $\sum$ |F$_o$| for all data with F$_o$ > 2 σ (F$_o$) excluding data to calculate R$_{free}$.
[d]R$_{free}$ $\sum$ |F$_o$ - F$_c$| / $\sum$ |F$_o$| for all data with F$_o$ > 2 σ (F$_o$) excluded from refinement.
[e]Root mean square deviation.
[f]Calculated by using MolProbity

## 2.9. Construction of directed evolution mutant library

The single point mutational library was generated using a commercially available GeneMorphII Random Mutagenesis Kit® (Agilent Technologies). The mutational rate was adjusted by varying the template concentrations in the PCR. The target mutational rate for the library of 1-2 single point mutations per gene was achieved when 20 ng of a template (*tet(X2)*/pUC19) was used in each reaction (**Figure 2.10**). Mutated *tetX2* PCR products were ligated into a pUC19 backbone, behind a constitutively active lacZ promoter. The ligation products were transformed into *E. coli* DH10B cells and plated on 100 μg/ml of ampicillin to estimate the library size and mutational rate. Screening for active mutants was performed on LB-agar plates containing 4 μg/ml of MCN. Plasmid DNA from 35 colonies was isolated and sent for sequencing of *tetX2* gene.
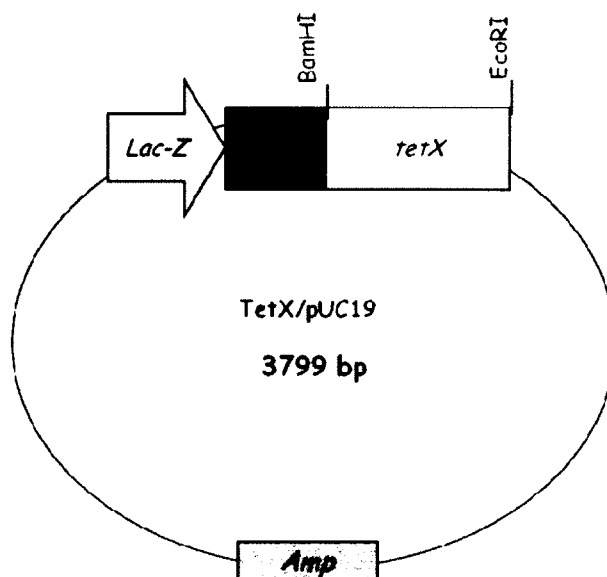


**Figure 2.10. Architecture of tetX2/pUC19 vector.**
The library mutants of tetX2 were cloned into a pUC19 backbone behind a constrictive promoter *lacZ* between restriction sites BamHI and EcoRI.

## 2.10. **Circular Dichroism (CD)**

CD experiments were performed using a Jasco J-815 spectrophotometer. Spectra were recorded at 200 - 250 nm at 20°C. The thermal stability of TetX2 and the mutant TetX2$_{T280A}$ was determined by monitoring absorption at 220 nm at the temperature range 20°C - 90°C using a scanning rate of 60°/hour in triplicate. Protein samples were prepared to the concentration of 20 µM in 20 mM Tris, pH 8.0. The thermal unfolding midpoint ($T_m$) could only be approximated as protein denaturation was largely irreversible.

# Chapter 3 Adaptation of TetX2

The objective of this project is to discover the biophysical principles that govern adaptation of naturally evolved *Escherichia coli* strains to tetracycline antibiotics and develop a model that can predict the successful evolutionary trajectories to TetX2 mediated resistance. We combined experimental evolution with *in vitro* error-prone mutagenesis to explore all potential single point mutations of TetX2 that could be candidates for natural selection and identify which physicochemical properties are most correlated to fitness in a population.

## 3.1. Experimental evolution of *E. coli* expressing a chromosomal copy of *tet(X2)*

Wild-type *tet(X2)* was chromosomally integrated into the *spc* operon of *E. coli* BW25113 between *prlA* (SecY) and *rpmJ* (L36) by recombineering, to produce a strain (BW25113$_{tet(X2)}$) carrying a single copy of *tet(X2)* under expression of the endogenous *spc* promoter. Chromosomal integration of *tet(X2)* eliminated concerns of variable gene copy numbers that can complicate interpretation of results. Confirmation PCR was performed using genomic DNA isolated from colonies recovered from tetracycline (10 μg/ml) LB-agar plates and confirmed by DNA sequencing. The minimum inhibitory concentrations (MIC) values towards MCN for the wild-type BW25113 strain and BW25113$_{tet(X2)}$ were measured to be at 2 and 16 μg/ml, respectively. Based on the MIC values, 10 μg/ml of MCN, a concentration that allowed growth of *E. coli* BW25513$_{tet(X2)}$

during a twelve hour period was set as a starting condition for the serial-passage experiment.

Serial-passage of *E. coli* (BW25113tet(X2)) grown in shaking flasks to gradually increasing MCN concentrations from 10 to 320 μg/ml was performed over 10 days at 37°C in LB (**Figure 3.1**).
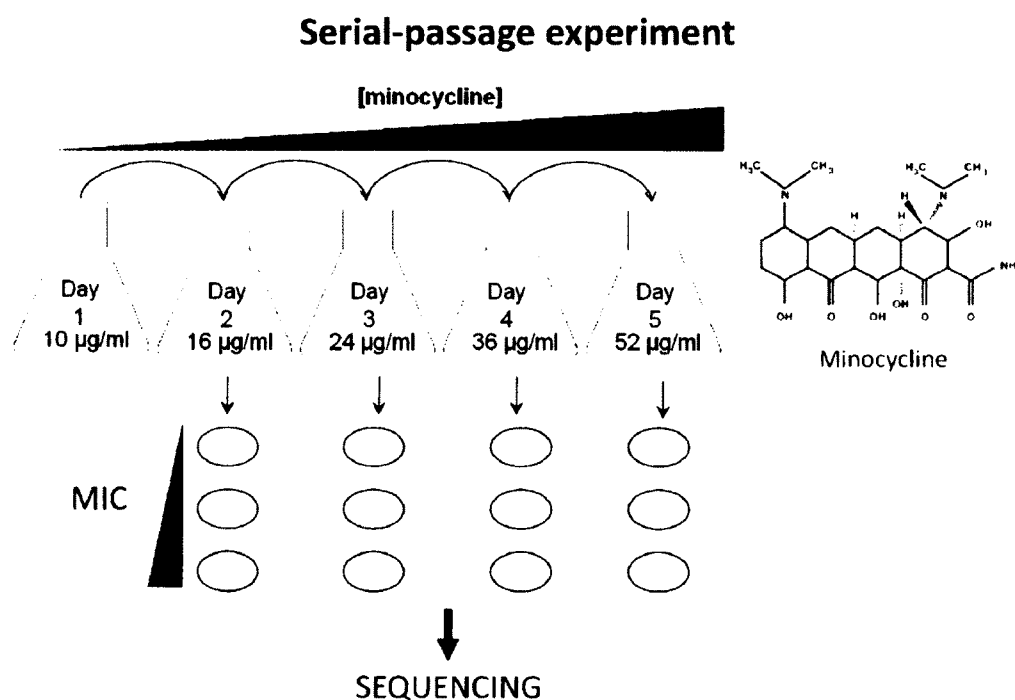


**Figure 3.1. Adaptation to minocycline experiment.**
*E. coli* with tetracycline resistant gene (*tetX2*) were subjected to increasing antibiotic concentrations (10-320 μg/ml) over 10 days. Each day, 50 μl of cells were transferred to fresh media with higher minocycline concentrations. Samples from each day were pooled and stored at -80°C. The *tet(X2)* gene was PCR amplified from colonies showing increased antibiotic resistance and the *tet(X2)* gene was sequenced.

DNA sequencing analysis of the *tet(X2)* allele from colonies isolated on different days of the experiment identified a single missense mutation (Thr to Ala) at position 280 on day 2, when MCN concentration in the media was 16 µg/ml (32 µM).

In order to determine the effect of Thr to Ala substitution in TetX2 on the overall resistance of the organism, this mutation was reintroduced chromosomally into the naïve BW25113-*tet(X2)* strain that had not been exposed to antibiotics and the MIC remeasured at 37°C (**Figure 2.1**). The unadapted strain with *tet(X2)*$_{T280A}$ (BW25113-*tet(X2)*$_{T280A}$) showed the expected increase in MIC to MCN at 37°C from 16 to 64 µg/ml (**Table 3.1**, **Figure 3.2**).

This increase in MIC is in good agreement with the selection conditions used on day 2 of the serial passage experiment. The MICs toward tetracycline and tigecycline were also determined for BW25113-*tet(X2)*$_{T280A}$ and showed a comparable increase in resistance towards both antibiotics (**Table 3.1, Figure 3.3**) although neither drug was included during the serial-passage suggesting general expansion in the range of tetracyclines.

**Table 3.1. Minimum Inhibitory Concentration (MIC)[a] of strains isolated during experimental evolution to tetracycline antibiotics[b].**

| | BW25113 | BW25113 *tet(X2)* | BW25113 *tetX2*$_{T280A}$ | BW25113 *tetX2*$_{T280A}$ Day 9 |
|---|---|---|---|---|
| Tetracycline | 4 | 128 | >128 | >128 |
| Minocycline | 4 | 16 | 64 | >128 |
| Tigecycline | 1 | 8 | 16 | >128 |

[a] MIC determined by agar-dilution method according to NCCLS guidelines
[b] Drug concentrations in µg/ml
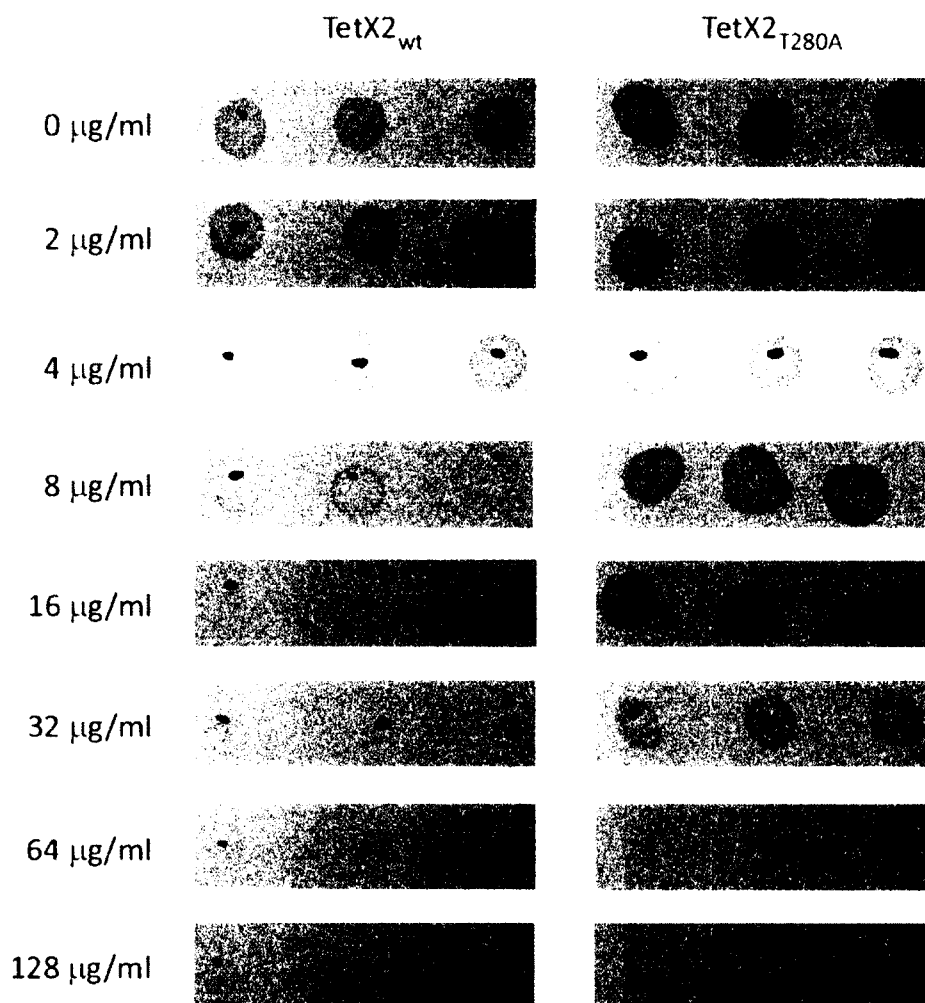
**Figure 3.2. Adaptive mutant of *tet(X2)*$_{T280A}$ isolated through experimental evolution has increased resistance to MCN by 4-fold at 37°C.**

MICs to MCN were measured by agar dilution test (see section 2.3 for methods) for three individual colonies from *E. coli* (BW25113) strains expressing either *tet(X2)* or *tet(X2)*$_{T280A}$. Growth of *E. coli* BW25113-*tet(X2)* was inhibited at 16 mg/ml and *E. coli* BW25113-*tet(X2)*$_{T280A}$ showed no growth only at 64 mg/ml, suggesting that the T280A mutation alone is sufficient for a 4-fold increase in the MIC for MCN.

TetX2<sub>wild-type</sub>   TetX2<sub>T280A</sub>

Minocycline
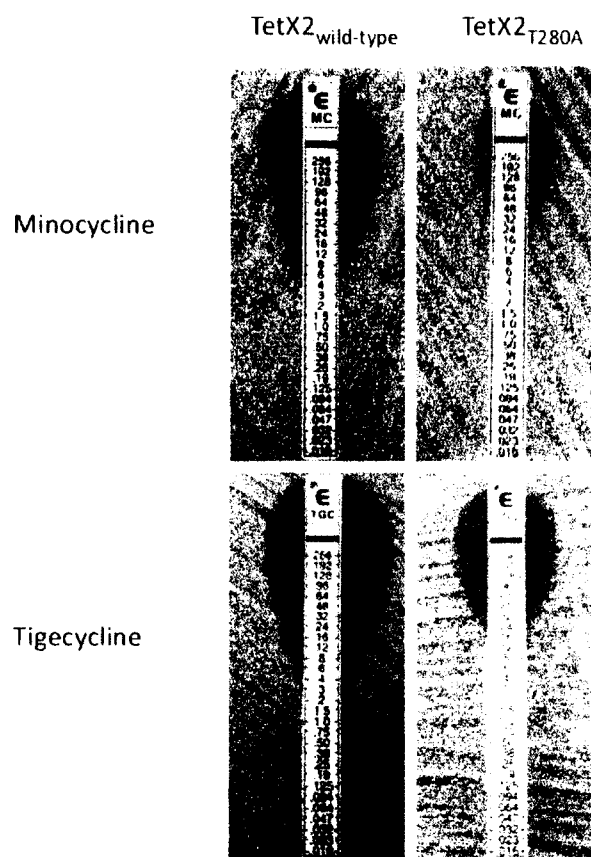
Tigecycline

**Figure 3.3. Adaptive mutant of *tet(X2)*<sub>T280A</sub> has increased resistance to MCN and tigecycline.**
The MICs measured by E-test for *E. coli* strains expressing *tet(X2)* or *tet(X2)*<sub>T280A</sub> show an increase in resistance (smaller inhibition zone) for both MCN (top panel) and tigecycline (lower panel), even though the cells were not exposed to tigecycline previously. This suggests that the adaptive mutation has broadened the specificity of the enzyme (see section 2.3 for methods).

T280A was the only mutation in *tet(X2)* identified during the experiment, although, resistance to MCN continued to increase beyond 320 μg/ml. These results suggest that an approximately 20 fold increase in resistance is a result of both changes within the *tet(X2)* gene as well as the *E. coli* chromosome.

In addition, we tested, whether *E. coli* BW25113 alone (without *tetX2*) could adapt to higher MCN concentrations. We determined that *E. coli* BW25113 could not be

readily adapted beyond an MIC of 4 µg/ml to MCN without having either a

chromosomally expressed copy of *tet(X2)*, a multicopy plasmid expressing *tet(X2)* or

another tetracycline resistance gene such as TetA on pBR322 (**Figure 3.4**) These findings

are consistent with other studies suggesting a more limited ability for many *E. coli* strains

to adapt to tetracyclines in the absence of horizontal gene transfer (Viveiros et al., 2007;

Zhang et al., 2008). Our results suggest that *tet(X2)* expression plays an important role in

facilitating adaptation to very high MICs. It is unclear how *tet(X2)* may mediate or

interact with these later chromosomal changes to raise resistance. The inability of the

strain without *tet(X2)* to adapt suggest that there is a substantial epistasis between *tet(X2)*

and the chromosomal changes.
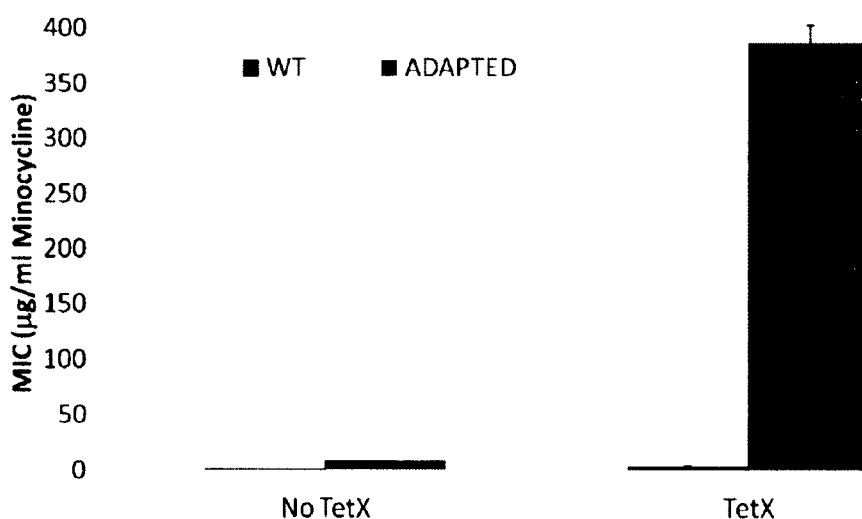


**Figure 3.4. Adaptation of *E. coli* BW25113 to high MCN concentration is highly dependent on expression of *tet(X2)*.**

The MIC of *E. coli* expressing *tet(X2)* increases from 16 µg/ml to < 320 µg/ml over the course of 10 days. *E. coli* without *tet(X2)* have an MIC of 4 µg/ml and don't have the ability to adapt to MCN unless *tet(X2)* or other tetracycline resistance determinant is present.

Serial passage of the highly resistant strain for 6 days in the absence of antibiotic showed no decrease in MIC, consistent with genomic changes rather than transient changes attributed to gene expression, suggesting an important role for chromosomal changes outside *tet(X2)* in adaptation.

## 3.2. *In vitro* evolution of TetX2

TetX2$_{T280A}$, a mutant with increased resistance to MCN, was the only TetX2 variant identified by experimental evolution. Based on these results, we asked whether T280A was the only single-point mutation of TetX2 that had a potential for success under selection towards MCN, or were there other single-point mutations that confer equal or greater resistance than the wild-type enzyme? To answer this question, we used *in vitro* error prone mutagenesis to construct a library of all the adaptive TetX2 single nucleotide polymorphisms (SNPs) that could serve as likely candidates for natural selection in a bacterial population undergoing adaptation to increasing MCN concentrations.

