Substrate adaptability of β-lactamase

Sun, J.

Chapter 1

Introduction

Courage does not always roar. Sometimes courage is the quiet voice at the end of the day saying, 'I will try again tomorrow'.

Mary Anne Radmacher

Protein evolution

A fundamental paradox in protein evolution is "Natural selection may explain the survival of the fittest, but it cannot explain the arrival of the fittest."[1] Proteins serve as the primary engines of life, intricate macromolecules that execute the essential functions of living organisms. Protein evolution is central to understanding natural selection, which is the driving force behind adaptation of an organism and subsequent fitness change. Over many eras, proteins have undergone gradual, incremental changes, leading to their specialized functions. Their diversity, complexity, and functional adaptability have intrigued biologists for generations. The fitness of a protein, its capacity to carry out its function effectively, is directly influenced by its structural and functional features, including primary, secondary, and tertiary structure, the binding sites, conformational changes, and stability.[2,3] The fitness landscape is rugged and multidimensional and each peak or valley represents a different protein phenotype. Proteins traverse the landscape through genetic mutations with natural selection as the guiding force, influencing stability and functionality. Mutations, insertions/deletions, and recombination lead to subtle alterations in protein structure, which can be adequate for the emergence of novel functions.

The effect of multiple mutations can lead to unexpected results, relating to several concepts: epistasis, compensatory mutations, robustness, and trade-off. Epistasis is the phenomenon that the effect of two mutations is not equal to the sum of the two individual mutations. The effect can be larger (positive) or smaller (negative) or lead to new, unpredicted effects. It is noted that mutations that are neutral for the function can still be important in evolution because they can become beneficial due to other mutations later. In other words, neutral mutations change the epistatic potential.[4,5] A mutation is compensatory when the negative impact on fitness of one mutation is remedied by another mutation.[6] Evolutionary robustness reflects the degree to which proteins can endure the accumulation of mutations without altering their structure, stability, or function. Evolutionary adaptability indicates how readily a protein can acquire new functions, i.e. functional plasticity. The latter two are related, as

adaptability cannot exist without a degree of robustness, but robustness is no guarantee for adaptability. The acquisition of a new function is often related to trade-offs in the original phenotype, for example, original vs. new function, stability vs. activity, and rate vs. specificity for enzymes. Trade-offs between monomeric state and oligomer assembly or between folding capability and the stability of final folded state might underlie the formation of entirely new proteins.[7] Rate-specificity trade-offs involve mutations that improve the catalytic efficiency at the cost of selectivity or affinity.[8] If the phenotypic protein variability has no significant effect on the fitness during the evolution, the variability is categorized as "molecular noise", including phenomena such as expression level, promiscuous protein conformations, and translational errors.

Temperature and pH adaptation

Enzymes have an optimal temperature range, in which they function most effectively. Extreme temperatures, either too hot (denaturation) or too cold, can disrupt the structure and affect the dynamics. Enzymes adapted to extreme conditions, such as thermophiles or psychrophiles, have evolved to be stable at high or low temperatures, respectively. Enzymes can adapt rapidly to new environments and traverse a wide range of temperatures by directed evolution,[9] exhibiting a trade-off during evolution between their catalytic activity at low temperatures (which tends to be high for enzymes from psychrophilic organisms but generally low for enzymes from thermophiles) and their thermostability (which is high for thermophilic enzymes but low for enzymes from psychrophiles). Enzymes that are both highly thermostable and highly active at low temperatures do not typically occur naturally, but laboratory evolution experiments demonstrate that it is possible to generate such a combination of properties with a few amino acids substitutions.[3,9-13]

Changes in pH can alter the charge distribution within an enzyme, affecting its folding and function or affect the ionization state of amino acid residues in the active site, influencing activity. Some proteins are adapted to work optimally at specific pH levels, while others, like digestive enzymes, need to function over a wide pH range.

