Doctoral Thesis

Regulation, effectors, and evolvability of chorismate mutases from Corynebacteriales

Author(s): Kamarauskaite, Jurate

Publication Date: 2016

Permanent Link: https://doi.org/10.3929/ethz-a-010795159

Rights / License: In Copyright - Non-Commercial Use Permitted
REGULATION, EFFECTORS, AND EVOLVABILITY OF CHORISMATE MUTASES FROM CORYNEBACTERIALES

A thesis submitted to attain the degree of

DOCTOR OF SCIENCE of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

JURATE KAMARAUSKAITE
MSc Biochemistry, University of Copenhagen
born on 11.08.1984
citizen of Lithuania

accepted on the recommendation of
Prof. Dr. Peter Kast, examiner
Prof. Dr. Stefanie D. Krämer, co-examiner

2016
CHAPTER 2


CHAPTER 5

CONTENTS

1 INTRODUCTION ...........................................................................................................................................17
  1.1 THE SHIKIMATE PATHWAY ......................................................................................................................17
    1.1.1 Regulation of DAHP synthases ............................................................................................................20
    1.1.2 Regulation of chorismate mutases ........................................................................................................23
  1.2 SHIKIMATE PATHWAY ENZYMES AS TARGETS AGAINST TUBERCULOSIS .................................................28
    1.2.1 The need of novel drugs against tuberculosis .......................................................................................28
    1.2.2 Chorismate mutases as anti-TB drug targets .......................................................................................30
    1.2.3 Small molecules or peptides? ..............................................................................................................32
  1.3 CHORISMAETE MUTASES AS MODELS TO INVESTIGATE ENZYME CATALYSIS ..............................34
    1.3.1 Directed evolution ...............................................................................................................................37
  1.4 AIMS OF THIS THESIS ............................................................................................................................41

2 INTER-ENZYME ALLOSTERIC REGULATION OF AROQ$_8$ CHORISMAETE MUTASES ..............................................43
  2.1 INTRODUCTION ...........................................................................................................................................44
  2.2 RESULTS ......................................................................................................................................................45
    2.2.1 CgCM is activated by complex formation with CgDS .........................................................................45
    2.2.2 Impact of the feedback inhibitors Phe, Tyr, and Trp on CM-DS complex activity ...............................47
    2.2.3 Assembly of the MtCM-MtDS complex in solution ............................................................................50
  2.3 DISCUSSION ...............................................................................................................................................52
  2.4 MATERIALS AND METHODS ....................................................................................................................58
    2.4.1 Materials and general procedures ........................................................................................................58
    2.4.2 Protein production and purification ......................................................................................................58
    2.4.3 Enzymatic assays ..................................................................................................................................59
    2.4.4 Size-exclusion chromatography .........................................................................................................60

3 IN SEARCH FOR SMALL MOLECULE INHIBITORS OF THE CHORISMAETE MUTASES FROM M. TUBERCULOSIS .................................................................61
  3.1 INTRODUCTION ...........................................................................................................................................61
3.2 RESULTS .................................................................................................................. 65

3.2.1 Optimization of the coupled MtCM-MtDS assay for the HTS format .................................................................................................................. 65

3.2.2 The HTS campaign and hit validation for the MtCM-MtDS complex using the coupled assay .......................................................................................... 66

3.2.3 Validation of the hit compounds for MtCM-MtDS and *MtCM using the direct assay ........................................................................................................ 68

3.3 DISCUSSION .............................................................................................................. 70

3.4 MATERIALS AND METHODS .................................................................................. 72

3.4.1 Materials and general procedures ........................................................................ 72

3.4.2 Protein production and purification ...................................................................... 72

3.4.3 HTS for MtCM-MtDS .......................................................................................... 72

3.4.4 Direct UV absorbance assay ................................................................................ 73

4 IN VIVO SELECTION FOR CYCLIC PEPTIDE MODULATORS OF THE
MtCM-MtDS COMPLEX ................................................................................................. 74

4.1 INTRODUCTION ...................................................................................................... 74

4.2 RESULTS .................................................................................................................. 77

4.2.1 Genetic selection system for CP modulators of the MtCM-MtDS complex ................................................................. 77

4.2.2 Construction of the cyclic peptide libraries .......................................................... 82

4.2.3 Selection for CP activators of the MtCM-MtDS complex ................................. 85

4.2.4 Genetic selection system for CP inhibitors of the MtCM-MtDS complex ............. 87

4.3 DISCUSSION ............................................................................................................ 92