### 3.2.1. Characterization of the error-prone library

In order to generate a library with mutational rate between 1-2 single point mutations per gene, the template (TetX2/pUC19) concentration was adjusted to 20 ng for each round of error prone PCR. Prior to selection, we estimated that in order to sample 95% of all possible single-point mutants, a library of 10,000-15,000 transformants had to be made and subjected to selection at drug concentration at which *tet(X2)*$_{wild-type}$ formed colonies within 24 hours. To estimate library size, the library was initially plated

on ampicillin. Based on the number of colonies counted ($2 \times 10^6$ cfu/ml), we estimated that this library had an eight-fold coverage of the minimum requirement. We felt confident that all possible single mutations were most likely to be sampled in the experiment. Results obtained from sequencing data based on plasmid DNA isolated from colonies selected for on ampicillin plates determined the mutational rate of the library to be 1.37 mutations/kB (see section 2.9). The distribution of mutations in the library is shown in **Figure 3.5**.
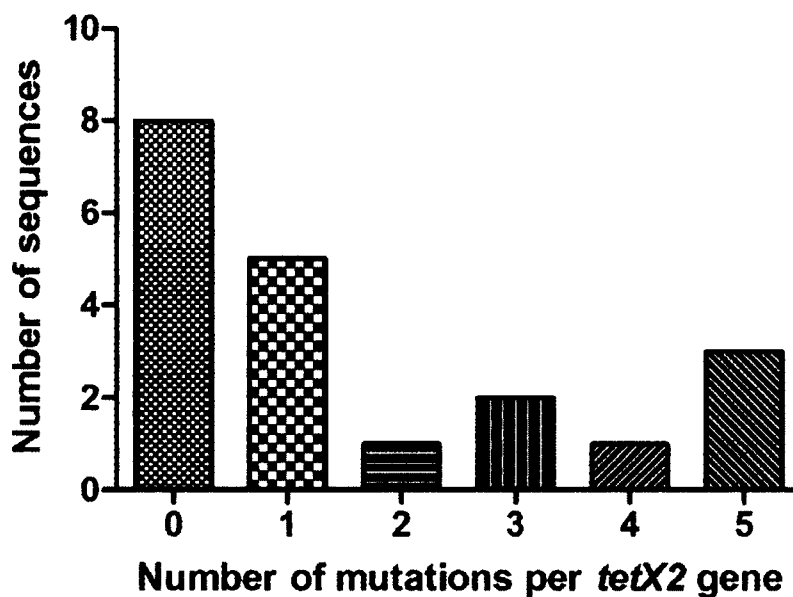


**Figure 3.5 Distribution of mutations in the error-prone library of *tet(X2)* mutants.** The average mutation rate of *tet*(X2) library was estimated at 1.37 mutations per kb. Most sequences contained zero or one mutation in *tet(X2)* gene.

### 3.2.2. Identification of functional TetX2 mutants

Selection for functional mutants expressed from a plasmid in a sensitive *E. coli* strain was performed on LB-agar plates containing 4 µg/ml of MCN, corresponding to the concentration of the drug at which both wild-type *tet(X2)* and *tet(X2)*$_{T280A}$ are able to sustain growth over 24 hours at 37°C.

Seven adaptive mutants were identified that had MIC to MCN equal to, or greater than, *tet(X2)*$_{wild-type}$ (**Table 3.2**) and included *tet(X2)*$_{T280A}$, the mutant identified previously by experimental evolution. Two out of the seven mutations, (T280A and K64R) arose from an A to G transition, while the remaining five mutations were the result of transversions. In addition, two sets of the mutations occurred at the same position in the protein sequence and included pairs T280A and T280S, as well as N371T and N371I.

**Table 3.2. Mutations recovered from error prone PCR generated library**

| Mutation site | DNA | MIC$_{MCN}$[a] | Transition/ Transversion |
|---|---|---|---|
| **WT** | N/A | 16 | N/A |
| **K64R** | AAA - AGA | 24 | Transition |
| **F235Y** | TTT - TAT | 32 | Transversion |
| **T280A** | ACG - GCG | 64 | Transition |
| **T280S** | ACG - TCG | 16 | Transversion |
| **S326I** | AGT - ATT | 32 | Transversion |
| **N371T** | AAC - ACC | 64 | Transversion |
| **N371I** | AAC - ATC | 64 | Transversion |

[a] MIC values were determined by E-test and are reported in µg/ml of MCN.

## 3.3. Fitness of adaptive mutants based on *in vivo* growth rates as a function of MCN concentration

Success of newly arising mutations in an evolving population strongly depends on the fitness benefit conferred by the mutations under selective environment. Because adaptive mutations of TetX2 have different effects on bacterial fitness when challenged with a range of MCN concentrations, MIC tests are not appropriate to assess their fitness *in vivo*. For example, the ability to form a colony after 24 hours is a poor quantitative measure of growth rate. Likewise, growth overnight in a liquid culture is also a very poor metric. Here, we use growth rate assays as a direct measurement of fitness because they provide more accurate link between physicochemical properties of TetX2 and bacterial growth than MICs.

Each mutant allele was introduced into the chromosome by recombineering performed following a protocol described for the generation of *E.coli* BW25113-*tet(X2)*<sub>wild-type</sub> strain. The growth rates were measured for all eight *E. coli* strains, each with a chromosomal copy of a different *tet(X2)* allele, over a range of MCN concentrations (**Figure 3.6, Table 3.4, section 2.4**) that correspond to FDA guidelines for the susceptibility testing of *Enterobacteria* at 37°C (www.accessdata.fda.gov). The breakpoint for resistance to MCN is 16 µg/ml (32 µM) and 4 µg/ml (8 µM) for susceptible strains.

Relative to the ancestral BW25113 strain, *E.coli* expressing either wild-type *tet(X2)* or a candidate *tet(X2)* mutant had equal growth rates within experimental error in

the absence of MCN, indicating that all strains had nearly identical fitness under non-selective conditions (**Table 3.3**).

**Table 3. 3. Growth rates[a,b] of *E.coli* BW25113 and the adaptive tet(X2) strains in the absence of antibiotics.**

| BW25113 | TetX2 | T280A | N371I | N371T | S326I | F235Y | K64R | T280S |
|---|---|---|---|---|---|---|---|---|
| 0.013 | 0.013 | 0.012 | 0.012 | 0.012 | 0.013 | 0.013 | 0.013 | 0.012 |

[a]Growth rates presented in the table have not been normalized and represent raw data
[b]Errors in the measurements were ≤ 0.001

As expected, MCN had the most dramatic effect on the growth of the ancestral strain, since it did not have a tetracycline resistant determinant, showing 40% decrease in growth at 4.1 µM (2 µg/ml) MCN. Beyond 24.3 µM (8 µg/ml), the growth of the ancestral strain was not detected. Two strains expressing *tet(X2)*$_{T280S}$ and *tet(X2)*$_{K64R}$ exhibited slower growth rates at approximately 12 µM (4 µg/ml) MCN, in comparison to the rest of the adaptive mutants (0.73 ± 0.03, 0.69 ± 0.03, respectively) and their growths were completely inhibited beyond 40 µM MCN. The performance of these mutants was very similar to wild-type *tet(X2)*, suggesting that both point mutations did not have significant effect on the function of the wild-type enzyme.

Two strains carrying *tet(X2)*$_{F235Y}$ and *tet(X2)*$_{S326I}$ exhibited intermediate fitness when compared the most successful mutants. For example, growth rates of both *tet(X2)*$_{F235Y}$ (0.58 ± 0.01) and *tet(X2)*$_{S326I}$ (0.54 ± 0.01) were approximately 40% lower than that of *tet(X2)*$_{T280A}$ (0.92 ± 0.02) at 32 µM MCN. Compared to the strain expressing wild-type *tet(X2)*, the growth rate advantage of these mutations was over 300%, which

strongly suggests that these mutations had a significant effect on the *in vivo* function of the enzyme.

Strain BW25113*tet(X2)*$_{N371T}$ showed greater resistance toward MCN exhibiting a significant 4-fold increase in growth rate when compared to the wild-type *tet(X2)*. Finally, the most successful mutations *tet(X2)*$_{T280A}$, *tet(X2)*$_{N371I}$ had nearly indistinguishable growth rates at lower MCN concentrations < 40 μM (20 μg/ml) where MCN would be employed clinically. A similar five-fold increase in growth rates for strains carrying these most resistant mutations implies that *tet(X2)*$_{T280A}$ and *tet(X2)*$_{N371I}$ have the largest effect on the kinetics of the enzyme. In addition, these results suggest that both mutants would have comparable success in a population during selection at concentrations where *Tet(X2)*$_{T280A}$ was identified (**Figure 3.6**).

Table 3.4. Relative growth rates[a] of ancestral strain and *E. coli* expressing chromosomal copies of wild-type and adaptive mutations.

| MCN ($\mu$M) | Ancestor *E.coli*$_{BW25113}$ | TetX2 | TetX2$_{T280A}$ | TetX2$_{N371I}$ | TetX2$_{N371T}$ | TetX2$_{S326I}$ | TetX2$_{F235Y}$ | TetX2$_{K64R}$ | TetX2$_{T280S}$ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1.00 ± 0.01 | 1.00 ± 0.01 | 1.00 ± 0.01 | 1.00 ± 0.01 | 1.00 ± 0.01 | 1.00 ± 0.04 | 1.00 ± 0.03 | 1.00 ± 0.03 | 1.00 ± 0.02 |
| 4.1 | 0.59 ± 0.01 | 1.01 ± 0.01 | 0.97 ± 0.01 | 0.98 ± 0.01 | 0.98 ± 0.01 | 0.95 ± 0.01 | 0.95 ± 0.01 | 0.97 ± 0.01 | 1.00 ± 0.01 |
| 8.1 | 0.39 ± 0.01 | 0.94 ± 0.03 | 0.99 ± 0.01 | 0.99 ± 0.01 | 0.98 ± 0.01 | 0.95 ± 0.01 | 0.96 ± 0.01 | 0.91 ± 0.05 | 0.92 ± 0.05 |
| 12.2 | 0.24 ±0.01 | 0.73 ± 0.03 | 1.00 ± 0.01 | 0.99 ± 0.01 | 0.98 ± 0.01 | 0.93 ± 0.02 | 0.94 ± 0.01 | 0.69 ± 0.03 | 0.73 ± 0.03 |
| 16.2 | 0.09 ± 0.05 | 0.56 ± 0.03 | 1.00 ± 0.01 | 0.99 ± 0.01 | 0.96 ± 0.01 | 0.87 ± 0.02 | 0.85 ± 0.03 | 0.49 ± 0.03 | 0.52 ± 0.02 |
| 20.3 | 0.04 ± 0.01 | 0.41 ± 0.01 | 1.00 ± 0.02 | 0.97 ± 0.07 | 0.95 ± 0.07 | 0.79 ± 0.02 | 0.77 ± 0.02 | 0.29 ± 0.02 | 0.30 ± 0.02 |
| 24.3 | ND | 0.27 ± 0.01 | 1.00 ± 0.01 | 0.95 ± 0.04 | 0.87 ± 0.04 | 0.70 ± 0.02 | 0.67 ± 0.02 | 0.20 ± 0.01 | 0.21 ± 0.02 |
| 28.4 | ND | 0.16 ± 0.01 | 0.93 ± 0.03 | 0.89 ± 0.02 | 0.73 ± 0.02 | 0.58 ± 0.01 | 0.55 ± 0.01 | 0.12 ± 0.01 | 0.12 ± 0.02 |
| 32.4 | ND | 0.17 ± 0.01 | 0.92 ± 0.02 | 0.89 ± 0.01 | 0.75 ± 0.01 | 0.58 ± 0.01 | 0.54 ± 0.01 | 0.12 ± 0.01 | 0.11 ± 0.02 |
| 40.6 | ND | 0.09 ± 0.01 | 0.78 ± 0.04 | 0.77 ± 0.01 | 0.62 ± 0.01 | 0.44 ± 0.01 | 0.40 ± 0.01 | 0.07 ± 0.01 | 0.06 ± 0.01 |
| 60.8 | ND | ND | 0.37 ± 0.05 | 0.55 ± 0.01 | 0.34 ± 0.01 | 0.20 ± 0.01 | 0.16 ± 0.01 | ND | ND |
| 101.4 | ND | ND | 0.07 ± 0.04 | 0.23 ± 0.02 | 0.12 ± 0.02 | 0.10 ± 0.02 | 0.03 ± 0.02 | ND | ND |

[a] Growth rates in the table are normalized for each strain tested to the growth rate determined at 0$\mu$M MCN.
ND = growth not detected
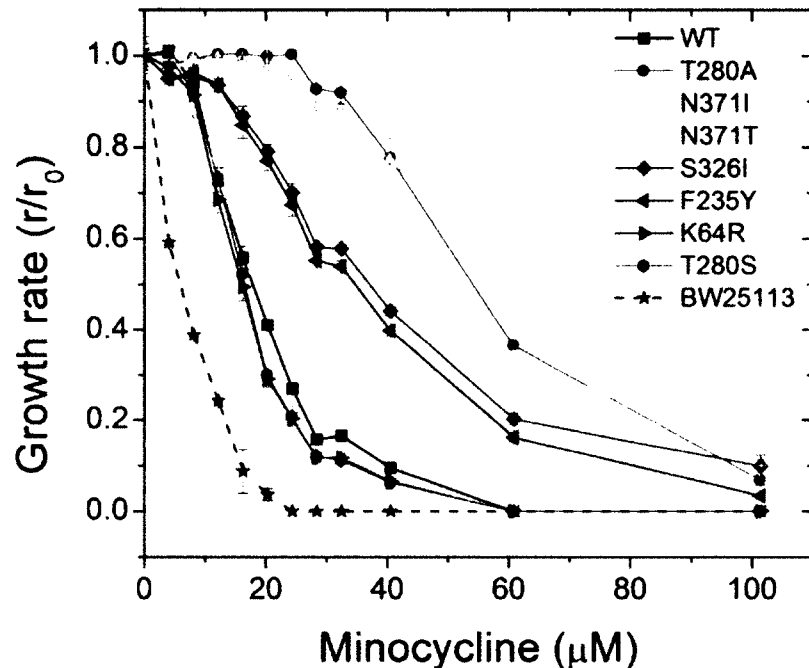
**Figure 3.6. Similar growth rate profiles for T280A and N371I suggest that both mutants should have equal opportunities for success in the evolution experiment.**

Growth of the ancestral strain *E. coli* BW25113 (no *tet(X2)*), strains with chromosomal copy of *tet(X2)*$_{\text{wild-type}}$ and seven variants were monitored at $OD_{600}$ for 24 hours. Growth rates determined at various MCN concentrations were plotted against the MCN concentrations (0-102 μM). The experiment was repeated for three individual colonies represented by error bars corresponding to the standard deviation between the measurements. At lower MCN concentrations, growth rates for all the variants are comparable, suggesting that all the variants have equivalent fitness. Under conditions where T280A was selected for (32 μM), the growth rates of N371I and T280A are nearly identical, while the growth rates of T280S and K64R are significantly lower and resemble the growth of wild-type TetX2. Growth rates (r) of each mutant at different MCN concentrations were normalized to its growth rate in the absence of MCN ($r_0$).

## 3.4. Comparison of lag times for the adaptive mutants during growth at various MCN concentrations.

Serial passage of populations in flasks allows for changes in fitness that occur from an ability to exit stationary phase. The faster an organism enters the log phase, the greater advantage it will have in an evolving population where millions of cells compete for success. In order to determine how much time was required for each strain carrying an adaptive mutant to begin exponential growth, we compared the lag times of each of the strains at various MCN concentrations (**Figure 3.7**).

Relative to a stain carrying the wild-type allele, all mutants had indistinguishable lag times at low MCN concentrations, which is consistent with similar growth rates at those MCN concentrations. Similarly to the wild-type, strains expressing $tet(X2)_{T280S}$ and $tet(X2)_{K64R}$ experienced extended lag times starting at >20 $\mu$M MCN. At 60 $\mu$M MCN, the time to enter exponential phase began to increase for all strains carrying mutant alleles that conferred higher resistance to MCN. At 32 $\mu$M MCN, which is the concentration of MCN at which T280A mutation was first observed, the lag time of strains carrying $tet(X2)_{F235Y}$, $tet(X2)_{S326I}$ and $tet(X2)_{N371T}$, $tet(X2)_{N371I}$ and $tet(X2)_{T280A}$ were identical. These data suggest that the mutants exhibiting higher resistance toward MCN show very similar growth patterns regarding their lag times at higher antibiotic concentrations.
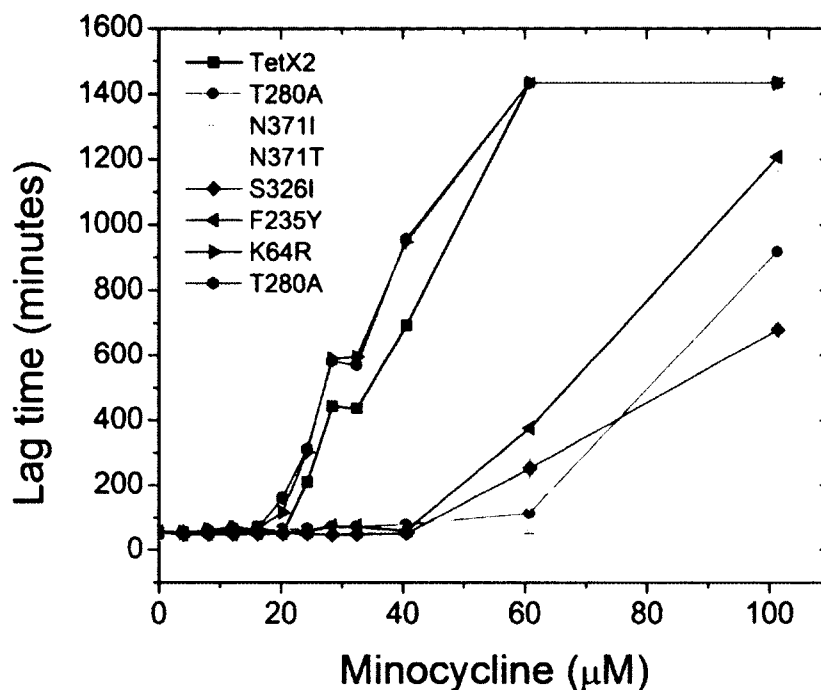
**Figure 3.7. Mutant strains of *tet(X2)* conferring higher resistance show similar lag time profiles at various MCN concentrations.**

The growth of *E.coli* strains with chromosomal copy of *tet(X2)*$_{wild-type}$ and seven adaptive variants were monitored at $OD_{600}$ for 24 hours. In addition to determination of growth rates (Figure 3.4) at various MCN concentrations, lag times of each mutant strain was estimated. The lag time described the time that it takes for bacteria to come out of lag phase and enter into log phase. The results were plotted against wide range of MCN concentrations (0-102 µM). The experiment was repeated for three individual colonies represented by error bars corresponding to the standard deviation between the measurements. At lower MCN concentrations, all variants take an equal amount of time to exit from the lag phase, which is in agreement with the growth rates data at low MCN concentrations. The differences in lag times are most apparent at higher MCN concentrations (≥60 µM MCN).