Evolution strategies

Protein evolution can be thought of as an adaptive walk on a fitness landscape in sequence space, where "fitness" is a quantitative measure of a physicochemical property of protein, such as thermal stability or enzymatic activity.[14,15] The adaptive walk involves evolutionary changes in the sequence caused by gene duplication, random mutation, transposition, gene recombination, or gene conversion.[16] *In vivo*, these processes usually operate slowly, eliciting changes of gene structures or functions over millions of years. To mimic the power of the natural evolution in the lab, the rate of random mutagenesis is enhanced. Numerous strategies exist, including oligonucleotide cassette mutagenesis, point mutagenesis through error-prone polymerase chain reaction (ER-PCR) and the use of mutator strains, as well as DNA shuffling. The two most relevant methods are discussed below.

Oligonucleotide-directed mutagenesis

Several techniques, such as site-directed and site-specific mutagenesis, can be utilized to randomize short segments of the genetic code of a protein. For one or a few codons, using oligonucleotides with partially randomized sequences can be used. Incorporation of the wild type sequence may be biased due to the greater stability of the template-oligonucleotide complex, which can be addressed by reducing the concentration of the wild type oligonucleotide.[17,18]

Error-prone polymerase chain reaction

EP-PCR is a simple and efficient method used to generate variants with random mutations. *Taq* polymerase is often employed, because of its naturally high error rate, but with an error bias for AT to GC base pair mutations. Recently, a newly developed polymerase was introduced to balance the bias (increase in GC to AT changes).[19,20] In error-prone PCR, the mutation frequency is increased by creating non-optimal conditions for the polymerase. The two most important factors to consider are mutation frequency and mutation pattern. A high concentration of $MgCl_2$ stabilizes non-complementary pairs and $MnCl_2$ is used to increase the error rate.[21] Mutation frequencies in the range of 0.11% to 2% (1 to 20 nucleotides per kilobase) can be achieved by adjusting the nucleotide ratios and the amount of $MnCl_2$,[22] while the average number of mutations per gene copy can be controlled by the number of PCR

cycles or the initial template concentration. The mutation frequencies also depend on the length and base composition of the template. The mutation pattern in EP-PCR is characterized by its randomness, such as deletions and insertions and point mutations, frameshift mutations, and multiple mutations, contributing to the diversity of the mutant library generated by EP-PCR.

Directed evolution

Systematic strategies for guiding protein evolution have been reported since the 1970s. Directed evolution has emerged as an invaluable and extensively employed technique in both molecular biology and protein engineering. It is a laboratory method to mimic the process of natural selection and it involves iterative rounds of genetic library screening. The key point in all directed evolution methods is to link the genetic code to the phenotype in some way and screen for improved phenotype. In this way, the library of genetic codes can be enriched for the best genes and in successive screening rounds beneficial mutations accumulate until the enzyme reaches the performance target. Biocatalysts have been engineered in this way, exhibiting heightened stability and increased activity, making them better suited for industrial processing settings and cell biocatalysis.[12,23,24] Directed evolution gives rise to new insights and deeper understanding of the relationship between protein sequence, structure and function.

β-Lactamase of Mycobacterium tuberculosis

Tuberculosis

The disease that we know as tuberculosis is one of the world's oldest and deadliest infectious diseases. Genetic studies indicate that tuberculosis has existed for at least 15,000 years.[25] The earliest evidence of tuberculosis in humans is from 2400-3400 BCE, based on Egyptian mummies having evidence of the disease in their spines.[26,27] Hippocrates coined the word "phithis," or "consumption," in 460 BCE because the disease caused significant weight loss.[28] Despite its frequency at the time, the cause of tuberculosis was unknown. As time went by, the disease reached epidemic proportions in Europe and North America during the 18th and 19th centuries. In the 19th century, tuberculosis was often associated with the romantic movement. Several poets, writers, and artists have suffered from this disease throughout the ages, including John Keats, Frédéric Chopin, and many others.[29] A fascination with tuberculosis influenced the aesthetic choices of the time among the artistic community. The use of pale complexions, languid poses, and a sense of fragility and vulnerability became fashionable among portraitists and artists.[30,31] Until 19th and 20th century, reduction of mortality was observed which was attributed to the development of the public healthcare and the deeper understanding of the disease. Nowadays, tuberculosis is still a global health concern, which was still the leading cause of death.[32]