4.4 MATERIALS AND METHODS .................................................................................. 97

4.4.1 Materials and general procedures ........................................................................ 97

4.4.2 Bacterial strains and plasmids ............................................................................ 97

4.4.3 Construction of the split intein gene .................................................................... 98

4.4.4 Construction of plasmids .................................................................................... 98

4.4.5 Identification of pKIMP-ACGQ-RBS6 ................................................................. 103
4.4.6 Construction of the cyclic peptide libraries........................................ 104
4.4.7 Selection for the cyclic peptide activators ........................................ 105
4.4.8 Penicillin enrichment for the cyclic peptide inhibitors......................... 106
4.4.9 Intein production and purification ................................................... 107
4.4.10 Cyclic peptide isolation and detection by LC-MS............................. 107
4.4.11 Enzymatic assay............................................................................... 108

5 DIRECTED EVOLUTION OF MtCM ..................................................... 109
  5.1 INTRODUCTION ............................................................................... 110
  5.2 RESULTS AND DISCUSSION ............................................................ 112
    5.2.1 Directed evolution via inter-subunit destabilization ......................... 112
    5.2.2 Directed evolution via terminal truncation ..................................... 120
    5.2.3 Biochemical characterization of the evolved MtCM variants ............ 131
    5.2.4 Crystal structures of the evolved MtCM variants ......................... 134
  5.3 CONCLUSIONS ................................................................................. 145
  5.4 MATERIALS AND METHODS ............................................................. 146
    5.4.1 Materials and general procedures ............................................... 146
    5.4.2 Strains, plasmids, and general procedures ..................................... 147
    5.4.3 Directed evolution of 3p5 via inter-subunit destabilization ............. 147
    5.4.4 Directed evolution of re4.7s11 via terminal truncation ................... 151
    5.4.5 Production and purification of His₆-tagged proteins ....................... 156
    5.4.6 Construction of the evolved variants containing the wild-type N-terminus........................................................................ 157
    5.4.7 Production and purification of untagged proteins ......................... 157
    5.4.8 Enzymatic assays........................................................................... 158
    5.4.9 Circular dichroism spectroscopy................................................. 158
    5.4.10 Crystallographic data processing and structure refinement .......... 158

6 CONCLUSIONS AND OUTLOOK............................................................ 160

7 APPENDICES ....................................................................................... 166
SUMMARY

The shikimate pathway couples the carbohydrate metabolism to the biosynthesis of aromatic amino acids. It begins with the aldol-like condensation of D-erythrose-4-phosphate and phosphoenolpyruvate to form 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP), catalyzed by DAHP synthase. Via six subsequent enzymatic steps DAHP is converted into chorismate, which is the branch point metabolite for the biosynthesis of L-phenylalanine (Phe), L-tyrosine (Tyr), L-tryptophan (Trp), and other vital aromatic compounds. Since the pathway only exists in bacteria, fungi, plants, and apicomplexan parasites but is absent in mammals, the enzymes from the shikimate pathway are promising targets for the development of novel antibiotics, fungicides, and herbicides.

The first committed step in the biosynthesis of Phe and Tyr is the conversion of chorismate to prephenate catalyzed by chorismate mutase (CM). The presented study focuses on the CM from *Mycobacterium tuberculosis* (MtCM) comprising a homodimer of intertwined three-helical subunits. Remarkably, the enzyme lacks certain otherwise strictly conserved active site residues and exhibits a 100-fold lower catalytic efficiency ($k_{cat}/K_m = 1.75 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$) than other natural CMs. However, MtCM can be activated via the transient interaction with DAHP synthase (MtDS) to reach a $k_{cat}/K_m$ of $2.4 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$.

Interestingly, formation of the MtCM-MtDS complex enables feedback regulation of MtCM since addition of MtDS’s allosteric inhibitors Phe and Tyr reduces the CM activity of the complex. In Chapter 2, we investigated the kinetic effects of the aromatic amino acids at physiologically relevant concentrations. We showed that Phe alone inhibits the CM activity of the MtCM-MtDS complex up to 20 %, whereas a combination of Phe and Tyr reduces the CM activity by 70 %. We employed size-exclusion chromatography to show that the reduction of CM activity in the presence of Phe and Tyr or all three aromatic amino acids is a result of complex disassembly.