### 3.5. Conclusions

*E. coli* carrying chromosomal copy of *tet(X2)* showed a dramatic ability to adapt to very high MCN concentrations in a small number of generations. Changes within both *tet(X2)* gene, as well as the *E. coli* chromosome contributed to the nearly 20 fold increase in resistance and were critical for the adaptation to such high MCN concentrations since the strain lacking *tet(X2)* did not evolve higher resistance using even much weaker selections (a more gradual increase in MCN concentration). There results strongly imply that *tet(X2)* confers increased resistance that is highly epistatic with chromosomally encoded genes and suggests that tet$^r$ alleles may also facilitate adaptation through other genes on the chromosome. The observation that *tet(X2)* was required to reach the highest level of resistance was surprising since it is clear that the *E. coli* have the ability to move from MICs of 64 to 320 μg/ml but do not seem able to make the initial adaptation from 16 – 64 μg/ml without TetX2 or TetA. In another strain of *E. coli*, Viveiros et al. showed that 110 days of slow adaptation was required to select for even modest resistance to tetracycline (Viveiros et al., 2005). It may be that *tet(X2)* and other tet$^r$ (TetA or TetM) alleles may be strong potentiators for higher resistance as well as directly conferring increased resistance. Although, we have shown that under laboratory settings, bacteria can evolve very high resistance, the goal of this study is to remain within the most clinically relevant settings. Therefore, the focus for the remainder of this study is to identify mechanisms that play a role in the initial adaption steps.

The success of the adaptive mutants generated by error-prone mutagenesis was assessed by the growth assays at clinically relevant MCN concentrations. Based on the

performance at various MCN concentrations, the mutants can be divided into three groups: weak (wild-type like), intermediate and strong. Weak mutants included $tet(X2)_{K64R}$ and $tet(X2)_{T280S}$, which behaved very similarly to the wild-type $tet(X2)$ at all MCN concentrations tested. The analysis of the fastest growth rates and the time each strains spent in the lag phase revealed that four mutants; $tet(X2)_{N371I}$, $tet(X2)_{N371T}$, $tet(X2)_{F235Y}$ and $tet(X2)_{S236I}$, in addition to $tet(X2)_{T280A}$ exhibited increases in growth rates at moderate MCN concentrations (**Figure 3.6** and **3.7**). Mutants with the strongest effects on growth rates compared to the wild-type strain included $tet(X2)_{N371I}$ and $tet(X2)_{N371T}$, with $tet(X2)_{N371I}$ exhibiting the same growth profile as $tet(X2)_{T280A}$. We did not identify any single-point mutants that had greater resistance than $tet(X2)_{T280A}$.

The fitness advantages of strains carrying $tet(X2)_{N371I}$ and $tet(X2)_{T280A}$ in the presence of MCN strongly suggest that both alleles had equal chances for success in the adaption experiment. Thus, it is puzzling that even though both $tet(X2)_{T280A}$ and $tet(X2)_{N371I}$ had identical growth rates at concentrations where $tet(X2)_{T280A}$ was isolated, (32 $\mu$M) yet we did not recover any isolates carrying $tet(X2)_{N371I}$ by experimental evolution. Is it simply not being made or does it appear but at much lower frequency, so it was not detected by sequencing because of a small sample size?

There are three likely scenarios that might have played a role in the evolutionary dynamics. One possibility is that $TetX2_{T280A}$ is a better enzyme at low MCN concentrations and as a result has an advantage over $TetX2_{N371I}$ early on in the adaption experiment (growth rates of $tet(X2)_{T280A}$ are not affected by MCN up to approximately 24 $\mu$M). It should also be noted that as MCN concentrations increase $TetX2_{N371I}$ should

have had a significant advantage over TetX2$_{T280A}$, however at higher MCN concentrations changes in the chromosome contribute to the increase in resistance.

Additionally, clonal interference and/or introduction of a bottle neck during daily transfer could have had a major influence on the evolutionary outcomes in the population. Clonal interference is common in asexually reproducing populations and occurs when beneficial mutation arise independently in different individuals (Elena and Lenski, 2003). As a result, clones carrying these beneficial mutations will compete and therefore inhibit each others' spread in a population. This often results in a loss of one of the mutations. In addition, the consequences of clonal interference are that the rates of spread of the beneficial mutation are much slower than predicted by the fitness benefit of a particular mutation in a clonal population (Gerrish and Lenski, 1998). Because of the similarity in fitness conferred by *tet(X2)*$_{T280A}$ and *tet(X2)*$_{N371T/I}$ their proportions within the population relative to one another will not change at low drug concentrations. However, if there is an initial bias related to mutational frequency, this bias will remain until a significant fitness difference is achieved by the less frequent genotype.

The reason for that initial bias in the *E. coli* population could be that mutation T280A results from a transition (ACG →GCG) whereas N371I and N371T were both a result of transversion (AAC→ATC, AAC→ACC, respectively). As mentioned briefly in the introduction, mutations arising from transitions are more common than transversions (Bulmer, 1986; Vogel and Rohrborn, 1966; Wielgoss et al., 2011). This phenomenon in evolutionary biology is called transition bias. Transition is a substitution of a purine to another purine (C↔T) or pyrimidine to another pyrimidine (A↔G), whereas transversion

is a substitution from purine to pyrimidine or vice versa (T↔A, C↔G, T↔G, A↔C). For each base there is one possibility to substitute for a transition, but two possibilities for transversions, therefore without bias, the expectation is that transversions would be twice as common as transitions. However, studies of DNA sequences from both pseudogenes and functional genes (Ochman, 2003; Vogel and Kopun, 1977; Vogel and Rohrborn, 1966) have shown that DNA substitutions are not random and that bias toward transitions exists. Based on mutation frequency studies on housekeeping genes sequenced in multiple strains of E. coli and S. enterica, it has been estimated that the ratio of transitions versus transversions is about 2.5:1 (Ochman, 2003). Therefore, it is possible that background generic variance (mutational supply) played a role in the evolutionary dynamics in the adapting E. coli population to MCN. It is possible TetX2$_{T280A}$ had a greater probability to arise in the starting population, and therefore had an advantage from the very beginning over TetX2$_{N371I}$ or TetX2$_{N371T}$ which resulted from transversions. The transition bias does not exist in the error-prone PCR generated library, because the polymerase and conditions used in this experiment were optimized to remove transition/transversions bias.

The second possibility is that a bottleneck was introduced during the daily transfer of a small fraction of a population to fresh media. Based on the mutation rate, size of tet(X2) gene (1140 bases), the population size (6.6×10$^9$ cfu/ml) and number of generations (~16) in the 50 mL flask, we estimated that 6.7×10$^4$ mutations in tet(X2) gene were generated in the first 24 hours. Transferring 1/1000$^{th}$ of total population to a new flask means that on average sixty seven tet(X2) mutations were transferred, which constitutes only 2% of all possible mutations. However, within the next twenty four hours

in the 50 ml culture, the sequence space of *tet(X2)* would be sampled approximately over two fold and new variants will be generated. It is important to note, that the mutation rates increase in response to stress (e.g. antibiotics) for *E. coli* (Rosenberg and Hastings, 2003) and thus this calculation is only a rough estimate since the actual mutation rate is unknown.

We tested the reproducibility of TetX2 mediated adaptation to MCN by repeating the serial passage experiment in ten parallel strains under the same conditions as outlined previously. If TetX2$_{T280A}$ was the only mutation or was more frequent early on we would be able to see this by increasing the sample size through performing the experiment in several populations. Deep sequencing of the ten replicate populations was used to quantitate the allelic frequency of the adaptive populations. In principle, this method will allow us to observe the earlier spectrum of mutations and the whether N371I/T are made at comparable frequencies to T280A.

# Chapter 4: Biophysical characterization of TetX2 and its adaptive mutants

We isolated seven adaptive mutants of TetX2 with single amino acid changes in protein sequence that confer equal or higher resistance towards MCN. Resistance was measured by growth assays at various antibiotic concentrations and then correlated to the catalytic properties of the wild-type and mutant enzymes.

TetX2 was originally isolated from transposons *Tn4351/Tn4400* found in *Bacteroides fragilis*, however its activity against tetracyclines was not realized until it was cloned and expressed in *E. coli* (Speer et al., 1991). Based on amino acid homology to NADPH – dependant oxidoreductases, it was suggested that TetX2 required NADPH and oxygen for its activity (Speer et al., 1991), which may explain why TetX2 was not functional in anaerobic *Bacteroides*. *In vitro* activity of TetX2 has been studied before using a spectrophotometric assay against a broad range of tetracycline derivatives, including the most recently FDA approved tetracycline derivative, Tygacil® (Pfizer) (Moore et al., 2005; Wangrong Yang, 2004). These studies showed that TetX2 belongs to a family of FAD-dependent monooxygenases and confirmed its requirement for NADPH, molecular oxygen and magnesium for catalytic activity.

Monooxygenases catalyze specific insertion of one oxygen atom into an organic substrate (van Berkel et al., 2006). This reaction relies on activation of molecular oxygen, which often requires a transition metal bound to a cofactor (e.g. heme in P450 monooxygenase) or a purely organic cofactor such as flavin (van Berkel et al., 2006).

Most flavoprotein monooxygenases contain non-covalently bound FAD as a prosthetic group that has to be in a reduced form to react with its substrate (Massey, 1994). Reduction most often occurs by reaction with NADPH, enabling the enzyme to react with molecular oxygen and leading to formation of reactive C4a-hydroperoxyflavin (Entsch et al., 1976). Because TetX2 is NADPH dependant but lacks the common NADPH binding fold, it was classified as a class A flavin monooxygenase (Speer et al., 1991). In this class of flavoproteins, C4a-hydroperoxyflavin is the oxygenation species that performs an electrophilic attack on the aromatic ring of its substrate (van Berkel et al., 2006). The most studied class A monooxygenase and thus a model enzyme for this subclass is para-hydroxybenzoate hydroxylase (PHBH) (Entsch and van Berkel, 1995). Although, TetX2 shares only 21% sequence identity with PHBH, three dimensional structure comparisons of these enzymes (Chapter 5) reveal substantial similarities in their overall folds. This structural homology between PHBH and TetX2 suggests some conservation of the kinetic mechanism for the two enzymes.

Previous studies of other FAD-dependent monooxygenases and TetX2 suggest that initially flavin is reduced by NADPH so it can react with molecular oxygen to form a C4a-flavin hydroperoxide, which regioselectively hydroxylates tetracycline (Wangrong Yang, 2004). However, detailed steady state kinetic analyses of the reaction mechanism were not performed.

## 4.1. Enzyme kinetics of TetX2 and TetX2$_{T280A}$

Previously reported kinetic parameters of TetX2 were determined by fitting initial velocities into classic Michealis-Menten equation which allowed determination of catalytic rates ($v_i$) and Michaelis constant ($K_M$) for tetracyclines (Wangrong Yang, 2004). However, the reaction catalyzed by TetX2 involves two additional substrates: NAPDH and molecular oxygen (**Figure 5.1**). The adaptive mutants could potentially affect interactions of any of the substrates with the protein, so we sought to determine the Michaelis-Menten constants and steady-state catalytic rates as function of both [MCN] and [NADPH]. Ideally, the third substrate, oxygen should also be taken into account but varying concentrations of oxygen is very challenging and studies of the catalytic mechanism of PHBH have shown that the rate of oxygen binding is very fast and is not the rate limiting step (Entsch and van Berkel, 1995). Moreover, the oxygen concentration during bacterial growth is held constant for all our fitness experiments. Therefore, we focused our efforts on understanding the interactions of the proteins with MCN and NADPH and their effects on the rates of catalysis.

We monitored the progress of TetX2 catalysis against MCN by varying both MCN and NADPH concentrations at 37 °C, to closely mimic the conditions under which adaptive mutants were selected (see **Section 2.6**). Because, tetracyclines have two absorption maxima at 260 nm and 363 nm and NADPH has an absorption maximum at 340 nm, the change in absorbance corresponding to MCN hydroxylation alone was monitored at 400 nm.
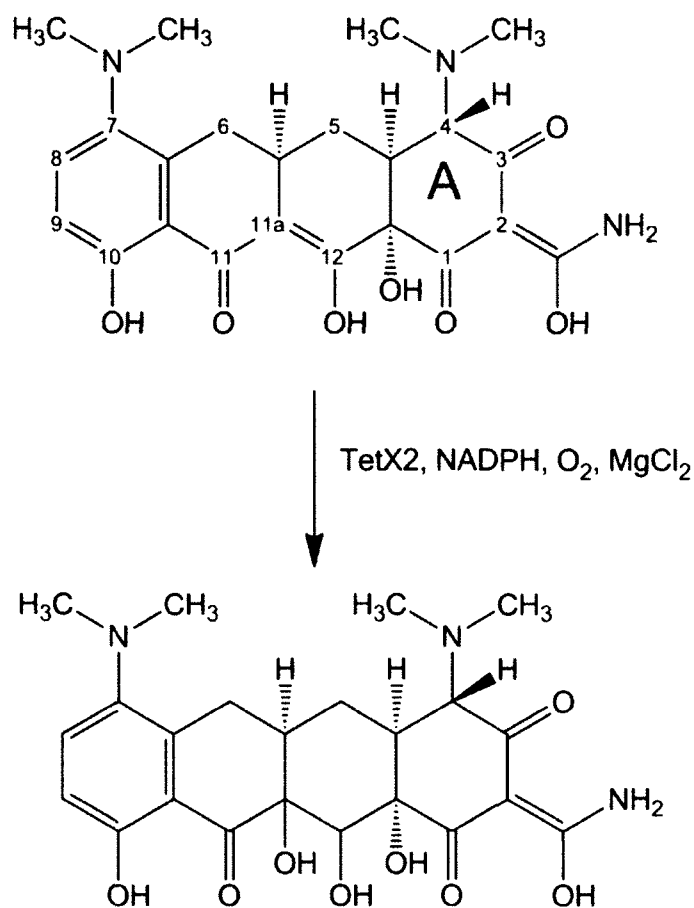
**Figure 4.1. Hydroxylation of minocycline catalyzed by TetX2 in the presence of NADPH, O₂ and MgCl₂.**

The inactivation of minocycline results in the formation of a unstable 11a-hydroxyminocycline by hydroxylation at position 11a. The product of this reaction is further degraded by non-enzymatic reactions to open ring structures that do not inhibit bacterial translation reactions (Wangrong Yang, 2004).

The initial velocities were measured at various NADPH (12.5, 25, 50, 100, 200

$\mu$M) and MCN (0, 5, 10, 25, 50, 100, 200 $\mu$M) concentrations (**Figure 4.2.A**). As shown

in **Figure 4.2.A**, the initial velocities at different MCN concentrations increase with

increasing NADPH which in turn results in an increase of $V_{max(obs)}$. In addition, the $K_{M(obs)}$

also increases with increasing NAPDH concentrations, which suggests that only binary

complexes of the enzyme with these substrates can occur and that MCN and NADPH can

not be bound at the same time in the active site (i.e. no ternary E:MCN:NADPH complex

forms). The double reciprocal plots (Lineweaver-Burke plots) of initial velocities versus

MCN concentrations were graphed to confirm the binary mechanism. The expectation

from this test is that parallel (nonintersecting) lines will be observed if a binary complex

mechanism is occurring (Lehninger, 2008). In contrast, the kinetic characteristic of

ternary complex mechanisms, are converging lines in double reciprocal plots, which

mean that the enzyme forms a complex with both substrates (ternary complex) and the

$K_M$ for the ternary complex has a finite value (Lehninger, 2008).

In the double reciprocal plots for TetX2, the lines do not intersect and appear to

have similar slopes (**Figure 4.2.B**). The slopes are similar because the $K_{M(obs)}$ for MCN

shows the same increase with [NADPH] as does the $V_{max(obs)}$, which is characteristic for

binary complex mechanisms, where the two substrates compete for the same site on the

enzyme. The kinetic parameters; $K_{M(MCN)}$, $K_{M(NADPH)}$, $k_{cat}$; reported in **Table 4.1** were

calculated by fitting initial velocities to a simple, rapid equilibrium binary complex

model for the two substrates MCN and NADPH (**Eq. 4.1, Scheme 4.1**).

$$\text{E}^{\text{red}} + \text{A} \xrightleftharpoons{K_A} \text{E}^{\text{red}}\text{A} \xrightarrow{k_1} \text{E}^{\text{ox}} + \text{P}$$

$$\text{O}_2$$

Scheme 4.1

$$\text{E}^{\text{ox}} + \text{B} \xrightleftharpoons{K_B} \text{E}^{\text{ox}}\text{B} \xrightarrow{k_2} \text{E}^{\text{red}} + \text{Q}$$
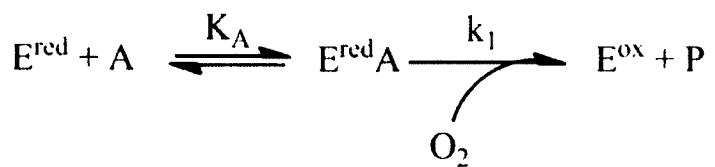
$$\text{where} \quad k_1 = \frac{k_{ox}[O_2]}{k_{02} + [O_2]}$$

$$v_o = \frac{V_{max}(obs)[A]}{[A] + K_M(obs)} \quad \text{where}$$

Eq.4.1

$$V_{max}(obs) = \frac{V_{max}[B]}{K_A + [B]}$$

Eq.4.2

$$\text{and} \quad K_M(obs) = \frac{K_A[B]}{K_B + [B]}$$

Eq.4.3

In the mechanism illustrated in **Scheme 4.1**, the substrate free enzyme in a reduced form ($\text{E}^{\text{red}}$) reacts with the first substrate A (MCN) with an equilibrium dissociation constant defined by $K_A$ to form the first binary complex ($\text{E}^{\text{red}}\text{A}$). This complex then reacts rapidly with $O_2$, which is held at a fixed concentration. In our simplified analysis, this hydroxylation step is represented by a simple pseudo-first order process defined by $k_1$. The net result is formation of the oxidized enzyme and dissociation of the hydroxylated and inactivated MCN. This step should, in principle depend on $[O_2]$.