Tuberculosis is caused by the bacterium *Mycobacterium tuberculosis* (Mtb). Théophile Laennec's research in the early 19th century marked the beginning of tuberculosis comprehension. Later, Jean-Antoine Villemin proved its communicability, and in 1882, Robert Koch identified the tubercle bacillus as the disease's causative agent.[33] Koch's endeavor to treat tuberculosis using tuberculin, an extract derived from the deceased tubercle bacillus, not only paved the way for rapid infection diagnosis methods but also earned him the Nobel Prize in 1905 for his contributions to the understanding of tuberculosis. Tuberculosis therapy started in 1944 with the discovery of medications including streptomycin[34] and para-

aminosalicylic acid.[35] In 1952, a triple therapy combination (streptomycin, para-aminosalicylic acid and isoniazid) was found, enabling to cure the disease.[36] In the 1970s, isoniazid and rifampin were recognized as drugs that could shorten the duration of treatment from 18 to 9 months,[37] and in the 1980s, pyrazinamide was found to shorten the duration of treatment by 6 months.[38] However, in resource-poor countries, these regimens are still too long and inconvenient to facilitate effective treatment. The emergence of drug-resistant strains also threatens Tuberculosis control in several regions around the world, including India, China, Russia, and the former Soviet Union.

β-Lactamases

β-Lactamases are the enzymes produced by certain bacteria that can confer resistance to β-lactam antibiotics, such as penicillin, cephalosporins, and carbapenems. The initial discovery of a β-lactamase dates to 1940 when it was identified in *Bacillus coli*.[39] These enzymes were first studied as model enzymes for chemists because of high production levels and kinetic properties, mainly from Gram-positive bacteria. The name "β-lactamase" was first used in the 1960s to distinguish enzymes hydrolyzing the β-lactam bond from the amidases that hydrolyzed the N-acylside chain on β-lactam rings.[40,41] Pre-1976, there were 2 distinct β-lactamases families based on Zn-dependent activity. Beyond substrate specificity, nomenclature was defined according to functional activity in the presence of EDTA or Zn^{2+}.[42] Later β-lactamases became appreciated as clinically important resistance factors, especially when transferred on plasmids. Classification of these resistance factors, or "R-factors", appeared before the classification of the enzymes.[43] There are five different types of resistance mechanism found in bacteria, including reduced permeability, modification of the target molecule, efflux pumps, altered metabolism, and enzymatic inactivation of antibiotics. Bacterial cell walls are composed of peptidoglycan that is made up of altering N-acetylmuramic acid (NAM) and N-acetylglucosamine (NAG) subunits. Enzymes named penicillin binding proteins (PBPs) are responsible for polymerization and modifications of the peptidoglycan chains. β-Lactam antibiotics bind the PBPs and act as inhibitors, leading to a

weakening of the cell wall. The cells eventually burst and die under osmotic pressure. β-Lactamases inactivate the antibiotics, one of the mechanisms enabling bacteria to resist these antibiotics, next to others, such as reduced uptake and enhanced efflux (Figure 1.1). Nowadays, there are over a thousand unique β-lactamase enzymes with a range of substrate profiles.[44] These enzymes are located in the periplasm of Gram-negative bacteria, whereas in Gram-positive bacteria they are bound to outside of the cytoplasmic membrane.[45]