A kinetic study performed on the closely related CM-DAHP synthase complex from the industrially important *Corynebacterium*
*glutamicum* showed similar effects of Phe and Tyr. We also discovered that Trp acts as a cross-pathway feedback CM activator of the CM-DAHP synthase complex. However, this occurs only at low protein concentrations relative to the dissociation constant of the complex, indicating that *K*<sub>d</sub> values and the intracellular concentrations of the complex components are important factors that govern the response to Trp within the cell.

Chapters 3 and 4 of this thesis present our efforts to identify potential inhibitors of the CMs from *M. tuberculosis*, the causative agent of tuberculosis. Such compounds are of particular interest due to a large and ever growing number of infected individuals and a dangerous spread of multi-drug resistant strains. *M. tuberculosis* offers two CMs as potential targets for drug development: the aforementioned MtCM-MtDS complex, and the secreted enzyme (*MtCM*). The former is responsible for the mycobacterial biosynthesis of Tyr and Phe, and due to its unique structure, which is only found in certain orders of Actinobacteria, guarantees that any developed drugs are safe to humans and most gut microbiota. The role of *MtCM*, however, is not yet known. The enzyme is assumed to be associated with *M. tuberculosis* pathogenicity; therefore, novel inhibitors could assist in elucidating its enigmatic function and be used as drugs that disrupt the host-pathogen interaction.

Previously, a screen for inhibitors of *MtCM* from a small molecule library of 32,032 compounds was performed using a coupled assay developed for high throughput screening (HTS). In Chapter 3, we describe the adaptation of the coupled assay and the screening of the same compound library for the MtCM-MtDS complex. All validated hit compounds from the HTS for *MtCM* and MtCM-MtDS were subsequently tested in-house using a direct CM assay; however, no strong inhibitors neither for *MtCM* nor for the MtCM-MtDS complex could be confirmed.

Identification of small molecule inhibitors for protein-protein interactions (PPIs), such as those in the MtCM-MtDS complex, is a challenging task. Due to the typically featureless interfaces of the interacting proteins, the focus is currently shifting to molecules that provide larger surfaces for interaction. Therefore, we resorted to a library of
constrained peptides to identify potential modulators of the MtCM-MtDS complex.

In Chapter 4, we used the MtCM-MtDS complex as a model to develop a novel *Escherichia coli*-based selection system for peptide modulators, potentially applicable to any essential metabolic enzyme. In this work, we generated head-to-tail cyclized peptides *in vivo*, relying on the previously described split-intein circular ligation of peptides and proteins system (SICLOPPS) and developed selection conditions that would allow identification of activators and inhibitors of CM activity. The former requires stringent conditions where CM-deficient *E. coli* cells can only survive if the activity of the target enzyme is enhanced. The latter is based on penicillin selection for Phe and Tyr auxotrophs that emerge due to inhibited CM activity. The two selection concepts were established and tested but revealed several critical issues resulting in the identification of various artifacts. Nevertheless, the new experimental approaches presented in this study lay the groundwork for the further development of ultra-high throughput selection systems for peptidic modulators of the MtCM-MtDS complex.

As presented earlier, MtCM undergoes a substantial activation upon complex formation with MtDS. The high catalytic activity does not require participation of MtDS residues and is solely dependent on a conformational rearrangement of active site residues, induced by the complex partner. This implies that MtCM holds an intrinsic capability of efficient catalysis, however, has evolved as a sluggish enzyme on its own to allow allosteric regulation. In Chapter 5, we thus employed directed evolution to investigate whether high catalytic activity of MtCM can be achieved via a mechanism other than interaction with the partner protein.

During two previously conducted evolution cycles, MtCM regions that are at the interface with MtDS were targeted by cassette mutagenesis. The best evolved variant exhibited a 92-fold increase in catalytic activity, but was still less active than the MtDS-activated wild-type enzyme. To enhance the catalytic efficiency of MtCM further, the available *E. coli*-based selection system had to be modified. We employed the strategies of inter-subunit destabilization and terminal truncation of the parent protein to cripple its enzymatic activity
thereby again increasing the stringency of the \textit{in vivo} selection. After two additional rounds of evolution we obtained the variant s10es4.15 ($k_{\text{cat}}/K_m = 4.7 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$) that even surpasses the wild-type MtCM in complex with MtDS. Structural studies revealed that the evolved variant adopts the catalytically beneficial conformation assumed only by the activated wild-type enzyme. Our results thus prove that, despite lacking active site residues conserved in related AroQ CMs of other subclasses, MtCM incorporates all the necessary features for efficient catalysis.