The second reaction involves formation of a binary complex between the oxidized enzyme ($E^{ox}$) and the reductant (B=NADPH) to regenerate the enzyme to its active form ($E^{red}$). The binding of NADPH to $E^{ox}$ is defined by the equilibrium constant $K_B$. This mechanisms shows that both A and B compete for the same active site on the enzyme and assumes that the binding of MCN and NADPH is much faster than the hydroxylation and electron transfer steps. It is probable that a more complex mechanism applies with respect to these assumptions. However, in all cases parallel lines will be observed if only binary complexes occur, and the key result for these fitness experiments is how the empirical values of $K_{M(MCN)}$, $K_{M(NADPH)}$ and $V_{MAX}$ are altered by mutagenesis.
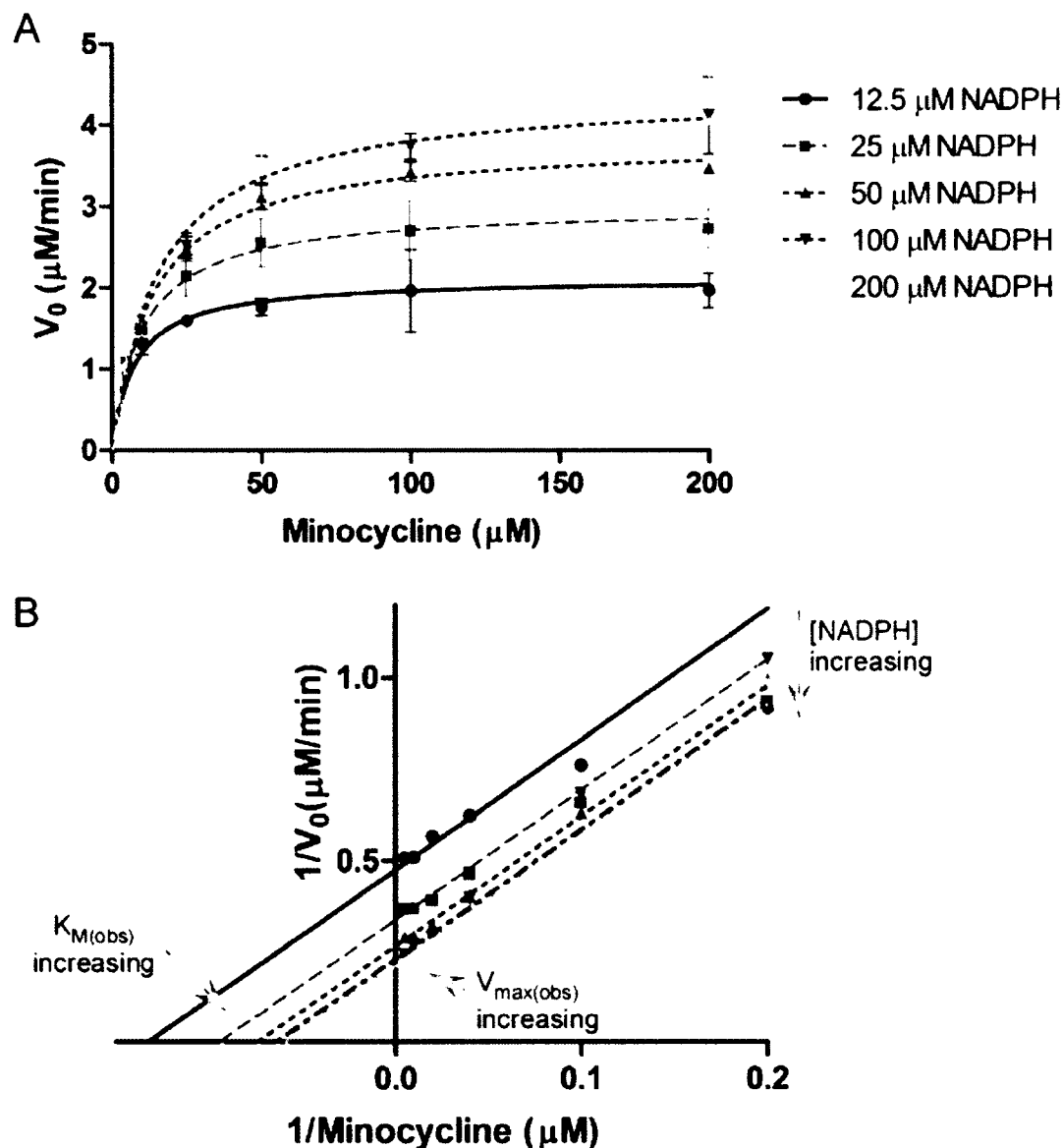
**Figure 4.2. The enzymatic profile of adaptive mutant TetX2$_{T280A}$ isolated by experimental evolution implies binary complex formation.**

A) Hydroxylation of MCN by TetX2$_{T280A}$ was monitored at 12.5, 25, 50, 100 and 200 μM of NADPH at 37°C for 5 minutes. The initial velocities were fit to the simple binary complex mechanism expression shown in Eq.4.1. B) The double reciprocal plot for the initial velocity of MCN hydroxylation at various NADPH concentrations exhibits characteristics of binary complex formation with parallel lines indicating identical slopes and therefore same dependence of $V_{max(obs)}$ and $K_{M(obs)}$ on NADPH. The error bars correspond to the standard deviation among three individual measurements.

**Table 4.1. Steady-state kinetic parameters for wild-type TetX2 and adaptive mutants isolated by directed evolution**

| | $K_{M(MCN)}$ ($\mu$M) | $K_{M(NADPH)}$ ($\mu$M) | $k_{cat}$ ($V_{max}/[E_{total}]$) ($s^{-1}$) | $k_{cat}/K_{M(MCN)}$ ($\mu M^{-1} s^{-1}$) | $k_{cat}/K_{M(NADPH)}$ ($\mu M^{-1} s^{-1}$) |
|---|---|---|---|---|---|
| **WT** | 35 ± 1.9 | 75 ± 4.1 | 0.34 ± 0.01 | 0.010 ± 0.0006 | 0.004 ± 0.0003 |
| **T280A** | 18 ± 0.9 | 18 ± 1.1 | 0.43 ± 0.01 | 0.024 ± 0.0013 | 0.024 ± 0.0016 |
| **T280S** | 30 ± 3.4 | 100 ± 14 | 0.18 ± 0.01 | 0.006 ± 0.0008 | 0.002 ± 0.0003 |
| **F235Y** | 54 ± 6.1 | 99 ± 11 | 0.32 ± 0.06 | 0.006 ± 0.0013 | 0.003 ± 0.0007 |
| **N371I** | 18 ± 1.9 | 64 ± 6.3 | 0.37 ± 0.02 | 0.020 ± 0.0024 | 0.006 ± 0.0006 |
| **N371T** | 24 ± 2.1 | 130 ± 11 | 0.40 ± 0.02 | 0.017 ± 0.0017 | 0.003 ± 0.0003 |
| **S326I** | 37 ± 2.8 | 73 ± 5.5 | 0.36 ± 0.01 | 0.010 ± 0.0008 | 0.005 ± 0.0004 |
| **K64R** | 36 ± 4.6 | 110 ± 15 | 0.32 ± 0.03 | 0.009 ± 0.0014 | 0.003 ± 0.0005 |

## 4.2. Enzyme kinetics of mutants isolated by directed evolution experiment

All mutant enzymes exhibited similar mechanistic behavior with modest changes in their kinetic parameters. The adaptive mutant T280A was the most enzymatically active enzyme among all mutants tested with the largest $k_{cat}$ (0.43 $s^{-1}$) and smallest $K_M$s for both MCN (18 $\mu$M) and NADPH (18 $\mu$M), resulting in 2 to 3-fold increases in enzymatic efficiency ($k_{cat}/K_{M(MCN)}$ or $k_{cat}/K_{M(NADPH)}$) when compared to the wild-type TetX2 ($K_{M(MCN)}$= 35.4 $\mu$M, $K_{M(NADPH)}$=75 $\mu$M, $k_{cat}$=0.34 $s^{-1}$). Both, TetX2$_{N371I}$ and

TetX2$_{N371T}$ have similar lower $K_M$ values (18 and 24 µM, respectively) for MCN when compared to TetX2$_{T280A}$, but the $K_{M(NADPH)}$ for TetX2$_{N371I}$ (64 µM) is similar to that of the wild-type enzyme and 3.5-fold higher than that for TetX2$_{T280A}$. The $K_{M(NADPH)}$ for TetX2$_{N371T}$ is 2-fold higher (134 µM) than the $K_{M(NADPH)}$ of the wild-type enzyme; however both the TetX2$_{N371I}$ and TetX2$_{N371T}$ mutants show slight increases in $k_{cat}$ values (0.37 and 0.40 s$^{-1}$, respectively) and as a result have similar $k_{cat}/K_{M(NADPH)}$) values and two-fold higher $k_{cat}/K_{M(MCN)}$ compared to wild-type TetX2.

Surprisingly, the kinetic parameters for the bacterial mutants with intermediate resistance to MCN, TetX2$_{S326I}$ and TetX2$_{F235Y}$, are very similar to those for the wild-type enzyme, showing no change in the $K_M$ (37.4 and 53.9 µM, respectively) for MCN and $k_{cat}$ (0.36 and 0.32 s$^{-1}$, respectively) and only a slight increase in the $K_{M(NADPH)}$ (73 and 99 µM, respectively). Finally, TetX2$_{T280S}$ and TetX2$_{K64R}$ show no change in the $K_{M(MCN)}$ (30 and 36 µM, respectively) but both have higher $K_{M(NADPH)}$ values (105 and 112 µM, respectively) and lower rates of catalysis, particularly for TetX2$_{T280S}$ (0.18 s$^{-1}$).

The catalytic parameters of TetX2$_{T280A}$ provide a rationale for the selection of T280A mutation in the evolution experiment, but don't explain why TetX2$_{N371I}$ and TetX2$_{N371T}$, which have similar enzymatic efficiency, are not observed in the selection experiments. However, the *in vitro* catalytic activity of TetX2$_{S326I}$ and TetX2$_{F235Y}$ alone could not be linked to their improved performance *in vivo* as shown by growth assays. In addition to enzymatic activity, adaptive mutations may also affect the physiology of the entire organism by altering regulation, expression levels and/or stabilities of proteins *in vivo* (Soskine and Tawfik, 2010). Thus, we investigated protein stability to determine if

increases in fitness that are not evident from *in vitro* activity assays could be due to increased resistance of the mutant enzymes to denaturation.

### 4.3. Stability of TetX2 and TetX2$_{T280A}$

In order to assess the effect of the adaptive mutation T280A on the amount of folded protein, the *in vitro* stabilities of wild-type and TetX2$_{T280A}$ were determined by thermal denaturation monitored by circular dichroism (CD). The stability data shows that both TetX2 and TetX2$_{T280A}$ have nearly identical unfolding profiles (**Figure 4.3.A**). However, the thermal unfolding midpoint ($T_m$) could only be approximated as protein denaturation was largely irreversible and at least two transitions were observed. The large transition occurs at 71.7 ± 0.5°C and 72.6 ± 0.5°C, for wild-type and TetX2$_{T280A}$, respectively (**Figure 4.3.A**). In addition, prior to the sharp transition, a gradual decrease in the CD signal of about 5 mdeg is observed between 25-45°C (**Figure 4.3.B**). Chemical unfolding in the presence of guanidinium chloride shows that the change in signal for fully unfolded TetX2 is expected to be significantly larger than what is observed at lower temperature (**Figure 4.3.C**). Although, it is unclear what exactly this small decrease in CD signal corresponds to, we postulate that it might be due to local unfolding, or partial loss of FAD cofactor. Even though, we were not able to assess the fraction of folded protein using CD thermal unfolding, we concluded that the T280A mutation does not dramatically alter the stability of the enzyme. Since the unfolding of TetX2 and TetX2$_{T280A}$ are irreversible and involve at least two transition states, the stability of the remainder of the mutants were not analyzed using this approach.

We shifted our focus from protein stability to *in vivo* steady-state protein levels as a probable cause of increased fitness despite a lack of increase in catalytic efficiency for some of the mutants.
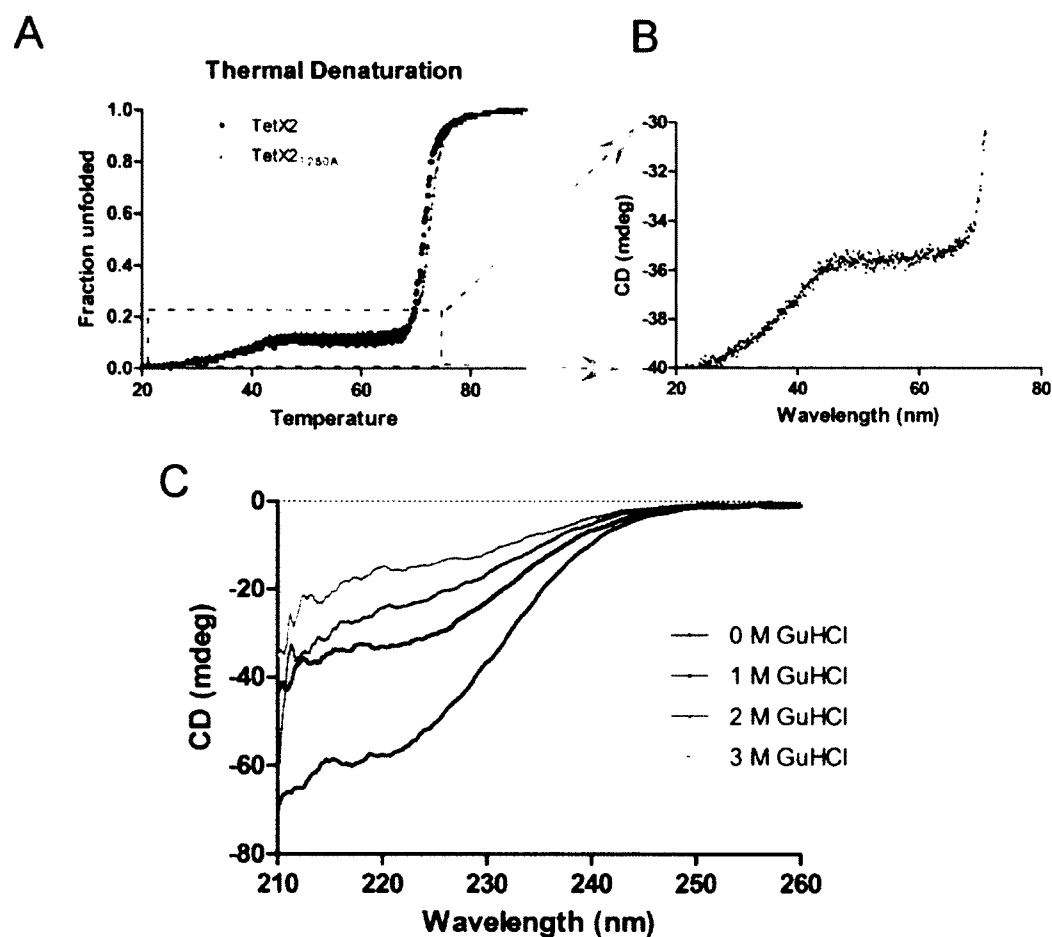


**Figure 4.3. T280A does not significantly alter the stability of TetX2.**
Thermal stability of TetX2$_{wild-type}$ and TetX2$_{T280A}$ was monitored by change in circular dichroism as function of temperature. A) Thermal denaturation midpoint (Tm) of unfolding is estimated at 71.7°C for TetX2$_{wild-type}$ (black solid circles) and at 72.6°C for TetX2$_{T280A}$ (red open triangles) for the sharp transition. B) Small change in CD signal during thermal unfolding of TetX2 occurs between 30-45°C and precedes sharp transition at higher temperature. C) Chemical unfolding in the presence of guanidinium chloride (GuHCl) of TetX2 shows a large change in CD between folded (0M GuHCl) and unfolded protein (3M GuHCl).

## 4.4. Relative steady-state concentrations of mutant enzymes

Although direct measurement of steady-state kinetic parameters correlated well with the *in vivo* success or failure for five of the seven adaptive TetX2 variants, TetX2$_{F235Y}$, and TetX2$_{S326I}$ had kinetic parameters comparable to wild type enzyme (**Table 4.1**) and yet significantly increased growth rates at higher MCN concentrations (Chapter 3, **Figure 3.5**). TetX2$_{F235Y}$, and TetX2$_{S326I}$ also displayed greater than expected growth rates when expressed from plasmids (data not shown) and led us to consider whether these two variants might be present at higher steady-state concentrations *in vivo*. Our preliminary mathematical model (work by Andres Benitez summarized in Chapter 6) relating bacterial growth rates to kinetic parameters of TetX2 had also suggested that *in vivo* steady-state concentrations of TetX2$_{F235Y}$, and TetX2$_{S326I}$ would likely be higher. To test this hypothesis, TetX2 activity was measured from cell extracts of *E. coli* expressing chromosomally integrated TetX2 variants. Based on the specific activities obtained from cell lysates (see **Section 2.7**), the relative enzyme concentrations of each mutant compared to the wild-type TetX2 was determined.

As shown in **Figure 4.3**, the steady-state protein levels of TetX2$_{F235Y}$ and TetX2$_{T280S}$ are ~2-fold higher than the wild-type, also TetX2$_{N371I}$, TetX2$_{N371T}$ and TetX2$_{S326I}$ appear at ~50% higher levels. The differences in the steady-state protein levels among the mutants are consistent with the growth rate data, which suggest that strains expressing TetX2$_{S326I}$ and TetX2$_{F235Y}$ are better at hydrolyzing MCN than wild-type TetX2. In addition, higher *in vivo* levels of TetX2$_{T280S}$ explain wild-type like

performance of this mutant in the growth assays, which could not be explained by kinetics due to lower catalytic rate.
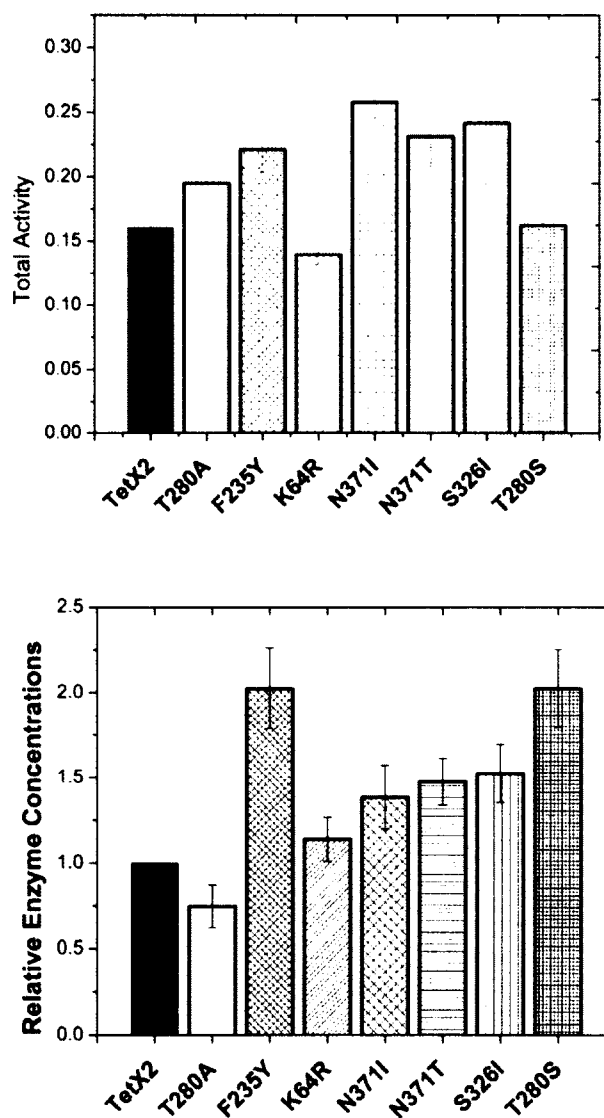
**Figure 4.4. Adaptive mutants show differences in their relative steady-state protein concentrations.**

A)      Specific activities of TetX2 and its potential adaptive mutants were determined from cell lysates of chromosomally expressed enzymes in *E. coli* BW25113. Total activity is plotted for strains carrying each mutant allele. B) Relative protein concentrations were determined by fitting initial velocities determined in the assay to Eq.4.1 and implementing kinetic parameters calculated in the *in vitro* kinetics assay. All data was normalized to the wild-type enzyme. Error bars correspond to standard-deviation between three independent measurements. Statistically significant differences in expression levels were observed for all mutants, when compared to the wild-type (p-value < 0.05) and were most noticeable for $TetX2_{F235Y}$ and $TetX2_{T280S}$.