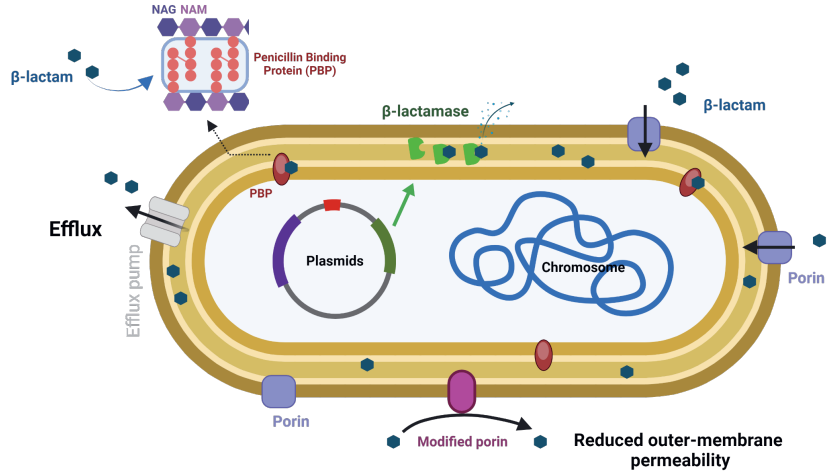

Figure 1.1 Mechanisms of β-lactam resistance in Gram-negative bacteria.

β-Lactamase classification

An early classification system was defined by the hydrolysis rates of penicillin and early cephalosporins.[46] Later, the β-lactamases were categorized in four molecular classes (Class A, B, C, and D), based on the molecular size and the homology of active-side amino acid motifs,[47-49] which is the most cited classification scheme recently. This system was first introduced in 1989 by Bush and updated in 2010.[50,51] A fundamental division is between serine β-lactamases (classes A, C, and D) and metallo-β-lactamases (class B), a heterogeneous group of zinc enzymes. The three classes of serine β-lactamases differ in sequence but share the acylation and deacylation mechanism. These enzymes feature a conserved Ser-Xaa-Xaa-Lys motif

and utilize the serine nucleophile in the acylation reaction.[52] In contrast, metallo-β-lactamases can be identified by the presence of a His-Xaa-Xaa-Asp motif which is important for metal coordination. Class A and D β-lactamases, such as TEM, SHV, CTX-M, and OXA are widely distributed and have evolved to extended-spectrum beta-lactamases (ESBLs), which are a major threat to the effectiveness of cephalosporin antibiotics.[53,54] The OXA family of Class D β-lactamases are considered to be carbapenemases and can evolve to break down carbapenems without requiring a deacylation water molecules.[55] Class C β-lactamases are generally encoded on the chromosome of Gram-negative bacteria and labeled as AmpC β-lactamases, such as CMY, FOX, and DHA. They mainly hydrolyze cephalosporins but also have resistance to benzylpenicillin.[50] The class B metallo-β-lactamases (MBLs) function differently from other β-lactamase classes, relying on zinc ions to facilitate the hydrolysis of β-lactam antibiotics.[56] They have a wide substrate range, including the ability to hydrolyze carbapenems[50,51,57] and limited activity against monobactams.[58,59]

BlaC

BlaC is a class A β-lactamases that is encoded by the *blaC* gene on the chromosome of *M. tuberculosis*.[60] It displays roughly 40% amino acid sequence homology with other class A β-lactamases.[61] It is an extended spectrum β-lactamase (ESBL), which possesses the ability to hydrolyze a wide range of substrates. It exhibits strong penicillinase activity[53,54] and can also hydrolyze second and third-generation cephalosporins.[61,62] In addition, BlaC is able to hydrolyze several carbapenems, such as imipenem and meropenem.[62] the presence of Gly132 rather than the conserved Asn132 in BlaC has been suggested to be the cause for carbapenemase activity.[62,63]
Like in all serine β-lactamases, the mechanism of inactivation of antibiotics comprises two steps, acylation and deacylation. During acylation, the catalytic Ser70 attacks the carbonyl carbon in the β-lactam ring, breaking the carbon-nitrogen bond and forming a covalent bond. Lys73 serves as general base for the acylation reaction by picking up the hydroxy proton of Ser70 (Figure 1.2a).[64,65] Ser130 acts as a proton relay station, picking up a proton from Lys73 and donating one to the nitrogen atom in the substrate.[66] During deacylation, a water molecule is activated by Glu166, which