In conclusion, this thesis expands the general understanding of mechanisms governing the biosynthesis of the essential aromatic amino acids Phe and Tyr in \textit{Corynebacteriales}. We describe members of the unique AroQ$_\delta$ subclass of CMs that exhibit unusually mediocre catalytic activity on their own, however, possess the intrinsic potential to become highly active catalysts. It is clear that natural evolution came up with these CMs to allow for inter-enzyme allosteric regulation, accomplished via interaction with the first shikimate pathway enzyme DAHP synthase. We hope that understanding of this unique system will open new paths for the development of antibacterial drugs or for improving the industrial production of aromatic compounds.
ZUSAMMENFASSUNG

Der Shikimat-Weg verbindet den Kohlenhydrat-Metabolismus mit der Biosynthese von aromatischen Aminosäuren. Er beginnt mit der durch DAHP-Synthase katalysierten Kondensation von D-Erythrose-4-Phosphat und Phosphoenolpyruvat zu 3-Desoxy-D-Arabino-Heptulosonat-7-Phosphat (DAHP). Durch sechs aufeinanderfolgende enzymatische Schritte wird DAHP in Chorismat umgewandelt, welches der Verzweigungspunkt-Metabolit für die Biosynthese von L-Phenylalanin (Phe), L-Tyrosin (Tyr), L-Tryptophan (Trp) und anderer, lebenswichtiger aromatischer Verbindungen ist. Darüber hinaus sind die Enzyme des Shikimat-Wegs vielversprechende Ziele für die Entwicklung neuer Antibiotika, Fungizide und Herbizide, da sie nur in Bakterien, Pilzen, Pflanzen und apicomplexen Parasiten existieren, bei Säugetieren jedoch fehlen.

Der erste irreversible Schritt bei der Biosynthese von Phe und Tyr ist die Umwandlung von Chorismat zu Prephenat, welche durch das Enzym Chorismat-Mutase (CM) katalysiert wird. Die vorliegende Studie konzentriert sich auf die CM von Mycobacterium tuberculosis (MtCM), einem Homodimer zusammengesetzt aus zwei verschlungenen Untereinheiten, welche je aus drei α-Helices bestehen. Bemerkenswert ist, dass dem Enzym mehrere normalerweise strikt konservierte Reste in der aktiven Stelle fehlen und es eine 100-fach geringere katalytische Aktivität \( \frac{k_{\text{cat}}}{K_m} = 1.75 \times 10^3 \text{ M}^{-1} \text{ s}^{-1} \) als andere natürliche CMs aufweist. Jedoch kann es über die transiente Interaktion mit dem ersten Enzym des Shikimat-Wegs, DAHP-Synthase (MtDS), aktiviert werden, um einen \( \frac{k_{\text{cat}}}{K_m} \) von \( 2.4 \times 10^5 \text{ M}^{-1} \text{ s}^{-1} \) zu erreichen.

Interessanterweise erlaubt die Bildung des MtCM-MtDS Komplexes eine Rückkopplungsregulation der MtCM, da eine Zugabe der allosterischen Inhibitoren Phe und Tyr von MtDS die CM Aktivität des Komplexes reduziert. In Kapitel 2 dieser Arbeit untersuchten wir die kinetischen Effekte der aromatischen Aminosäuren bei physiologisch relevanten Konzentrationen. Wir konnten zeigen, dass Phe allein die CM-Aktivität des MtCM-MtDS Komplexes um bis zu 20 % verringert, während eine Kombination von Phe und Tyr die CM-Aktivität sogar um 70 % reduziert. Wir verwendeten ausserdem die
Grössenausschluss-Chromatographie um zu zeigen, dass die Reduktion der CM-Aktivität in Gegenwart von Phe und Tyr oder von allen drei aromatischen Aminosäuren das Ergebnis einer Komplex-Dissoziation ist.


Eine niedermolekulare Chemikalienbibliothek bestehend aus 32’032 Verbindungen wurde zuvor mittels eines gekoppelten und für Hochdurchsatz-Screening (HTS) entwickelten Assays auf Inhibitoren von *MtCM überprüft. In Kapitel 3 beschreiben wir die Anpassung des
gekoppelten Assays und das Screening der gleichen Substanzbibliothek für den MtCM-MtDS Komplex. Alle bestätigten Hits aus dem HTS für *MtCM und MtCM-MtDS wurden anschliessend intern unter Anwendung eines direkten CM Assays getestet. Jedoch wurden hiermit weder für *MtCM noch für den MtCM-MtDS Komplex starke Inhibitoren bestätigt.