## 4.5. Discussion

The steady-state kinetic parameters determined for the wild-type and mutant enzymes provide important insights into the catalytic mechanism of MCN hydrolysis. Based on these results and what has been known from studies on the related monoxygenase, PHBH, we propose that the mechanism of MCN degradation consists of at least two binary complex reactions, one reductive and one oxidative (Entsch and van Berkel, 1995). In the reductive step, the enzyme containing oxidized flavin reacts with NADPH to form of a binary complex. This mechanism is in contrast to what is observed for PHBH, where both NAPDH and p-OHB form a ternary complex (Entsch et al., 1976). The parallel lines, shown in the double reciprocal plots for our experiments (**Figure 4.2**), clearly eliminate the possibility of ternary complex formation suggesting that MCN is never present in the active site along with NADPH. Thus, after reduction of FAD, $NADP^+$ is released from the active site before MCN interacts with the reduced flavin.

In the second oxidative step, the enzyme with reduced flavin and bound MCN rapidly reacts with molecular oxygen to form a C4a-hydroxyperoxide. The position of oxygen transfer is conserved among most flavomonooxygenases that react with aromatic substrates (Entsch and van Berkel, 1995). The C4a-hydroxyperoxide then reacts rapidly with MCN, which leads to formation of a hydroxylated product that is unstable at neutral pH. NMR studies with another substrate, oxytetracycline, under acidic conditions showed

formation of a monohydration product at C11a of this substrate, which is further broken down through a series of non-enzymatic reactions (Wangrong Yang, 2004).

Overall, TetX2 shares homology with PHBH in terms of function and structure (discussed in Chapter 5), however as shown above there are several differences in the details in their mechanisms of catalysis. In this study, the mechanism of MCN hydroxylation by TetX2 was studied under steady-state conditions, however to truly elucidate the fate of FAD, NADPH, oxygen and MCN, a more thorough analysis should be performed using rapid reaction techniques such as stop-flow spectroscopy to examine the detailed kinetics of the half reactions and the reaction of the $E^{red}MCN$ complex with $O_2$.

The *in vitro* stability studies have shown that the T280A mutation does not significantly change the stability of the enzyme. This may explain why compensatory mutations that serve to stabilize proteins would not be necessary for TetX2 and were not observed in the adaptation experiment.

*In vivo* success of the adaptive mutants was reflected by *in vitro* catalytic performance of the isolated enzymes (except $TetX2_{F235Y}$ and $TetX2_{S326I}$). For example, when the most enzymatically active mutant, $TetX2_{T280A}$, was introduced into the chromosome of *E. coli*, we observed a very significant improvement in fitness (240-575%) at moderate MCN concentrations (10-16 µg/ml) compared to a strain carrying wild-type enzyme. The same held true for less active enzymes. The *in vitro* catalytic performance of $TetX2_{K64R}$ closely resembled that of the wild-type TetX2, which suggested that the growth profiles of these two enzymes should be very similar. As

expected, the fitness of $TetX2_{K64R}$ and TetX2 were nearly identical at all MCN concentrations tested, showing a tight link between *in vitro* enzyme kinetics and organismal fitness.

We have asked whether there were any parallels in the way in which these mutations influenced the kinetic parameters of TetX2. We observed that cells expressing $tet(X2)_{N371I}$, $tet(X2)_{N371T}$ and $tet(X2)_{T280A}$ had the fastest growth rates when compared to the wild-type stain and that all three mutations caused a 2-fold decrease in $K_{M(MCN)}$ and slight increases in $k_{cat}$, but had a much smaller effect on the $K_M$ for NADPH. The mutant $TetX2_{K64R}$ has growth rates comparable to wild-type TetX2 but causes an increase in the $K_M$ of NADPH while not affecting any other parameters. This result suggests that beneficial mutations that alter the efficiency of MCN hydrolysis have larger effects on overall fitness and that the bacteria are less sensitive to changes affecting the NADPH oxidation reaction, perhaps because the [NADPH] is always above the $K_{M(NADPH)}$ value in rapidly growing cells.

The molecular mechanisms underlying greater fitness of strains carrying $TetX2_{F235Y}$ and $TetX2_{S326I}$ could not be directly linked to the *in vitro* activity of the enzymes but rather to the steady-state protein levels in the cell. The effective concentrations of proteins *in vivo* depend on several factors including RNA and protein stability, rates of protein folding, as well as changes in protein expression patterns due to mutations in regulatory elements (Soskine and Tawfik, 2010; Tokuriki and Tawfik, 2009b). Therefore, measuring the relative steady-state protein levels from cell lysates provide a means to evaluate whether the adaptive mutations affected any of these

molecular processes and to relate them to the *in vivo* performance of the enzymes with respect to antibiotic resistance in growth assays.

We have found that relative to strains carrying wild-type TetX2, bacteria chromosomally expressing TetX2$_{F235Y}$ and TetX2$_{T280S}$ showed 2-fold higher steady-state protein levels, whereas bacteria carrying TetX2$_{N371I}$, TetX2$_{N371T}$ and TetX2$_{S326I}$ had only 50% higher protein levels. This result shows that single point mutations can cause small changes in the effective steady-state protein levels that in turn have large effects on the success of particular mutations in the population.

Overall, the *in vitro* kinetic studies show that all adaptive mutations exhibit surprisingly small changes in kinetic parameters for TetX2 enzymatic reaction in relation to the *in vivo* fitness benefits conferred by bacteria expressing these variant alleles. For example, the best mutants TetX2$_{T280A}$, TetX2$_{N371I}$ and TetX2$_{N371T}$ conferred 240% fitness benefit over TetX2 at 10 μg/ml of MCN but showed less than 30% increases in $k_{cat}$ and about 2-fold decrease in their respective $K_{M(MCN)}$ values. In addition, the efficiencies of the two catalytic steps, first FAD reduction by NADPH and second hydroxylation of MCN, were altered at the most by 5-fold (i.e. $k_{cat}/K_{M(MCN)}$, $k_{cat}/K_{M(NADPH)}$) for T280A, Table 4.2 ).

Adaptive mutations with large effects on the kinetics of TetX2 were not isolated from the adaption experiment or the error-prone library. Because most proteins are only marginally stable, mutations with greatest effects on activity are typically highly destabilizing (Romero and Arnold, 2009). Additionally, these destabilizing mutations can be costly to an organism because they often lead to loss of function and protein

misfolding which is cytotoxic and requires energy for refolding. As a consequence, more substitutions with smaller effects on either stability or activity exist (Soskine and Tawfik, 2010), and those are often selected for because they increase enzymatic efficiency and/or total activity, providing a fitness gain and playing an important role in evolutionary dynamics. In our experiments, success or failure of a particular mutant at various antibiotic concentrations was based on moderate differences in both *in vivo* steady-state protein levels and protein activity. As shown by growth assays, these subtle changes can have large consequences on the overall growth fitness of the organisms. Therefore, it is likely, that mutations with large compensatory effects on physicochemical properties of TetX2 are not observed because they are rare and don't provide net increases in enzymatic activity and thus a fitness advantage to the organism.

# Chapter 5: Structural characterization of TetX2

The prose, figures and data contained in this chapter regarding TetX2$_{wild-type}$ structure have been previously presented by Walkiewicz et al. (2011) in the journal of Proteins: Structure, Function and Bioinformatics. Quotation marks and references to this work have been omitted.

At the molecular level, it is the interplay between genetic changes and its effect on the three dimensional fold and function of a protein that lies at the basis of adaptation to a particular stressor. In order to relate mutations identified in the evolution experiment to their biophysical origins, we set off to determine high resolution structures of the wild-type TetX2 and the best adaptive mutant TetX2$_{T280A}$ in complex with minocycline (MCN) and tigecycline (TIG). The high resolution structure provides invaluable information about the atomic structure of TetX2 and reveals mechanistic details regarding mechanisms involved in substrate recognition and inactivation, which can be informative for future drug design. It addition, the atomic structure of TetX2 serves as a template for mapping the adaptive mutations.

## 5.1. Overall structure of TetX2$_{wild-type}$

The final TetX2 model consists of 4 molecules (A, B, C, D) with r.m.s.d. = 0.1 Å between copies, 84 water molecules and R and R$_{free}$ of 24.0 and 28.2 % respectively. Electron density for residues 11, 248-249 and 6 C-terminal residues was weak and could not be modeled. TetX2 shows a very low overall sequence identity with other FAD-

dependant hydroxylases (21%) (**Figure 5.1**), but shares several conserved motifs associated with formation of an FAD binding pocket and similar overall topology to the class A family of oxidoreductases. TetX2 is a monomer comprised of two domains binding one FAD molecule (**Figure 5.1.A**). The larger (N-terminal) domain is mostly involved in the interactions with the adenosine ring of FAD, while the smaller (C-terminal) domain is predicted to play a major role in substrate recognition.

GXGXXG

```
TetX2    1   MTMRIDTDKQM LSDK            PV      K  N    VS YER  DR A  F GT
PHBH     1   ---------------KT V    A P   L Q   K     I- I ER    V GRIRA
PHZS     1   ------------MSEP D L A   IG   C A      GK- T L S S I P  V
SAL      1   ---------MK NK G R         I  A    E CRYSHL - Q F CA A GE  A V

TetX2   61   D HKGS Q    KA  LQT YD   MG----- NI  E N  L --T N  F R D P
PHBH    46   G LEQ M  L   A  DRRM RD  VHEG--V A  G R  DLKRLS G TVTV
PHZS    48   N Q - A    A L  P L AT      ----H R I  S AT WSEP  EA A P Y
SAL     51   SFG -N  A  VGL   E  LQ DR SEPWE  W EWR   DA YLGAT A  --V  S

TetX2  114   E N N R     S ENDTVI R ----  VM  PGKKKWT T ENKPS T----A   
PHBH   104   E T DL E REAC ATTVYQAA      ------- H LQ RPY TF RD  LR DC Y A
PHZS   103   S    QM  A RER   QA TGLG ER   D R L GAR H KPQA GA  V
SAL    108   S  A FLDA  H PKG A RK ----ATQ  Q G  Q  A  TEH ----C

DG
TetX2  166   L N GM K  KF T TE EE G-TFN QAD HQP INCPGFFQ CNG-----NRL ASHQ
PHBH   157   C F GIS QS PA R K FERVY FGWLGLLA TPPVS E  Y    ---------- PR
PHZS   163       A HP  RP  HGGIT R TE  RF  G T  V     -----   WSR V
SAL    160       K   S  L G GLAPQ--V R S TCA RGM  SLH RE YRAQGI    VD P

TetX2  220   NL F       NN  A HFG S KT--P EWK  Q VD  Q  NSV  F   E S
PHBH   207   A CS --------- SA RSR YVQ -L K ED  D R T----- LKA LP  VA
PHZS   218    P S R-----HAAEGK     CM  -- AA G LD   D   GR   PF  
SAL    218   QMY GLDGHILTF V   R   V A SDR  PKP PA  P V VSQR   DA  G G

GD
TetX2  272    YK LIHTTL FVG A R FPLEK WKSKRPLP         F   V S  V
PHBH   253   LVTGP  EK  AP R  EP------MQ   F A        T K L L AS V
PHZS   271   LG F I D  TRNQL LQ P  RD  H GR      Y   A   S   L
SAL    278   ----AA   EC PTPTL A H LAE  G V   V        A L HQ A  A  E

TetX2  332   I SDN   -----------KFNSI  V NYE Q FI G AQ----------------
PHBH   307   ST Y  LK YR-- GR    R SAICLR IW AE  WWMTS LH FPDTD FSQ --
PHZS   331   E  AA  RN----AD  AA RE      T   ILAN  R---- E ----------
SAL    335   YF     TQVDA N    A   L     C   Q T  TGE Y FR PVV ANE LL

TetX2  366   ---------------- ESTQN IE FKP F  QQ  NV---
PHBH   363   ---------- QQT   YYL S  G AT AEN V  PYEEIE
PHZS   373   ----------- AAASRPKTEKS  E  TG  NQ ERP--
SAL    395   GENLATRFDW  SH  ADL EAR RL W NGS   RQG--
```

**Figure 5.1. TetX2 shares common sequence motifs with other FAD-dependant monooxygenases.**

Multiple sequence alignment of TetX2 from *Bacteroides thetaiotaomicron* (TetX2), p-hydroxybenzoate hydroxylase (PHBH) from *Pseudomonas fluorescence*, phzS from *Pseudomonas aeruginosa* (PHZS), and salicylate hydroxylase (SAL) from *Pseudomonas putida* shows that TetX2 shares conserved sequence motifs, GXGXXG, DG, modified GD, with other FAD-monooxygenases. Sequences were aligned with program CLASTALW and represented by BOXSHADE (Thompson et al., 1994).
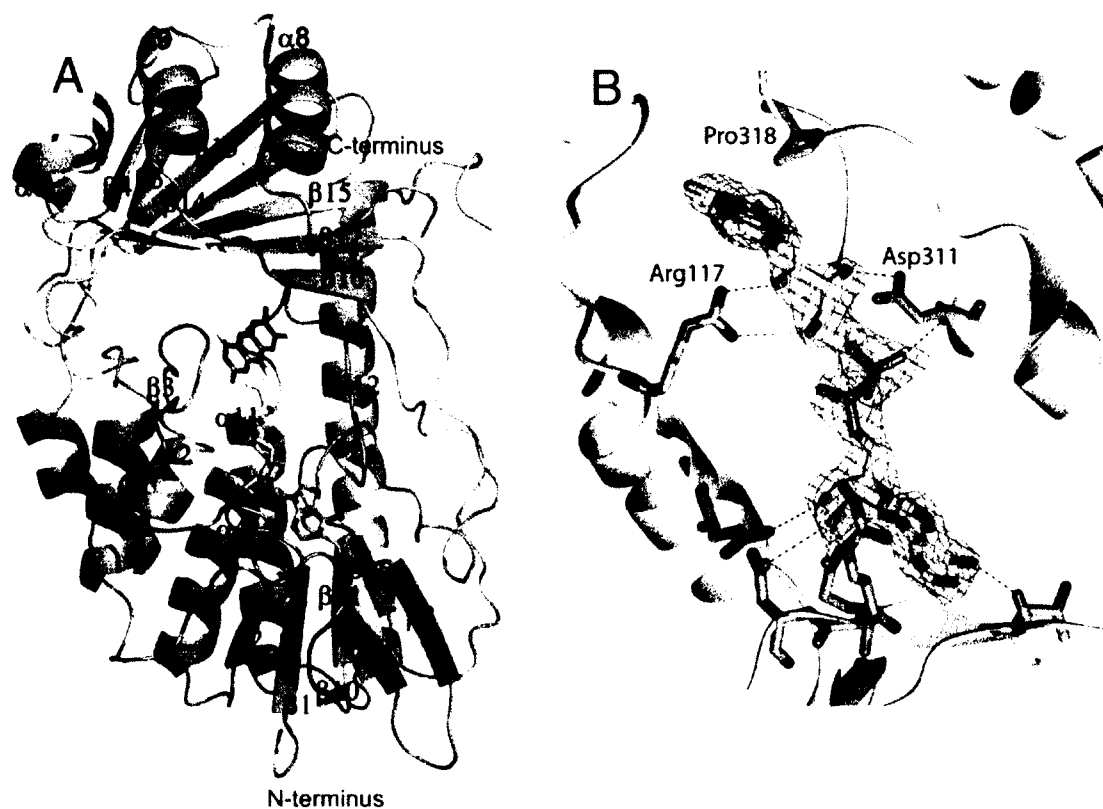
**Figure 5.2 Crystal structure of wild-type TetX2 from *Bacteroides thetaiotaomicron*, a FAD-dependent monooxygenase determined at 2.8 Å.**

A) Ribbon representation of crystal structure of TetX2 with bound FAD (yellow).B) FAD binding is coordinated through a number of hydrogen interactions with residues surrounding the FAD-binding pocket.

β1      α1    β2      β3    α2

MNLLSDKNVAI IGGGPVGLTMAKLLQQNGIDVSVYERDNDREARIFGGTLDLHKGSGQEA
13      20      30      40      50      60      70

α3      β4    β5    α5    β6      α4

MKKAGLLQTYYDLALPMGVNIADKKGNILSTKNVKPENRFDNPEINRNDLRAILLNSLEN
71      80      90      100      110      120      130

β7   β8      β9    β10   β11          β12

DTVIWDRKLVMLEPGKKKWTLTFENKPSETADLVILANGGMSKVRKFVTDTEVEETGTFN
131      140      150      160      170      180      190

α6    α7    β13    β14    β15

IQADIHQPEINCPGFFQLCNGNRLMASHQGNLLFANPNNNGALHFGISFKTPDEWKNQTQ
191      200      210      220      230      240      250

α8    α9    β16      β17

VDFQNRNSVVDFLLKKFSDWDERYKELIHTTLSFVGLATRIFPLEKPWKSKRPLPITMIG
251      260      270      280      290      300      310

α10        α11        α12

DAAHLMPPFAGQGVNSGLVDALILSDNLADGKFNSIEEAVKNYEQQMFIYGKEAQEESTQ
311      320      330      340      350      360      370

NEIEMFKPDFTFQQLLNV
371      380

**Figure 5.3. TetX2 secondary structure.**

TetX2 is composed of 30% helices and 24% β-sheets.