attacks the same carbon and hydrolyses the ester bond with the oxygen of Ser70.[67-69] In cephalosporins, the denoted R_2 group is split off during the enzymatic reaction (Figure 1.2a). The loss of the leaving group has been suggested to occur after the opening of the β-lactam ring and may hamper fast catalysis of several serine β-lactamases.[70,71]

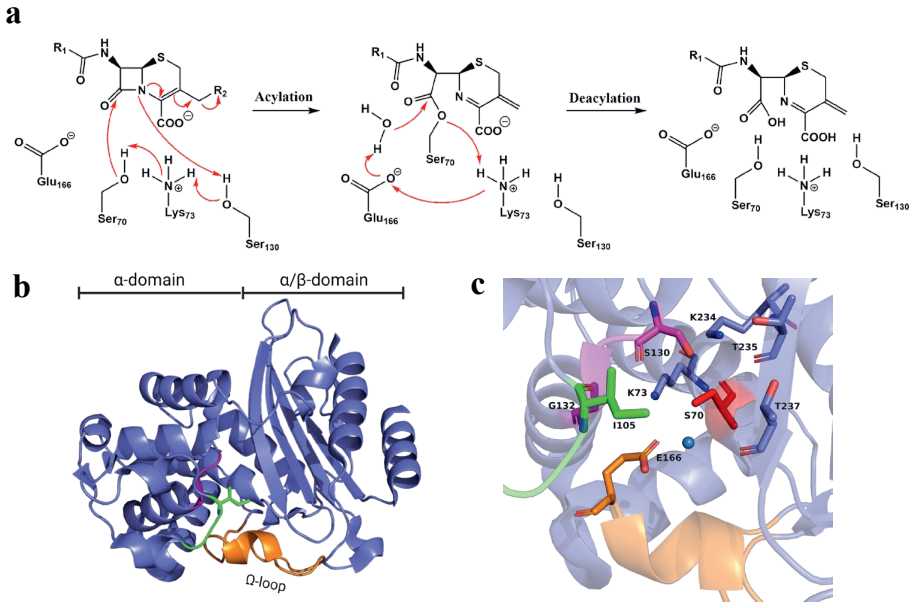

Figure 1.2 a) General mechanism of β-lactam hydrolysis; **b)** BlaC structure (PDB ID: 2GDN[72]) showing several important regions, including the SDG loop (magenta), gatekeeper region (green), and omega loop (orange); **c)** Detail of the active sites of BlaC (2GDN[72]) with important residues shown in sticks and a water as blue sphere.

The structure of the BlaC comprises two structural domains, the α-domain and α/β-domain (Figure 1.2b). The α-domain contains six long and three short α-helices, while the α/β-domain includes five antiparallel β-strands. The overall structure closely resembles that of other class A β-lactamases. Several highly conserved motifs can be distinguished. The conserved residues SX_2K (residue 70 to 73 according to ambler numbering[73]) are part of the active site and play a crucial role in the acylation (see above).[74,75] On one side of the active site, the conserved SDN motif (residues 130 – 132)

is important for substrate binding. Of note, Asn132 is substituted by Gly in BlaC.[74-76] The carboxylate binding site KTGT (234 - 237) contributes to interactions with the substrates, which all have such a negative group.[75,76] This site is also capable of binding phosphate ions.[77] A unique structural feature of BlaC, outside the active site is a glycine-rich insertion within the loop between H7 and H8. The role of this insertion, if any, is not known. Moreover, the "gatekeeper" loop (residue 103 - 106) and the Ω-loop (residue 161 - 179) are highly conserved (Figure 1.2c). The "gatekeeper" Ile105 covers the active site of BlaC, and affects substrate binding and protein stability.[78] The Ω-loop is essential for the substrate recognition and participates in stabilizing the acyl-enzyme intermediate during the hydrolysis.