Die Identifizierung von Kleinmolekül-Inhibitoren für Protein-Protein-Wechselwirkungen (PPIs) wie diejenigen im MtCM-MtDS Komplex ist eine sehr schwierige Aufgabe. Aufgrund der konturlosen Oberflächen der interagierenden Proteine verschiebt sich der Fokus momentan in Richtung Moleküle, welche ausgedehntere Flächen zur Interaktion anbieten. Aus diesem Grund griffen wir auf eine Bibliothek von konformationell eingeschränkten Peptiden zurück, um potentielle Modulatoren des MtCM-MtDS Komplexes zu identifizieren.


Wie bereits dargestellt, erfährt MtCM eine wesentliche Aktivierung bei der Komplexbildung mit MtDS. Darüber hinaus benötigt die hohe katalytische Aktivität keine Beteiligung von MtDS-
Resten und ist allein abhängig von Konformationsänderungen im aktiven Zentrum, welche durch den Komplexpartner induziert werden. Dies lässt vermuten, dass MtCM die intrinsische Fähigkeit zur effizienten Katalyse zwar besitzt, sich jedoch zu einem schwachen eigenständigen Enzym entwickelt hat, um allosterische Regulation zu ermöglichen. In Kapitel 5 wandten wir daher gerichtete Evolution an, um zu untersuchen, ob eine hohe katalytische Aktivität von MtCM über einen anderen Mechanismus als die Interaktion mit dem Partnerprotein erreicht werden kann.

Während zwei zuvor durchgeführten Evolutionszyklen wurden MtCM Regionen an der Kontaktfläche mit MtDS mit Hilfe der Kassetten-Mutagenese gezielt untersucht. Die beste daraus hervorgegangene Variante zeigte einen 80-fachen Anstieg der katalytischen Aktivität, war aber immer noch weniger aktiv als das MtDS-aktivierte Wildtyp-Enzym. Um die katalytische Aktivität von MtCM weiter zu verbessern, musste das vorhandene E. coli-basierte Selektionssystem modifiziert werden. Wir benutzten die Strategien der Destabilisierung von Untereinheiten-Wechselwirkungen und der terminalen Verkürzung des Elternproteins, um seine enzymatische Aktivität zunächst zu reduzieren um damit die Stringenz des in vivo Selektionssystems wieder zu verstärken. Nach zwei weiteren Evolutionszyklen konnten wir die Variante s10es4.15 ($k_{\text{cat}}/K_m = 4.7 \times 10^5 \text{M}^{-1} \text{s}^{-1}$) gewinnen, welche sogar das MtDS-aktivierte Wildtyp-Enzym MtCM übertrifft. Strukturuntersuchungen enthielten, dass die neu evolvierte Variante ähnliche katalytisch vorteilhafte Konformationen einnimmt, wie sie zuvor nur im MtDS-aktivierten Wildtyp beobachtet wurden, Unsere Resultate beweisen, dass MtCM trotz fehlender katalytischer Reste in der aktiven Stelle, welche sonst in AroQ CMs anderer Subklassen hochkonserviert sind, alle notwendigen Eigenschaften für eine effiziente Katalyse in sich trägt.

In ihrer Gesamtheit erweitert diese Studie das allgemeine Verständnis für die Mechanismen, welche die Biosynthese der essentiellen aromatischen Aminosäuren Phe und Tyr in Corynebacteriales regeln. Wir beschreiben die einzigartige AroQs Unterklassene der CMs, deren Mitglieder alleine eine ungewöhnlich mittelmässige katalytische Aktivität aufweisen, jedoch das intrinsische Potential besitzen, zu hochaktiven Katalysatoren zu werden. Die
Evolution brachte diese CMs offensichtlich hervor, um eine allosterische Inter-Enzym-Regulation zu ermöglichen, welche über die Wechselwirkung mit dem ersten Enzym des Shikimat-Wegs DAHP Synthase erreicht wird. Wir hoffen, dass das Verständnis dieses einzigartigen Systems neue Wege für die Entwicklung antibakterieller Medikamente oder die Verbesserung der industriellen Produktion von aromatischen Verbindungen eröffnet.
1 INTRODUCTION

1.1 The shikimate pathway

The shikimate pathway metabolizes a significant portion of organic carbon for the biosynthesis of aromatic compounds in bacteria, fungi, algae, plants, and protozoan parasites.\cite{1,2,3} The central products of the pathway are the aromatic amino acids L-phenylalanine (Phe, F), L-tyrosine (Tyr, Y), and L-tryptophan (Trp, W). The spectrum of secondary metabolites produced depends on the biochemical needs of the host organisms and includes lignin and phytoalexins in plants or iron chelators (siderophores) in bacteria.\cite{1,4} Mammals, however, lack the shikimate pathway and supplement their requirement for Phe and Trp via their diet, while Tyr is obtained by hydroxylation of Phe.\cite{5}