## 5.2. FAD binding in the structure of TetX2wild-type

FAD is stabilized in the active site through a network of hydrogen interactions with residues from both N - and C - terminal domains (**Figure 5.2.B**). In TetX2, FAD is in an extended ("IN') conformation with adenosine positioned close to the Rossmann fold and the isoalloxazine ring extending toward the smaller domain near the putative substrate binding site. The flavin binding pocket also contains a highly conserved GD motif (residues 310-311) found in most FAD-dependent oxidoreductases (Eggink et al., 1990). The residues in the GD motif form a specific pocket for the flavin molecule

through the formation of hydrogen bond interactions between the $O\delta$ of Asp311 and the

O3' of ribose, and $O_{P2}$ of the pyrophosphate of the FAD to the backbone amide of

Asp311. In TetX2, a third DG motif involved in the binding the pyrophosphate moiety of

FAD (Eppink et al., 1997), that is found in most FAD – hydroxylases, has an Asn168 in

place of Asp resulting in a slightly altered FAD binding pocket. In addition, Arg117

forms highly conserved hydrogen bonds to the ribityl chain of FAD that may orient the

flavin toward the substrate in the TetX2 active site (Greenhagen et al., 2008). TetX2

lacks an independent NADPH binding domain which is a characteristic of class A

monooxygenases (Eppink et al., 1997).


## 5.3. Putative active site in the structure of TetX2$_{wild-type}$

The interface between antiparallel β-sheets (strands $\beta_{5-6}$, $\beta_{16-20}$) of the second

domain and the isoalloxazine ring of FAD comprises the binding site for the substrate.

We observed additional strong electron density in the (2Fo-Fc) SIGMAA weighted

composite maps near the isoalloxazine ring of FAD that corresponds to the predicted

substrate binding site though the identity of the molecule remains unknown (**Figure 5.3**,

**Figure 5.4**). In the TetX2$_{wild-type}$ structure, unknown electron density was found for both

the $P2_1$ and $P_1$ (data not shown) crystal forms of TetX2. Although, we were unable to

determine the identity of the unknown molecule that gives rise to the strong density seen

in the composite electron density map, we speculate based on its position near the

isoalloxazine ring of FAD and similar location to the substrate of PHBH that this

molecule is a natural *E. coli* metabolite that co-purified with TetX2. The TetX2 structure

bound to an unknown substrate is in a 'closed' state and the strong yellow color of the crystals suggests that the flavin is oxidized and therefore unable to carry out further catalysis. This might also explain the extended conformation of FAD which resembles more the FAD conformation of PHBH with substrate bound rather than the FAD in the unbound PhzS structure (Greenhagen et al., 2008).

### 5.4. Comparison of the active sites between TetX2$_{wild-type}$ and PHBH.

Superposition of TetX2 to the well characterized PHBH (Fig. 2C) (PDB ID: 1cj3) and *Pseudomonas aeruginosa* PhzS (PDB ID: 3c96) results in an r.m.s.d. of 4.1 Å and 2.9 Å, respectively. As shown in **Figure 5.4.A**, there are substantive differences in the structure compared to PHBH in the N-terminal domain consistent with the modest sequence identity although the general topology is maintained. The overall comparison of the active sites between crystal structures of TetX2 and substrate bound PHBH shows that the FAD cofactor is oriented in a similar position in both proteins (**Figure 5.4.B**) but that the specific residues surrounding the predicted substrate binding site of TetX2 compared to PHBH are poorly conserved, which is not surprising considering that the enzymes bind significantly different substrates. TetX2 maintained a highly conserved proline (Pro318) that is analogous to Pro293 in PHBH that interacts with the phenolic oxygen of the substrate of PHBH. This residue has been suggested to be important for maintaining the proper FAD conformation (Palfey et al., 2002).

The location of the unknown density in the TetX2 structure closely resembles the position of the substrate in PHBH. Arg213 and Met215 in the TetX2 structure are in positions to make hydrogen bonds to the unknown substrate in a manner analogous to residues Ser212 and Tyr201 in PHBH. In addition, the sidechain of His234 and the backbone amide of Gly58 can hydrogen bond with the substrate similarly to Arg214 in PHBH. We speculate that Phe224 is similar to Tyr222 in PHBH, which is also important for substrate binding. A significant shift of a loop (residues 54-59) between the substrate binding site and the isoalloxazine ring of FAD is observed in TetX2 when compared to PHBH loop (residues 42-47), which results in an expanded binding pocket that might be required to accommodate a large substrate, such as tetracycline. While the natural substrate of TetX2 remains unknown, the ability of TetX2 to inactivate tetracyclines in a flavin dependent manner together with the structure presented here confirm the role of TetX2 as a Class A oxidoreductase.
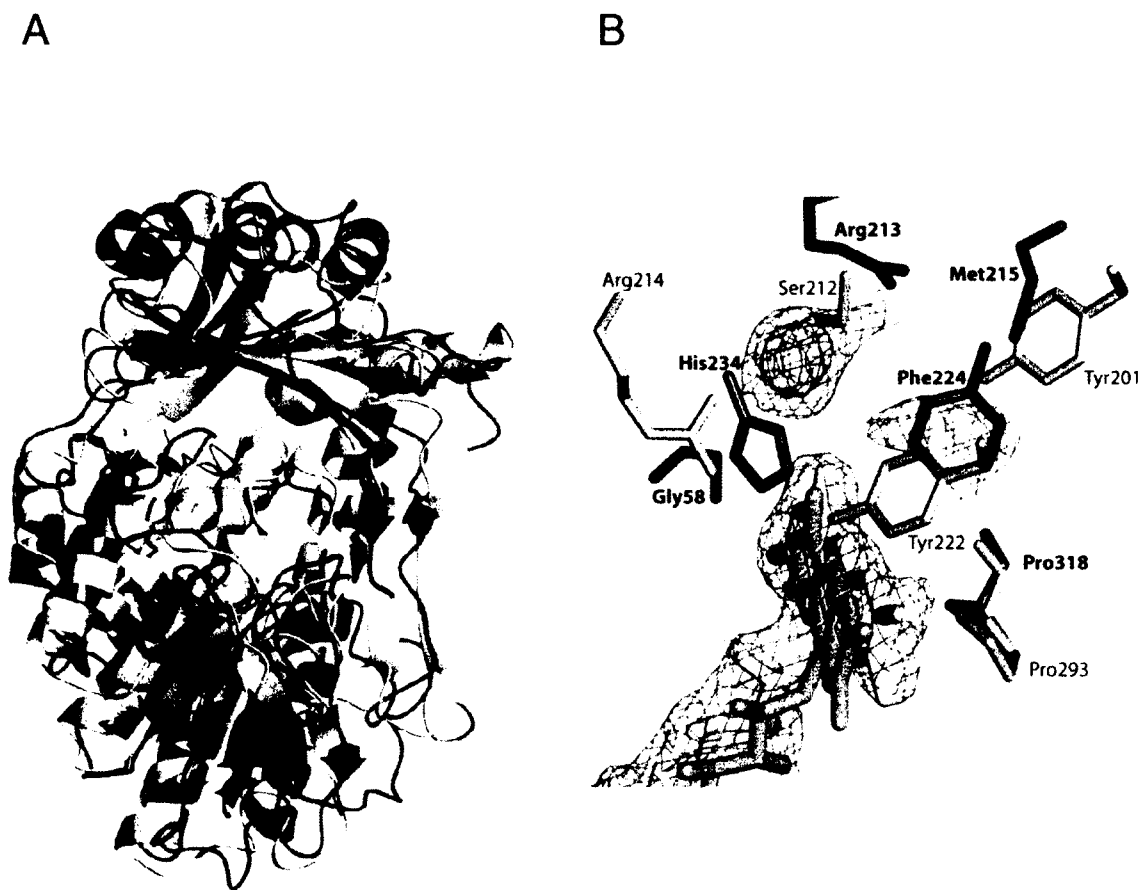
**Figure 5.4 Structural comparison of TetX2 and PHBH.**
A) Superposition of structures of TetX2 (blue) and PHBH (orange) (4.1 Å r.m.s.d.). The overall topology of TetX2 is similar to PHBH and other FAD-dependent hydroxylases, forming two domains with one FAD molecule and the absence of an independent NADHP binding domain. B) Overlay of active sites of TetX2 (blue residues) and PHBH (orange residues) with strong unknown density shown in blue and cyan. Residues surrounding the active site are not well conserved with the exception of Pro318 (Pro293 in PHBH). Final (2Fo-Fc) SIGMAA weighted electron density map corresponding to FAD and the unknown density was contoured at 1.5σ (blue) and 3σ (cyan).

## 5.5. Crystal structure of TetX2$_{T280A}$ in complex with minocycline and tigecycline

In order to better understand the role of TetX2 in tetracycline resistance and the potential structural basis for the rise of adaptive mutants isolated from the directed evolution experiment, we determined the crystal structures of the TetX2$_{T280A}$ in complex with MCN at 2.7 Å resolution (**Figure 5.5**). We compared the crystal structure of MCN:TetX2$_{T280A}$ complex to previously solved native unbound (Walkiewicz et al., 2011) and substrate bound (7-iodtetracycline,7-ITc and 7- chlortetracycline, 7-ITc) (Volkers et al., 2011) TetX2 structures. The overall structure of TetX2$_{T280A}$ is very similar to the wild-type TetX2 (r.m.s.d. = 0.34 Å; PDB ID: 3P9U) and 7-iodtetracycline - TetX2 complex structures (r.m.s.d.= 0.24 Å; PDB ID: 2Y6Q) (Volkers et al., 2011). As observed previously, FAD adopts an elongated "closed" (IN) conformation suggesting that it does not undergo conformational change upon substrate binding (Volkers et al., 2011). In a large cavity at the interface of the two domains of the protein near FAD, we identified strong electron density in (2Fo-Fc) SIGMAA weighted composite maps that corresponds to the substrate binding pocket, where we fitted MCN (**Figure 5.5.A**).

MCN was found in a similar conformation in the active site to previously described substrates 7-ITc and 7-ClTc with the hydrophilic portion of substrate oriented towards the isoalloxazine ring of FAD (Volkers et al., 2011). Despite large similarities in the substrates' orientation, the residues directly involved in MCN recognition were altered. In the structure of TetX2$_{T280A}$, the MCN interactions with the enzyme occur via hydrogen bond network formed between O2 and O5 of MCN and side chains of Q192 and R213. However, the interactions mediated by residues H234 and backbone of G236

that were important in the recognition of 7-ITc (PDB ID: 2Y6Q), were not conserved in our structure. Based on comparison of $TetX2_{T280A}$:MCN complex to $TetX2_{wild-type}$:MCN complex (unpublished data from Prof. Hinrichs laboratory, University of Greifswald, Germany) we concluded that these differences are most likely due to the slight changes in substrate properties.

The structure of $TIG:TetX2_{T280A}$ complex was also solved, however at slightly lower resolution of 2.9 Å. The electron density for TIG was visible in copies A and C and was weaker in copies D and B. The electron density for the aliphatic chain of TIG was visible only in copy A, although it was weak (**Figure 5.5.B**). This might be due to lower resolution, but also because this portion of TIG molecule is solvent exposed and might be more disordered. The orientation of TIG in the active site very closely resembles the position of MCN and the other two substrates. Nevertheless, the specific residues involved in substrate recognition seem to be slightly altered. Based on our $TIG:TetX2_{T280A}$ complex structure, it appears that TIG is stabilized in the binding pocket through interactions between D-ring of TIG and backbone carbonyl oxygen of Ala320, as well as A-ring amine of TIG and the sidechain of Thr59 (OG). Again, it is difficult to conclude whether the differences in substrate recognition are due to specific properties of this substrate or because of the resolution limits of this crystal structure.

The orientation of both MCN and TIG towards the FAD hydroxylation site (C4a) appears to be conserved among both of our substrates as well as 7-ClTc and 7-ITc, and is achieved by interactions of N5 and O4 of FAD with O7 and O6 of MCN and TIG. In addition, binding of the substrates is stabilized by hydrophobic interaction of residues

M215, F224, M375, F380 and F382 with the hydrophobic portion on MCN (**Figure 5.6**).

These residues form a hydrophobic cavity which was speculated to constitute the putative

oxygen binding site (Volkers et al., 2011).
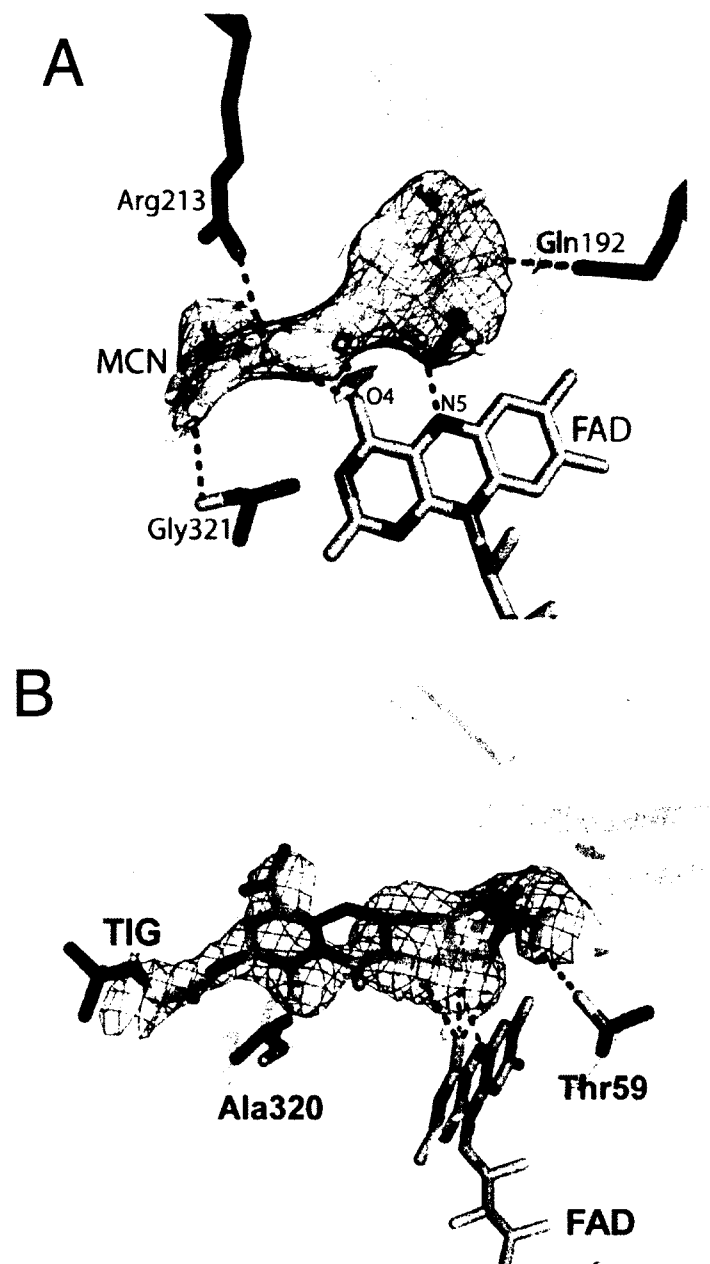


**Figure 5.5 Active site of TetX2 in complex with MCN and TIG.**
The residues involved in MCN (yellow) and TIG (cyan) recognition are shown in green.
The final SIGMAA-weighted electron density maps were contoured at 1σ.
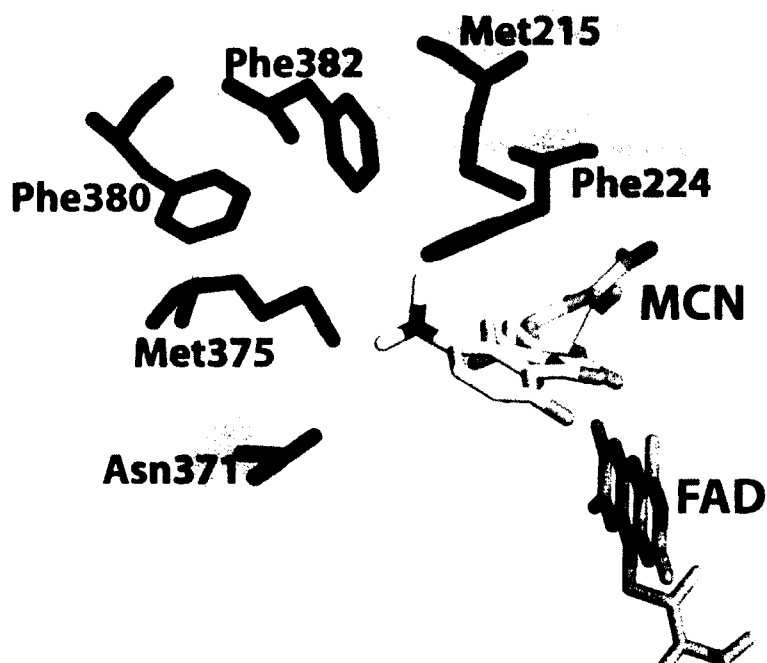
**Figure 5.6. Active site with the hydrophobic residues shown in purple.**
Residues surrounding the substrate binding pocket form a non-polar cavity which forms a putative oxygen binding site.

T280A is positioned on the second (smaller) domain of TetX2 that is mainly involved in substrate recognition (**Figure 5.8**, red sphere). Precisely, A280 is located on a loop (residues 280-284) near α-helix 9 (residues 272-279) and is ~18 Å away from the active site of the enzyme. In the wild-type structure, the backbone (N) and side chain (OG) of T280 interact with the backbone carbonyl of E276 on the same helix and

backbone carbonyl of N201 on the opposite helix (**Figure 5.6**). In the mutant, although

the interaction with N201 and side chain (OG) of residue 280 is lost, the backbone amine

of AT280 hydrogen bonds with carboxyl oxygens of E276 and L277 on the same helix.