β-Lactam antibiotics

The rise of antibiotic resistance in *M. tuberculosis* makes finding new treatments against TB urgent. Large pharmaceutical companies are not very interested in this area due to the poor business case. Existing β-lactam antibiotics, a group that includes multiple subgroups and compounds, offer some advantages such as low cost and user-friendliness. However, they are not commonly used for treating TB because the bacteria produce BlaC, making them resistant to these antibiotics. The first β-lactam antibiotic penicillin was discovered in 1929 by Alexander Fleming.[79] Following the successful discovery of penicillin, cephalosporins and carbapenems were developed from 1948 and 1976, respectively (Figure 1.3a).[80,81] Penicillins display high efficacy and low toxicity. The five-membered thiazolidine ring connected to the β-lactam ring forms the core of the structure. The broad-spectrum penicillins, such as amoxicillin, ampicillin, are efficient against both Gram-positive and Gram-negative bacteria.[82] Penicillins are distinguished by the substituent at C6 position (Figure 1.3b). Cephalosporins are also highly effective and commonly used for wide range of infectious diseases. These compounds are characterized by a bicyclic core combining a six-membered dihydrothiazine ring and a β-lactam ring. Variable sidechains have been introduced at carbon positions C3 and C7 (Figure 1.3b), extending antibacterial activities and improving the compound stability. Carbapenems are β-lactam antibiotics with the broadest spectra of antibiotic activity.[83] The five-membered pyrroline ring connected to a β-lactam ring is the core structure. The stable scaffold and saturated atoms enhance its resistance against most β-lactamases.[83]