The shikimate pathway is an ancient pathway that is localized in the cytosol of fungal and the cytoplasm of prokaryotic cells. Higher plants produce most of the required enzymes with the transit sequence directing them into the plastids.\cite{3} It has been shown, that in eukaryotes the shikimate pathway originates from horizontal gene transfer, gene fusion, and endosymbiotic gene transfer from cyanobacterial progenitor or other eubacterial genomes.\cite{6}

The pathway received its name after the intermediate metabolite shikimate (Japanese shikimi-no-ki), which was isolated from Chinese star anise (Illicium anisatum) in 1885 by Eijkman.\cite{7} It starts with the aldol-like condensation of erythrose 4-phosphate (E4P) from the non-oxidative branch of the pentose phosphate pathway and phosphoenolpyruvate (PEP) from the glycolysis pathway to form 3-
deoxy-D-arabinose-7-phosphate (DAHP) catalyzed by DAHP synthase (DS, EC 2.5.1.54) (Figure 1.1). The action of six other enzymes leads to chorismate, which is a central precursor for the biosynthesis of primary and secondary metabolites. The conversion of chorismate into prephenate, catalyzed by chorismate mutase (CM; EC 5.4.99.5) is the first committed step for the biosynthesis of aromatic the amino acids Phe and Tyr. The subsequent prephenate conversion depends on the host organism and proceeds either via 4-hydroxyphenylpyruvate (Tyr biosynthesis) and phenylpyruvate (Phe biosynthesis) or via arogenate (both Tyr and Phe biosynthesis). Trp is synthesized via an independent route and starts with chorismate conversion into anthranilate catalyzed by anthranilate synthase (Figure 1.1).

Being essential and resource intensive, the shikimate pathway is tightly regulated at the transcriptional, translational, and enzyme activity level. The most famous example of regulation at the genetic level is the trp operon, which was the first operon to be discovered in Escherichia coli. However, while the slow regulation at the transcriptional level is more predominant in plants, bacteria usually exploit regulation at the protein level allowing an immediate response to the changing metabolite concentrations.

The main focus of this thesis deals with the mechanisms for regulation of DAHP syntheses and CMs, representing prominent examples for illustrating the high diversity of allosteric regulation exhibited by the components of the shikimate pathway.

**Figure 1.1 Biosynthesis of aromatic amino acids.** The shikimate pathway starts with the condensation of phosphoenolpyruvate (PEP) with D-erythrose-4-phosphate (E4P) and ends with chorismate. Chorismate is a precursor for the biosynthesis of Trp via anthranilate, and Tyr/Phe via prephenate. In bacteria Phe and Tyr are synthesized via prephenate, arogenate, or both routes. In fungi, such as *S. cerevisiae* and *N. crassa*, the pathway usually branches after prephenate. Cyanobacteria, coryneform bacteria, *Actinomycetes* and plants use preferably the arogenate route.
Chapter 1: Introduction

**Figure 1.1 (legend previous page)**
1.1.1 Regulation of DAHP synthases

DAHP synthases catalyze the first step of the shikimate pathway to produce DAHP and inorganic phosphate (Figure 1.2). The carbonyl group of E4P, which has been activated by the metal ion, is attacked by the electron-rich enol alkene of PEP. The resulting oxocarbenium ion is attacked by water to form hemiketal intermediate, which subsequently loses the phosphate group to produce the C2 carbonyl. Product release is followed by spontaneous cyclization to give rise to cyclic DAHP.[14]

Figure 1.2 Reaction catalyzed by DAHP synthase. The proposed mechanism of aldol-like condensation of phosphoenolpyruvate (PEP) with D-erythrose-4-phosphate (E4P); H-Enz, acidic residue from the enzyme.[14]

DAHP synthases are at the gateway of the shikimate pathway and, therefore, are typically feedback regulated in prokaryotes. In plant cells, however, the regulation of DAHP synthases occurs at the genetic level and allosteric regulation of these enzymes is still questionable.[3]