In addition, the side chain (CB) of A280 interacts with the hydrophobic core made of

residues I195, F235 and L277. Structural analysis clearly shows that residue at position

280 is not directly involved in the catalytic mechanism of the enzyme. This strongly

suggests that T280A, rather than having local effects on enzyme activity, has a more

global effect.
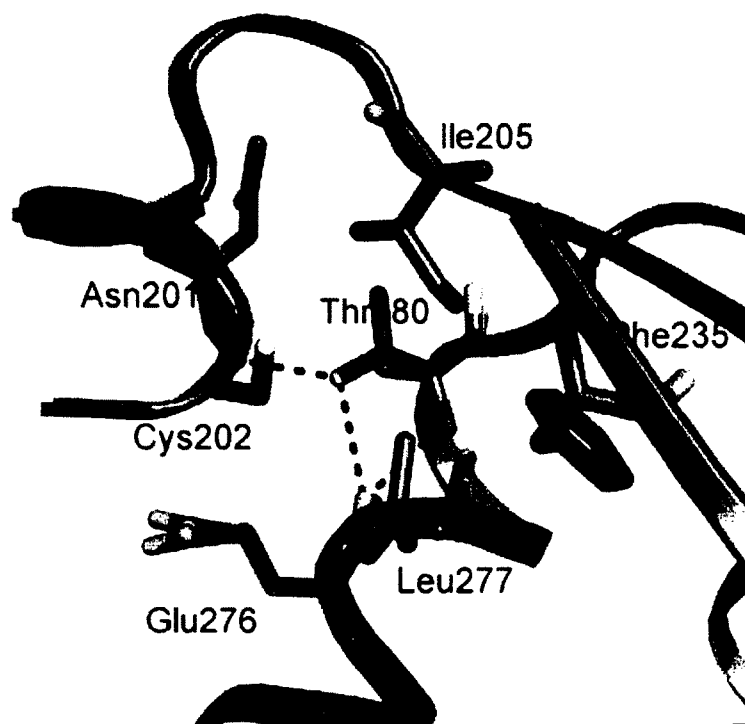


**Figure 5.7 Overlay of crystal structures of TetX2_wild-type and TetX2_T280A.**
Crystal structure of TetX2_T280A (red) does not show significant differences in the overall structure of the enzyme. In the wild-type structure (cyan), the side chain (OG) of T280 interacts with backbone carbonyl of N201 and L277. In the mutant, the side chain interactions are lost, but the sidechain (CB) of Ala280 interacts with hydrophobic core made of residues I205, F235 and L277.

## 5.6. Library mutants mapped onto the crystal structure of TetX2$_{T280A}$

Mapping of the mutations onto the crystal structure of TetX2$_{T280A}$:MCN complex (**Figure 5.8**), showed that all of the mutated residues, except K64R and S326, are located on the second domain of the protein which is implicated largely in substrate recognition. Besides the previously described T280A substitution, which was also isolated from the directed evolution library, another mutation was identified at position 280, where threonine was replaced by serine (**Figure 5.8**, shown in red). Since serine and threonine are similar in size, both have hydroxyl group at the side chains and are located about 18 Å away from the active site, we might expect that this conservative change has a minor effect on substrate binding and enzyme kinetics.

The mutation found at position 235 (**Figure 5.8**, shown in purple), where Phe was substituted to Tyr, is also located on the second domain of TetX2, near the substrate binding pocket cavity on β-sheet 15. While, the side chain of F235 is oriented about 180° away from the substrate and it makes van der Waal contacts with hydrophobic residues A193, F205, I237, L277, and F284 in the adaptive mutant F235Y these residues must now accommodate a polar hydroxyl group (**Figure 5.9.A**). As the backbone of Cα of F235 is ~5Å away from the A-ring of MCN, changes to accommodate the F to Y mutation may alter main chain positioning.

In contrast, the side chain of N371 is oriented directly towards the D-ring of MCN (~4.7 Å) (**Figure 5.8**, green sphere, **Figure 5.9.B**)) at the putative entrance site for tetracyclines. Substitution of a non-polar residue (I) or a shorter polar side chain (T) at position 371 near the apolar pocket (**Figure 5.6.**) results in smaller $K_{M(MCN)}$ and modest

increases in $k_{cat}$. In addition, position 371 is oriented at a hydrophobic cavity which was speculated to constitute the putative oxygen binding site for TetX2 and thus changes in this area may have effects on the utilization of oxygen for catalysis (Volkers et al., 2011).

S326I (**Figure 5.8**, shown in yellow) is located on α-helix 11 parallel to the ribose portion of FAD. The non-polar side chain of Ile is perhaps well tolerated since it can interact with hydrophobic side chains of M316 and A364 on α-helix 12, and thus stabilize the structure at this position (**Figure 5.9.C**).

Finally, K64R (**Figure 5.8**, shown in blue, **Figure 5.9.D**) is located on a loop at the surface of the molecule. It is difficult to predict the effect of this very conservative mutation on catalytic mechanism of the enzyme; however, one possibility is that the loop moves as MCN or NADPH enter the substrate binding cavity. Although there is no data revealing the NADPH binding site, a previous study hypothesized based on comparison to PHBH that NADPH binding occurs opposite to the MCN end of the active site (near R54) (Volkers et al., 2011).

**Figure 5.8. Ribbon representation of TetX2$_{T280A}$ structure with mutation sites shown in spheres and FAD shown in yellow.**

The spheres represent the following residues: red 280, green 371, blue 64, yellow 326, and purple 235. Two mutations, N371I and F235Y are located directly in the active site of TetX2$_{T280A}$.
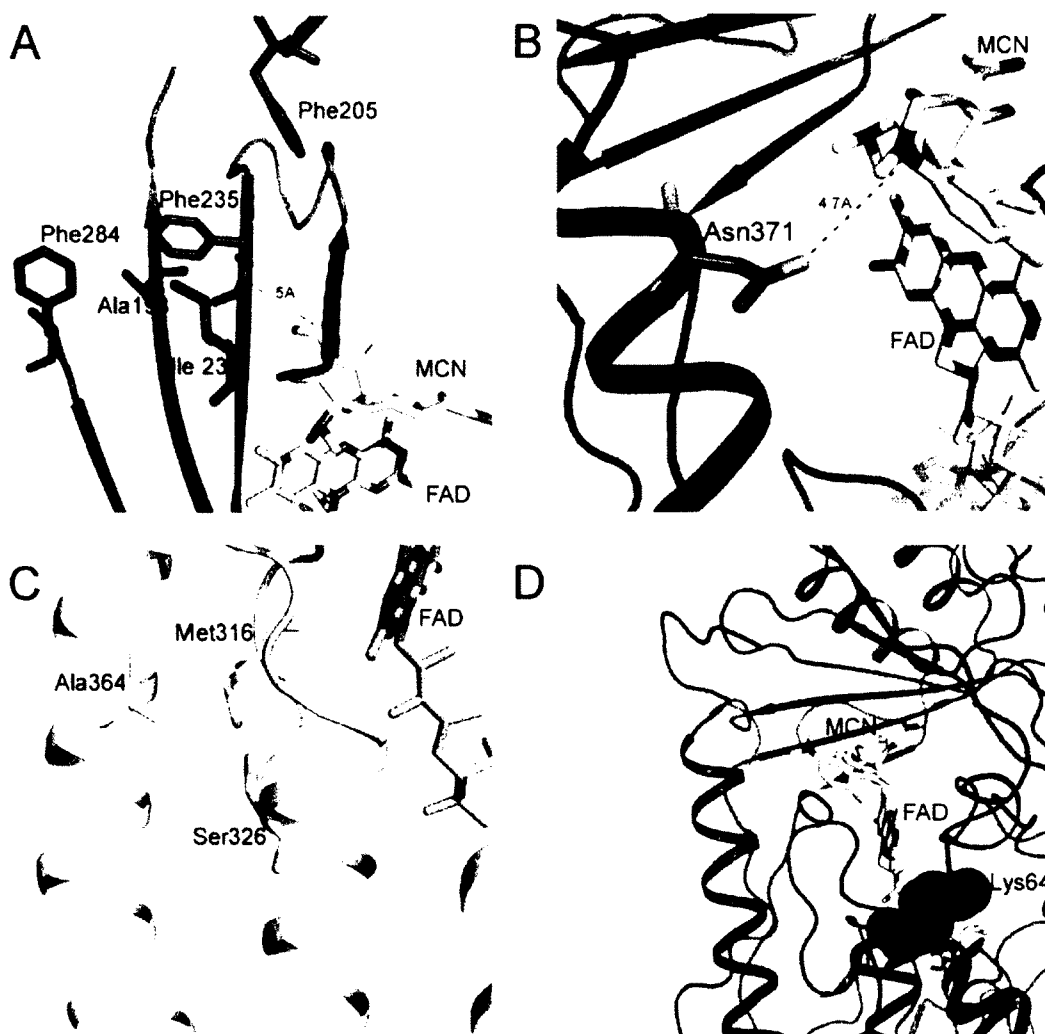
**Figure 5.9. Adaptive mutation sites mapped onto the crystal structures of TetX2$_{T280A}$:MCN complex.**

A) F235 is one of the two mutated residues located near the MCN binding site, approximately 5Å away from its backbone Cα. Mutation at this position to Y is likely to result in local changes since F235 interacts with in the hydrophobic core made of A193, F205, I237, L277 and F284. B) The side chain of Asn371 is located approximately 4.7Å from ring D of MCN, suggesting that shorter residues (T) or non-polar (I) could have a direct effect on the interaction with MCN. C) S326 is located away from the active site, but on a helix involved in stabilizing the FAD molecule. Substitution to I at position 325 is most likely well tolerated since this residue would be able to interact with M316 and A364. D) K64 is at the surface of the protein located on the loop at the putative MCN entry to the active site.

## 5.7. Discussion

Analysis of the native TetX2 crystal structure shows that TetX2 is a class A FAD - dependent monooxygenase, and, like its homologues, contains conserved GXGXXG, GD and modified DG motifs that are part of FAD binding pocket (van Berkel et al., 2006). FAD is in an extended conformation with adenosine positioned close to the Rossmann fold and the isoalloxazine ring extending toward the smaller domain near the putative substrate binding site. As previously predicted, TetX2 lacks an independent NADPH binding domain which is a characteristic of class A monooxygenases and is consistent with the hypothesis that NADPH is immediately released upon FAD reduction (Eppink et al., 1997).

Additional electron density was observed near the isoalloxazine ring of FAD for both the $P2_1$ and $P_1$ (data not shown) crystal forms of TetX2. Interestingly, the additional electron density was located in the putative active site of the enzyme and fits with the prediction for the substrate binding site based on comparison of active sites for TetX2, PHBH, and PhzS. We speculate that TetX2 crystallized in the presence of a substrate analogue that might be a natural *E. coli* metabolite. The TetX2 structure bound to an unknown substrate is in a 'closed' state and the strong yellow color of the crystals suggests that the flavin is oxidized and therefore unable to carry out further catalysis. This might also explain the extended conformation of FAD which resembles more the FAD conformation of PHBH with substrate bound rather than the FAD in the unbound PhzS structure (Greenhagen et al., 2008).

Comparison of the active sites between TetX2 and substrate bound PHBH shows little similarity between the residues that might be involved in substrate binding. One exception is Pro318 which is highly conserved among most of FAD-dependent hydroxylases and plays a role in substrate binding and orienting the flavin in the conformation required for catalysis (Palfey et al., 2002). The low conservation of residues in the active site is not surprising considering the large diversity of substrates that these enzymes act upon, especially TetX2 that modifies tetracyclines which are larger in size.

To gain a better understanding of the mechanisms underlying the increase in fitness of the TetX2$_{T280A}$ enzyme, we solved its crystal structure in complex with MCN and TIG. Determination of high resolution structures of TetX2$_{T280A}$ was critical for our studies because it served as a template for mapping the mutations of the altered proteins and it provided insights regarding the mechanistic details of its catalytic mechanism.

Perhaps, the most surprising finding was that there were no significant differences between crystal structures of native TetX2 and substrate bound mutants. T280A is positioned on a loop (residues 280-284) connecting $\alpha$-helix 9 (residues 272-279) and $\beta$-sheet 20 (residues 285-288) and is ~18 Å away from the active site of the enzyme. From its location on the three dimensional structure it is evident that residue at this position is not directly involved in the catalytic mechanism of the enzyme. Although there are many cases where the adaptive mutations occur at the active site of the enzymes, it is not uncommon to find mutations that alter activity or stability of an enzyme but are distal from the active site (Counago et al., 2006). Nevertheless, this finding strongly suggests

that T280A, rather than having local effects on enzyme activity, has a more global effect. Preliminary molecular dynamics studies on the T280A mutant support a role for the mutation in altering protein dynamics (not shown).

The position of the mutant residues in the three dimensional structure of TetX2$_{T280A}$ in relationship to the substrate cavity together with changes in different kinetics parameters of the enzymes suggest widely varying mechanisms of adaptation through altered enzymatic activity and *in vivo* stability. The mutations isolated through the experimental evolution mapped mostly (except K64R and S326I) to the smaller domain of the enzyme that encompasses the active site for the reduction and hydroxylation steps in catalysis. However, only residues at position 235 (**Figure 5.7**, violet sphere, TetX2$_{F235Y}$) and 371 (**Figure 5.7**, green sphere, TetX2$_{N371I}$, and TetX2$_{N371T}$) are within 5Å of MCN and likely to be directly involved in substrate binding or catalysis. The remainder of the mutants appears to perturb the function of TetX2 via indirect mechanisms, perhaps affecting the flexibility of the enzyme. Wang et al. have shown that dynamics of the related PHBH protein play a critical role in its catalysis (Wang et al., 2002). As described previously, while TetX2 does not have a distinct NADPH binding domain, it is evident that the active site must undergo significant conformational changes to allow the entry and exit of the large NADPH molecule to reduce flavin and allow subsequent binding and modification of tetracyclines.This is not the first time that altered flexibility and catalysis were directly linked to antibiotic resistance. A recent study of metallo-β-lactamases showed that flexibility is an essential trait that natural selection acts upon (Tomatis et al., 2008).

Even though TetX2 requires magnesium for its activity, electron density for magnesium was not observed in the wild-type TetX2 or the complex structures shown by us and others (Volkers et al., 2011). Magnesium has been shown to be essential for binding of tetracyclines to the ribosome (Chopra and Roberts, 2001). Crystal structures of tetracycline bound to the 30S ribosomal subunit show that the magnesium ion is bound by oxygen atoms at carbons 11 and 12 of tetracycline (11a is the hydroxylation site) and oxygen atoms from the backbone of the ribosome (Brodersen et al., 2000; Pioletti et al., 2001). Why magnesium is required for tetracycline degradation is puzzling and as of today its specific role in the catalysis of tetracyclines antibiotics mediated by TetX2 remains unknown.

# Chapter 6: Summary of Findings and Future Directions

In the work presented here, experimental evolution and biophysical characterization were used to identify and link the physicochemical properties of TetX2 and the adaptive mutations to increased resistance to MCN. The success of TetX2 mutants during adaptation to MCN is reflected by moderate alterations in *in vitro* catalytic activity and *in vivo* steady-state protein levels. Our results clearly indicated that genetic changes resulting in very subtle alterations in the molecular properties of enzymes (e.g. 2-fold changes in $K_M$ or expression) can have large effects on fitness and thereby play an important role in antibiotic resistance.

Another interesting finding is that adaptation of TetX2$_{T280A}$ broadens specificity of the enzyme such that resistance has increased to another newer tetracycline derivative, tigecycline, which was not used in the experimental selection. It has been previously demonstrated that protein adapting to one substrate can evolve 'promiscuous' function, such as binding of new substrates (Aharoni et al., 2005; Tahlan et al., 2007; Tracewell and Arnold, 2009). The ability to perform a new function while maintaining activity towards the original substrate reflects on evolvability and robustness of the enzyme (Romero and Arnold, 2009). In addition, 'promiscuous' function even if several orders of magnitude lower provides an opportunity for new adaptations (Arnold, 2009).

Kinetic studies showed that TetX2$_{T280A}$ was the most enzymatically active mutant with the lowest $K_M$ for both NADPH and MCN suggesting that this enzyme was most efficient at low MCN concentrations and explains its success in the adaptation

experiment. Although, more detailed kinetic studies are required to accurately elucidate the mechanisms of MCN hydrolysis by TetX2, this is the first study that provides evidence for a binary complex mechanism. The kinetic parameters determined from these experiments show that TetX2 forms a binary complex with NADPH which dissociates from the active site prior to second binary complex formation between reduced enzyme and MCN. This finding was supported by crystallographic studies on $TetX2_{wild-type}$ and the best adaptive mutant, $TetX2_{T280A}$, in complex with MCN or TIG. Currently, the NADPH recognition and binding site with TetX2 have not been elucidated. Crystallographic studies of TetX2 in complex with NADPH would provide invaluable insight into this reaction, especially that the kinetic mechanism seems to differ from PHBH.

Although, during the course of this study, crystal structures of TetX2 in complex with two different substrates were solved (Volkers et al., 2011), the complex structures with tigecycline presented in this thesis represents the first high resolution protein structure with this novel antibiotic. TIG remains a very effective tetracycline derivative. Due to increased size and decrease in polarity efflux pumps are ineffective in transporting tigecycline out of the cell. Analysis of the TIG binding pocket in $TetX2_{T280A}$ shows that the aliphatic portions of TIG, which is the unique modification on this tetracycline derivative, is solvent exposed and does not make specific contacts with the protein. In fact, most of the interactions of TetX2 with its substrates are mediated through the conserved portions of tetracycline antibiotics, it is therefore not surprising that tigecycline and other tetracycline derivative can be accommodated in the TetX2 active site. This finding implies that even though TetX2 has not been identified in clinical settings, it still

has the potential to provide very broad resistance to a new generation of pathogens. Crystal structures presented here can therefore provide guidelines and assist in the design of novel tetracycline derivatives that would serve as poor substrates for this novel mechanism of resistance.

The success of TetX2$_{T280A}$ was predicted by experimental evolution; however, the *in vivo* performance in the presence of MCN of two other mutations also predicted their success. Whether clonal interference, transition bias or introduction of a bottleneck played a role in the outcomes of the natural evolution experiment remains an open question. Deep sequencing using DNA barcoding to quantitate the allelic frequency of the seven most successful *in vitro* *tet(X2)* mutants in replicate populations should address this question and assess the role for mutation supply in the evolutionary dynamics of this system. We speculate that if TetX2$_{T280A}$, TetX2$_{N371I}$ and TetX2$_{N371T}$ have similar chances for success, we expect to observe equal frequencies of each of the mutants within populations. Although this experiment will answer the question regarding the allelic frequencies in the evolving population, it will not reveal the frequencies of mutations in the starting populations. Alternatively, we may find that TetX2$_{T280A}$ is indeed the most successful. If so, this would suggest an important role for the $K_M$ to NADPH at the low concentrations of MCN.

To assess the background genetic variation, another experiment would have to be performed where *E. coli* population expressing *tet(X2)* would be grown in a continuous culture (e.g. turbidostat) rather than flasks. Performing deep sequencing analysis on samples derived from a turbidostat allows propagating a larger population for many

generations in a short time with the benefit of decreasing the effects of bottlenecks. Deep sequencing of samples derived from these experiments would allow determining the allelic frequencies in the starting population.

The ultimate goal of this project was to test the hypothesis that based on *in vitro* physicochemical properties of potential adaptive mutants; we can build a mathematical model that accurately predicts the evolutionary outcomes in populations during adaptation to antibiotics. Currently, Andres Benitez (graduate student in Dr. Yousif Shamoo's laboratory) is developing a mathematical model to quantitatively describe the success and failure of each adaptive TetX2 mutation over a range of MCN concentrations. The model incorporates the classic Michaelis-Menten kinetics ($K_{M(MCN)}$, $K_{M(NADPH)}$ and $k_{cat}$). In addition, it uses a hill function determined by the inhibition of MCN on growth rates of *E. coli* BW25113 that is transformed onto the growth rates of the TetX2 variants using the steady state equation and fitted using measured kinetic parameters as well as three unknown parameters; the diffusion rate of antibiotic, NADPH concentration, and the concentration of active protein for each of the individual mutants. If a model that accurately predicts the success of TetX2 mutants is developed will it be applicable to other systems? This idea could be tested on another bacterial system that utilized antibiotics by enzymatic inactivation such as chloramphenicol acetyltransferase which is known to confer resistant to choramphenicol.