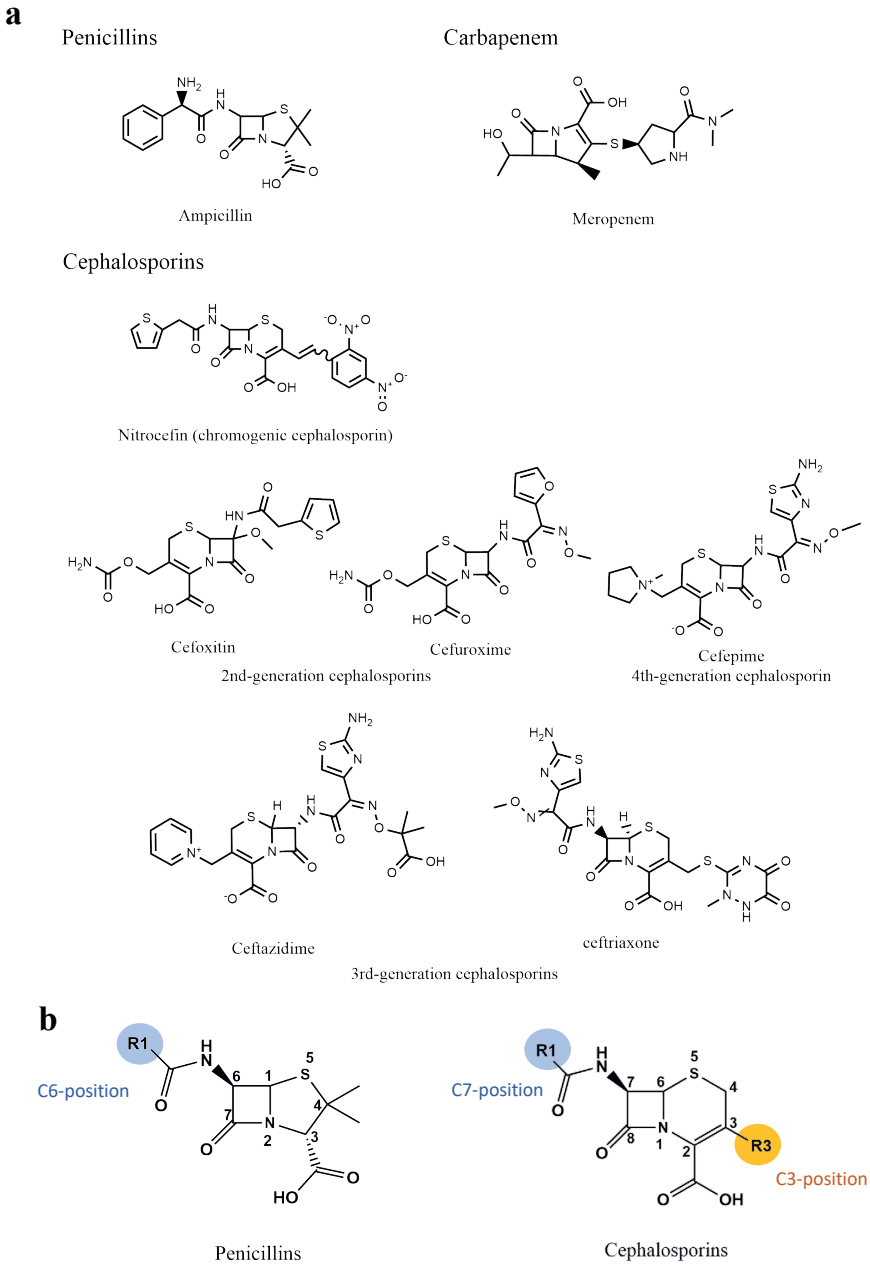

Figure 1.3 a) β-Lactamase substrates discussed in the thesis. **b)** The main difference between penicillin and cephalosporins. Penicillin contains a single side chain at the 6-position, while cephalosporins have two side chains at C3 and C7 position.

Classification of cephalosporin antibiotics

Four generations of cephalosporins have been developed. The first-generation cephalosporins, such as cefalexin, are active against Gram-positive bacteria, but have relatively weak activity against the Gram-negative bacteria. For this reason, the second-generation cephalosporins were developed, including cefoxitin and cefuroxime. This generation exhibited reduced hydrolysis by β-lactamases, especially from Gram-negative bacteria. The reduced dosage requirements and extended half-life of second-generation cephalosporins bring added benefits for clinical practice.[84] Most of first and second generation cephalosporins have acylamido groups similar to those in the earlier generations of penicillins. The introduction of the third-generation, e.g., ceftazidime and ceftriaxone, expanded the spectrum of antibiotic activities, particularly against Gram-negative bacteria such as *Enterobacter* or *Haemophilus influenzae*.[85] These cephalosporins are also frequently employed in the treatment of the sepsis from an unidentified origin. The fourth generation cephalosporins, such as cefepime, retain the activity against Gram-negative bacteria of the third generation and have improved Gram-positive activity. Third and fourth-generation cephalosporins feature bulky acylamido side chains at the C7 position, containing an oxyimino group, which sterically interferes with β-lactamase catalysis.[86]

Dynamics of β-lactamase

It is well known that proteins are dynamic molecules,[87] and that their dynamical properties are linked to function.[88] Proteins exist as a dynamic ensemble of conformations and several local energy minima may exist in the sampled conformational landscape, each with its own population. The exchange between different conformations can occur on a wide range of timescales, from femtoseconds to hours. The very fast timescales relate to vibrations and rotations of single bonds, whereas sidechain rotations and loop motions can occur on the scale from the picosecond to microsecond or even longer. Conformational dynamics can result in multiple structures and functions based on a single amino acid sequence and can be important for enzyme catalysis.[89-91] In general, crystallography is used to derive a

structural snapshot of the lowest energy state of an enzyme. Dynamic behavior of enzymes can be observed by nuclear magnetic resonance (NMR) spectroscopy. In the presence of dynamics, individual atom spins experience various chemical environments, and the rates of exchange between states can be measured by correlating them with the time-dependent phenomena of NMR, such as spin-spin interactions and relaxation. In the case of slow exchange on the chemical shift time scale, states with populations surpassing the detection limit can be observed separately.