Bacterial DAHP synthases have been divided into two structural groups based on phylogenetic reconstruction: the smaller Type I (subdivided into Iα and Iβ) of approximately 30-40 kDa, and the larger Type II enzymes of >50 kDa. Despite very low sequence similarities, these subtypes share a common triosephosphate isomerase (TIM) barrel fold (α/β)$_8$ and the requirement for a divalent metal ion for catalytic activity. However, depending on the allosteric regulation mechanisms, their oligomeric organization and the presence of sequence extensions and insertions vary largely (Figure 1.3).[10]

Type Iβ DAHP synthases share high structural similarity to 3-deoxy-D-manno-octulosonate-8-phosphate synthase (EC 2.5.1.55), an enzyme from lipopolysaccharide biosynthesis in Gram-negative
bacteria, which catalyzes the condensation of PEP and arabinose-5-phosphate.\footnote{15} The basic prototype of Type I\(\beta\) enzymes is the unregulated DAHP synthase from \textit{Pyrococcus furiosus} that exists as a dimer in solution and does not have any effector binding site\footnote{16} (Figure 1.3A). The more sophisticated members of type I\(\beta\) DAHP synthases contain an N- terminal ACT (or ferredoxin-like) domain or N- (or C-) terminal CM-like domains. These extensions form the effector binding sites upon tetramer formation and allow feedback regulation by undergoing structural rearrangements. The Phe/Tyr or chorismate/prephenate responsive Type I\(\beta\) DAHP synthases can be exemplified by the enzymes from \textit{Thermotoga maritima}\footnote{17, 18} and \textit{Listeria monocytogenes}\footnote{19} (Figure 1.3C and B), respectively.

The Type I\(\alpha\) DAHP synthases have a \(\beta\)-hairpin insertion in addition to the structural elements of the unregulated Type I\(\beta\) enzymes. Terminal extensions induce tetramer formation and, together with the \(\beta\)-hairpin, form a cavity for effector binding as illustrated by DAHP synthase from \textit{E. coli}\footnote{20} (Figure 1.3D). In this case, the effector binding does not induce major domain movements as in I\(\beta\) DAHP synthases but rather causes subtle conformational changes in the catalytically important 2\(\beta\)/2\(\alpha\) loop. Organisms often have multiple I\(\alpha\) isoenzymes, each specific for one of the three aromatic amino acids; e.g \textit{E. coli} has all three\footnote{21} and \textit{Saccharomyces cerevisiae} has two (lacks Trp specific isoform) isoenzymes.\footnote{22}

Finally, the largest and most sophisticated Type II DAHP synthases are found mostly in plants and are often called plant-type DAHP synthases. They contain an N- terminal extension and a helical insertion as compared to the prototypic DAHP synthase of Type I\(\beta\) and form a tetrameric structure, which is unrelated to the I\(\alpha\) DAHP synthases (Figure 1.3E).

Type II enzymes can be unregulated like DAHP synthase from \textit{Helicobacter pylori}\footnote{23} or regulated. The latter enzymes can be responsive to Trp (\textit{Neurospora crassa} DAHP synthase\footnote{24}), a combination of Trp and chorismate (\textit{Xanthomonas campestris} DAHP synthase\footnote{25}) or synergistically regulated by all three aromatic amino acids. For instance, the DAHP synthase from \textit{Mycobacterium tuberculosis} (MtDS), is only weakly inhibited by individual aromatic
Type Iβ

Unregulated

B.

Chorismate/prephenate

C.

Phenylalanine/Tyrosine

Type Iα

Phenylalanine

D.

Type II

Phenylalanine/Tyrosine/Tryptophan

E.

Figure 1.3 (legend next page)
amino acids,[26] but combinations of Phe, Tyr, and Trp strongly inhibit the enzyme.[27] Interestingly, Webby and co-authors did not observe a significant inhibition by a combination of Phe and Tyr in the absence of Trp,[27] in contrast to observations by our group.[26, 28] However, both studies report that the strongest inhibition of MtDS occurs in the presence of all three aromatic amino acids.[26, 29, 28]

1.1.2 Regulation of chorismate mutases

CMs catalyze the intramolecular 3,3-sigmatropic (Claisen) rearrangement of chorismate to prephenate (Figure 1.4). The conversion is highly exergonic and can be regarded as essentially irreversible in the cell.[30] Therefore, tight feedback regulation of CM activity is necessary to prevent continuous drain of cellular resources once the needs for the pathway end products have been met. As for the DAHP synthases, CMs exhibit a variety of structural organizations that fall into two general structural classes (Figure 1.5). The AroH class enzymes are characterized by a trimeric pseudo α/β-barrel fold, whereas AroQ class CMs typically exhibit a dimeric α-helical structure. The more abundant AroQ enzymes have been further divided into
subclasses AroQ\textsubscript{a} through AroQ\textsubscript{6} to describe their larger structural diversity.[\textsuperscript{31, 32}]

The CMs that are not subject to feedback regulation are found among both, the AroH and AroQ class enzymes. They are typically monofunctional catalysts, as illustrated by the intracellular CMs from *Bacillus subtilis* (BsCM, from the AroH class), *Methanococcus jannaschii* (MjCM, from the AroQ\textsubscript{a} subclass), and the secreted CM from *Mycobacterium tuberculosis* (*MtCM*, from the AroQ\textsubscript{f} subclass).