# Bibliography

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Aharoni, A., Gaidukov, L., Khersonsky, O., Mc, Q.G.S., Roodveldt, C., and Tawfik, D.S. (2005). The 'evolvability' of promiscuous protein functions. Nat Genet *37*, 73-76.

Alekshun, M.N., and Levy, S.B. (2007). Molecular mechanisms of antibacterial multidrug resistance. Cell *128*, 1037-1050.

Andersson, D.I., and Hughes, D. (2010). Antibiotic resistance and its cost: is it possible to reverse resistance? Nat Rev Microbiol *8*, 260-271.

Arjan, J.A., Visser, M., Zeyl, C.W., Gerrish, P.J., Blanchard, J.L., and Lenski, R.E. (1999). Diminishing returns from mutation supply rate in asexual populations. Science (New York, NY *283*, 404-406.

Arnold, F.H. (2009). How proteins adapt: lessons from directed evolution. Cold Spring Harb Symp Quant Biol *74*, 41-46.

Barlow, M., and Hall, B.G. (2003). Experimental prediction of the natural evolution of antibiotic resistance. Genetics *163*, 1237-1241.

Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E., and Kim, J.F. (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature *461*, 1243-1247.

Betancourt, A.J., and Bollback, J.P. (2006). Fitness effects of beneficial mutations: the mutational landscape model in experimental evolution. Current opinion in genetics & development *16*, 618-623.

Bloom, J.D., Labthavikul, S.T., Otey, C.R., and Arnold, F.H. (2006). Protein stability promotes evolvability. Proceedings of the National Academy of Sciences of the United States of America *103*, 5869-5874.

Bloom, J.D., Silberg, J.J., Wilke, C.O., Drummond, D.A., Adami, C., and Arnold, F.H. (2005). Thermodynamic prediction of protein neutrality. Proceedings of the National Academy of Sciences of the United States of America *102*, 606-611.

Brodersen, D.E., Clemons, W.M., Jr., Carter, A.P., Morgan-Warren, R.J., Wimberly, B.T., and Ramakrishnan, V. (2000). The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit. Cell *103*, 1143-1154.

Bull, J., and Wichman, H. (1998). A revolution in evolution. Science (New York, NY *281*, 1959.

Bulmer, M. (1986). Neighboring base effects on substitution rates in pseudogenes. Mol Biol Evol *3*, 322-329.

Burdett, V. (1991). Purification and characterization of Tet(M), a protein that renders ribosomes resistant to tetracycline. J Biol Chem *266*, 2872-2877.

Camps, M., Herman, A., Loh, E., and Loeb, L.A. (2007). Genetic constraints on protein evolution. Crit Rev Biochem Mol Biol *42*, 313-326.

Chopra, I., and Roberts, M. (2001). Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. Microbiol Mol Biol Rev *65*, 232-260 ; second page, table of contents.

Cohen, M.L. (2000). Changing petterns of infectious disease. Nature *406*, 762-767.

Connell, S.R., Tracz, D.M., Nierhaus, K.H., and Taylor, D.E. (2003). Ribosomal protection proteins and their mechanism of tetracycline resistance. Antimicrobial agents and chemotherapy *47*, 3675-3681.

Conrad, T.M., Lewis, N.E., and Palsson, B.O. (2011). Microbial laboratory evolution in the era of genome-scale science. Molecular systems biology *7*, 509.

Cooper, V.S., and Lenski, R.E. (2000). The population genetics of ecological specialization in evolving Escherichia coli populations. Nature *407*, 736-739.

Counago, R., Chen, S., and Shamoo, Y. (2006). In vivo molecular evolution reveals biophysical origins of organismal fitness. Molecular cell *22*, 441-449.

Counago, R., and Shamoo, Y. (2005). Gene replacement of adenylate kinase in the gram-positive thermophile Geobacillus stearothermophilus disrupts adenine nucleotide homeostasis and reduces cell viability. Extremophiles *9*, 135-144.

Counago, R., Wilson, C.J., Pena, M.I., Wittung-Stafshede, P., and Shamoo, Y. (2008). An adaptive mutation in adenylate kinase that increases organismal fitness is linked to stability-activity trade-offs. Protein Eng Des Sel *21*, 19-27.

D'Costa, V.M., King, C.E., Kalan, L., Morar, M., Sung, W.W., Schwarz, C., Froese, D., Zazula, G., Calmels, F., Debruyne, R., *et al.* (2011). Antibiotic resistance is ancient. Nature *477*, 457-461.

Darwin, C. (1859). On the Origin of Species (John Murray).

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America 97, 6640-6645.

Dean, A.M., Dykhuizen, D.E., and Hartl, D.L. (1986). Fitness as a function of beta-galactosidase activity in Escherichia coli. Genet Res 48, 1-8.

Dean, A.M., and Thornton, J.W. (2007). Mechanistic approaches to the study of evolution: the functional synthesis. Nature reviews 8, 675-688.

DePristo, M.A., Weinreich, D.M., and Hartl, D.L. (2005). Missense meanderings in sequence space: a biophysical view of protein evolution. Nature reviews 6, 678-687.

Doublie, S. (1997). Preparation of selenomethionyl proteins for phase determination. Methods Enzymol 276, 523-530.

Eggink, G., Engel, H., Vriend, G., Terpstra, P., and Witholt, B. (1990). Rubredoxin reductase of Pseudomonas oleovorans. Structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints. Journal of molecular biology 212, 135-142.

Elena, S.F., Cooper, V.S., and Lenski, R.E. (1996). Punctuated evolution caused by selection of rare beneficial mutations. Science (New York, NY 272, 1802-1804.

Elena, S.F., and Lenski, R.E. (2003). Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nature reviews 4, 457-469.

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

Entsch, B., Ballou, D.P., Husain, M., and Massey, V. (1976). Catalytic mechanism of p-hydroxybenzoate hydroxylase with p-mercaptobenzoate as substrate. J Biol Chem 251, 7367-7369.

Entsch, B., and van Berkel, W.J. (1995). Structure and mechanism of para-hydroxybenzoate hydroxylase. Faseb J 9, 476-483.

Eppink, M.H., Schreuder, H.A., and Van Berkel, W.J. (1997). Identification of a novel conserved sequence motif in flavoprotein hydroxylases with a putative dual function in FAD/NAD(P)H binding. Protein Sci 6, 2454-2458.

Gerrish, P.J., and Lenski, R.E. (1998). The fate of competing beneficial mutations in an asexual population. Genetica 102-103, 127-144.

Goldsmith, M., and Tawfik, D.S. (2009). Potential role of phenotypic mutations in the evolution of protein expression and stability. Proceedings of the National Academy of Sciences of the United States of America 106, 6197-6202.

Gould, S.J. (1989). Wonderful life:The burgess shale and the nature of history (New York, Norton).

Greenhagen, B.T., Shi, K., Robinson, H., Gamage, S., Bera, A.K., Ladner, J.E., and Parsons, J.F. (2008). Crystal structure of the pyocyanin biosynthetic protein PhzS. Biochemistry 47, 5281-5289.

Hajibabaei, M., Singer, G.A., Hebert, P.D., and Hickey, D.A. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet 23, 167-172.

Hall, B.G. (2004). Predicting the evolution of antibiotic resistance genes. Nat Rev Microbiol 2, 430-435.

Kacser, H., and Burns, J.A. (1981). The molecular basis of dominance. Genetics 97, 639-666.

Lang, G.I., Botstein, D., and Desai, M.M. (2011). Genetic variation and the fate of beneficial mutations in asexual populations. Genetics 188, 647-661.

Lee, D.H., and Palsson, B.O. (2010). Adaptive evolution of Escherichia coli K-12 MG1655 during growth on a Nonnative carbon source, L-1,2-propanediol. Applied and environmental microbiology 76, 4158-4168.

Lee, H.H., Molla, M.N., Cantor, C.R., and Collins, J.J. (2010). Bacterial charity work leads to population-wide resistance. Nature 467, 82-85.

Lehninger, A., Nelson, D.L., and Cox, M.M. (2008). Lehninger principles of biochemistry, 5 edn (New York, W. H. Freeman and Company).

Lenski, R.E., and Travisano, M. (1994). Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. Proceedings of the National Academy of Sciences of the United States of America 91, 6808-6814.

Levy, S.B., and Marshall, B. (2004). Antibacterial resistance worldwide: causes, challenges and responses. Nature medicine 10, S122-129.

Li, W. (1997). Molecular evolution. In Sunderland, Sinauer Associates, Inc.

Long, F., Vagin, A.A., Young, P., and Murshudov, G.N. (2008). BALBES: a molecular-replacement pipeline. Acta Crystallogr D Biol Crystallogr 64, 125-132.

Massey, V. (1994). Activation of molecular oxygen by flavins and flavoproteins. J Biol Chem *269*, 22459-22462.

McRee, D.E. (1999). XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. J Struct Biol *125*, 156-165.

Moore, I.F., Hughes, D.W., and Wright, G.D. (2005). Tigecycline is modified by the flavin-dependent monooxygenase TetX. Biochemistry *44*, 11829-11835.

Novella, I.S., Duarte, E.A., Elena, S.F., Moya, A., Domingo, E., and Holland, J.J. (1995). Exponential increases of RNA virus fitness during large population transmissions. Proceedings of the National Academy of Sciences of the United States of America *92*, 5841-5844.

Ochman, H. (2003). Neutral mutations and neutral substitutions in bacterial genomes. Mol Biol Evol *20*, 2091-2096.

Orencia, M.C., Yoon, J.S., Ness, J.E., Stemmer, W.P., and Stevens, R.C. (2001). Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. Nat Struct Biol *8*, 238-242.

Orencia, M.C., Yoon, J.S., Ness, J.E., Stemmer, P.C., Stevens, R.C. (2001). Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. Nature Structural Biology *8*, 238-242.

Orr, H.A. (2003). The distribution of fitness effects among beneficial mutations. Genetics *163*, 1519-1526.

Otwinowski, Z.a.M., W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods in Enzymology *276*, 307-326.

Palfey, B.A., Basu, R., Frederick, K.K., Entsch, B., and Ballou, D.P. (2002). Role of protein flexibility in the catalytic cycle of p-hydroxybenzoate hydroxylase elucidated by the Pro293Ser mutant. Biochemistry *41*, 8438-8446.

Pena, M.I., Davlieva, M., Bennett, M.R., Olson, J.S., and Shamoo, Y. (2010). Evolutionary fates within a microbial population highlight an essential role for protein folding during natural selection. Molecular systems biology *6*, 387.

Pflugrath, J.W. (1999). The finer things in X-ray diffraction data collection. Acta Crystallogr D Biol Crystallogr *55*, 1718-1725.

Pioletti, M., Schlunzen, F., Harms, J., Zarivach, R., Gluhmann, M., Avila, H., Bashan, A., Bartels, H., Auerbach, T., Jacobi, C., *et al.* (2001). Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3. The EMBO journal *20*, 1829-1839.

Poelwijk, F.J., Kiviet, D.J., Weinreich, D.M., Tans, S.J. (2007). Empirical fitness landscapes reveal accesible evolutionary paths. Nature *445*, 383-386.

Romero, P.A., and Arnold, F.H. (2009). Exploring protein fitness landscapes by directed evolution. Nat Rev Mol Cell Biol *10*, 866-876.

Rosenberg, S.M., and Hastings, P.J. (2003). Microbiology and evolution. Modulating mutation rates in the wild. Science (New York, NY *300*, 1382-1383.

Salverda, M.L., Dellus, E., Gorter, F.A., Debets, A.J., van der Oost, J., Hoekstra, R.F., Tawfik, D.S., and de Visser, J.A. (2011). Initial mutations direct alternative pathways of protein evolution. PLoS Genet *7*, e1001321.

Skopalik, J., Anzenbacher, P., and Otyepka, M. (2008). Flexibility of human cytochromes P450: molecular dynamics reveals differences between CYPs 3A4, 2C9, and 2A6, which correlate with their substrate preferences. J Phys Chem B *112*, 8165-8173.

Smith, J.M. (1970). Natural selection and the concept of a protein space. Nature *225*, 563-564.

Soskine, M., and Tawfik, D.S. (2010). Mutational effects and the evolution of new protein functions. Nature reviews *11*, 572-582.

Spahn, C.M., Blaha, G., Agrawal, R.K., Penczek, P., Grassucci, R.A., Trieber, C.A., Connell, S.R., Taylor, D.E., Nierhaus, K.H., and Frank, J. (2001). Localization of the ribosomal protection protein Tet(O) on the ribosome and the mechanism of tetracycline resistance. Molecular cell *7*, 1037-1045.

Speer, B.S., Bedzyk, L., and Salyers, A.A. (1991). Evidence that a novel tetracycline resistance gene found on two Bacteroides transposons encodes an NADP-requiring oxidoreductase. Journal of bacteriology *173*, 176-183.

Speer, B.S., Shoemaker, N.B., and Salyers, A.A. (1992). Bacterial resistance to tetracycline: mechanisms, transfer, and clinical significance. Clinical microbiology reviews *5*, 387-399.

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. Proceedings of the National Academy of Sciences of the United States of America *103*, 8060-8065.

Tahlan, K., Ahn, S.K., Sing, A., Bodnaruk, T.D., Willems, A.R., Davidson, A.R., and Nodwell, J.R. (2007). Initiation of actinorhodin export in Streptomyces coelicolor. Molecular microbiology *63*, 951-961.

Thaker, M., Spanogiannopoulos, P., and Wright, G.D. (2010). The tetracycline resistome. Cell Mol Life Sci *67*, 419-431.

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 research *22*, 4673-4680.

Tokuriki, N., Stricher, F., Serrano, L., and Tawfik, D.S. (2008). How protein stability and new functions trade off. PLoS computational biology *4*, e1000002.

Tokuriki, N., and Tawfik, D.S. (2009a). Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature *459*, 668-673.

Tokuriki, N., and Tawfik, D.S. (2009b). Protein dynamism and evolvability. Science (New York, NY *324*, 203-207.

Tomatis, P.E., Fabiane, S.M., Simona, F., Carloni, P., Sutton, B.J., and Vila, A.J. (2008). Adaptive protein evolution grants organismal fitness by improving catalysis and flexibility. Proceedings of the National Academy of Sciences of the United States of America *105*, 20605-20610.

Toprak, E., Veres, A., Michel, J.B., Chait, R., Hartl, D.L., and Kishony, R. (2012). Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. Nat Genet *44*, 101-105.

Tracewell, C.A., and Arnold, F.H. (2009). Directed enzyme evolution: climbing fitness peaks one amino acid at a time. Curr Opin Chem Biol *13*, 3-9.

Travisano, M. (1997). Long-term experimental evolution in Escherichia coli. VI. Environmental constraints on adaptation and divergence. Genetics *146*, 471-479.

Travisano, M., Mongold, J.A., Bennett, A.F., and Lenski, R.E. (1995). Experimental tests of the roles of adaptation, chance, and history in evolution. Science (New York, NY *267*, 87-90.

van Berkel, W.J., Kamerbeek, N.M., and Fraaije, M.W. (2006). Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. Journal of biotechnology *124*, 670-689.

Viveiros, M., Dupont, M., Rodrigues, L., Couto, I., Davin-Regli, A., Martins, M., Pages, J.M., and Amaral, L. (2007). Antibiotic Stress, Genetic Response and Altered Permeability of E. coli. PLoS ONE *2*, e365.

Viveiros, M., Jesus, A., Brito, M., Leandro, C., Martins, M., Ordway, D., Molnar, A.M., Molnar, J., and Amaral, L. (2005). Inducement and reversal of tetracycline resistance

in Escherichia coli K-12 and expression of proton gradient-dependent multidrug efflux pump genes. Antimicrobial agents and chemotherapy *49*, 3578-3582.

Vogel, F., and Kopun, M. (1977). Higher frequencies of transitions among point mutations. J Mol Evol *9*, 159-180.

Vogel, F., and Rohrborn, G. (1966). Amino-acid substitutions in haemoglobins and the mutation process. Nature *210*, 116-117.

Volkers, G., Palm, G.J., Weiss, M.S., Wright, G.D., and Hinrichs, W. (2011). Structural basis for a new tetracycline resistance mechanism relying on the TetX monooxygenase. FEBS Lett *585*, 1061-1066.

Walkiewicz, K., Davlieva, M., Wu, G., and Shamoo, Y. (2011). Crystal structure of Bacteroides thetaiotaomicron TetX2: a tetracycline degrading monooxygenase at 2.8 A resolution. Proteins *79*, 2335-2340.

Walsh, C. (2000). Molecular mechanisms that confer antibacterial drug resistance. Nature *406*, 775-781.

Wang, J., Ortiz-Maldonado, M., Entsch, B., Massey, V., Ballou, D., and Gatti, D.L. (2002). Protein and ligand dynamics in 4-hydroxybenzoate hydroxylase. Proceedings of the National Academy of Sciences of the United States of America *99*, 608-613.

Wangrong Yang, I.F.M., Kalinka P. Koteva, David C. Bareich, Donald W. Hughes, and Gerard D. Wright (2004). TetX Is a Flavin-dependent Monooxygenase Conferring Resistance to Tetracycline Antibiotics. The Journal of Biological Chemistry *279*, 52346-52352.

Watt, W.B., and Dean, A.M. (2000). Molecular-functional studies of adaptive genetic variation in prokaryotes and eukaryotes. Annu Rev Genet *34*, 593-622.

Weinreich, D.M., Delaney, N.F., Depristo, M.A., and Hartl, D.L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. Science (New York, NY *312*, 111-114.

Weinreich, D.M., Watson, R.A., and Chao, L. (2005). Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. Evolution; international journal of organic evolution *59*, 1165-1174.

Whittle, G., Hund, B.D., Shoemaker, N.B., and Salyers, A.A. (2001). Characterization of the 13-kilobase ermF region of the Bacteroides conjugative transposon CTnDOT. Applied and environmental microbiology *67*, 3488-3495.

Wielgoss, S., Barrick, J.E., Tenaillon, O., Cruveiller, S., Chane-Woon-Ming, B., Medigue, C., Lenski, R.E., and Schneider, D. (2011). Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution Experiment With Escherichia coli. G3 (Bethesda) *1*, 183-186.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., *et al.* Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr *67*, 235-242.

Zhang, D.F., Jiang, B., Xiang, Z.M., and Wang, S.Y. (2008). Functional characterisation of altered outer membrane proteins for tetracycline resistance in Escherichia coli. International journal of antimicrobial agents *32*, 315-319.