Unlike enzymes for which conformational changes are known to be important for function, β-lactamase is thought to be in a single, well-defined and rigid state. Rigidification of active site or stabilization of the ground-state conformation is often used as a means to improve catalytic activity.[92] Explaining of how mutations alter substrate specificity revolves around the potential interaction between the substrate and the substituted residues. Simulation techniques have been used to capture lowly populated, hidden conformations to connect such high-energy structures to function and to predict the effects of mutations.[93-95] Although these conformations are lowly populated in the wild-type enzyme, mutations can shift the balance of populations, so that such alternative conformation becomes dominant, resulting in the shift towards favoring new activities and even allowing completely new enzyme functions to emerge. As a result, enzymes can greatly expand the functional diversity on the basis of a relatively small repertoire of sequences.[96] A better understanding of the link between conformational diversity and evolution is useful, not only for grasping the fundamental mechanisms of biochemistry, but also for designing new enzymes with tailored properties.

Research objectives and thesis outline

The research aims to explore the evolutionary adaptability of enzymes and the impact of temperature on protein evolution pathways, using *M. tuberculosis* β-lactamase BlaC as the object of study. Enzymes inherently embody a delicate balance between activity and stability, and the acquisition of new enzymatic functions is often accompanied by trade-offs, such as decreased stability or reduction of the original activity. Probing evolutionary adaptability of BlaC with laboratory evolution in combination with structural characterization can provide information about the mechanisms of rapid adaptations observed for β-lactamases in the clinic. The role of temperature as a conventional selection pressure in such evolutionary adaptation is unclear. The cooperative nature of enzyme unfolding over a narrow temperature trajectory raises the question whether evolution at temperatures well below the melting point is influenced by temperature. The approach used in this work to answer these questions is by simulating evolution under different selection pressures and characterize the variant enzymes in terms of activity, structure, dynamics and melting temperature. The research makes clear how enzyme kinetics and dynamics vary with different selection pressures and maps the evolutionary path that enzymes may take. The underlying structural mechanisms are established to provide a rationale for the observed effects.

Chapter 1 of this thesis provides the relevant background information about the field of protein evolution, β-lactamase BlaC, β-lactam antibiotics, and the research questions probed in this work. In ***Chapter 2*** an experimental study involving lab evolution and selective screening at various temperatures is described, yielding a mutant BlaC P167S with increased resistance to the antibiotic ceftazidime. The resting state of mutant P167S displays two forms that exchange on the minute's timescale. It is characterized in cells and *in vitro*. Crystal structures give insight into the structures of the two forms and NMR spectroscopy and kinetic experiments and modelling help to explain the mechanism of ceftazidime hydrolysis. In ***Chapter 3*** the question is addressed why the highly conserved Glu166 can be mutated to Ala while

keeping the ability of BlaC to hydrolyze ceftazidime. It is found that this residue is not required for the deacylation reaction. Interestingly, wild type BlaC is found to exist in two states at high pH resembling the states of mutant P167S. The pH-dependent switch between two states is caused by the loss of a low-barrier hydrogen bond in Ω-loop, which can as a potential starting point for further optimization through evolution under ceftazidime selection pressure. ***Chapter 4*** returns to the main research question on the effect of temperature on evolution. Mutant P167S/D240G is used as a template to test how higher ceftazidime resistance evolves at different temperatures. New mutants are found that show remarkable increase in ceftazidime resistance after several rounds of screening. At 23 °C and 37 °C, different mutations are found, suggesting that temperature has influence on the evolutionary landscape to some degree. ***Chapter 5*** provides a general discussion of the work in this thesis and outlook on further research.