**Figure 1.4 Chorismate mutase reaction.** CM catalyzes the Claisen rearrangement of chorismate to prephenate. The reaction proceeds via an endo-oxa-bicyclic transition state with a chair-like geometry. Below shown is Bartlett’s endo-oxa-bicyclic transition state analog (TSA), \textsuperscript{[33]} which is a good inhibitor of CMs.

BsCM exhibits the typical AroH homotrimERIC structure with three shared active sites located at the interfaces of its subunits\textsuperscript{[34]} (Figure 1.5A). MjCM is an \(\alpha\)-helical dimer, which forms two shared active sites provided by its intertwined subunits, similar to the known structure of the closely related CM domain of the CM-prephenate dehydratase from *E. coli*\textsuperscript{[35]} (Figure 1.5D). The secreted *MtCM* exhibits a more complex \(\alpha\)-helical structure with only one active site per subunit\textsuperscript{[31]} (Figure 1.5B). Despite being structurally more sophisticated than AroQ\textsubscript{a} subclass enzymes, *MtCM* exhibits no response to the aromatic amino acids\textsuperscript{[36]} This could be related to its still enigmatic function as an
extracellular enzyme, which cannot be responsible for meeting the ever-changing metabolic demands of the producing organism.

The allosterically regulated CMs usually exhibit the structural architectures of the AroQ class and occur as mono- or multifunctional enzymes. Typical examples for feedback-sensitive monofunctional CMs are found in *S. cerevisiae* (ScCM) and *Arabidopsis thaliana* (AtCM). These enzymes belong to the AroQβ subclass and exist as α-helical dimers, similar to *MtCM*. ScCM binds Tyr or Trp resulting in a less active (T-state) or a more active (R-state) conformation, respectively[37] (Figure 1.5C). *A. thaliana* has three CM isoenzymes: the cytosolic AtCM2 is non-allosteric, while the plastidic AtCM1 and AtCM3 respond to aromatic amino acids. Phe and Tyr inhibit the AtCM1 isoenzyme, whereas Trp activates both AtCM1 and AtCM3. Surprisingly, Cys and His could also activate AtCM3, but to a minor extent.[38]

Representatives for the feedback-sensitive bifunctional enzymes are the CM-prephenate dehydratase (CM-PDT, encoded by pheA) and CM-prephenate dehydrogenase (CM-PDH, encoded by tyrA) from *E. coli*. The CM-PDT is a three-domain protein involved in Phe biosynthesis (Figure 1.1). It contains an N-terminal CM domain (EcCM, Figure 1.5D), a central PDT domain, and a C-terminal ACT domain responsible for feedback control by Phe. The binding of Phe causes a shift in the oligomeric state from the active dimer to the less active tetramer or octamer.[39] The two-domain CM-PDH is involved in Tyr biosynthesis (Figure 1.1) and is feedback regulated by Tyr. The mechanism for the activity regulation has been disputed since it was suggested that Tyr either acts as a competitive inhibitor in the dehydrogenase reaction,[40] or binds to an allosteric site.[41] However, the former hypothesis might be more plausible since Tyr has been found in the PDH active site of the closely related TyrA protein from *Haemophilus influenzae* with which TyrA from *E. coli* shares 57% sequence identity.[42] Moreover, similarly to EcCM, which improves its substrate affinity in the context of the other PheA domains, the activity of the CM domain from CM-PDH is 10-fold lower when the PDH domain is removed.[43] This suggests that interactions between the CM and PDH domains play a role in CM regulation.
Regulation, effectors, and evolvability of chorismate mutases from *Corynebacteriales*

![Diagram of chorismate mutases](image)

**Figure 1.5 (legend next page)**
Figure 1.5 Structural and allosteric diversity of chorismate mutases. (A) The typical structure of unregulated AroH enzymes is represented by the CM from B. subtilis (BsCM) (PDB ID 2CHT). It forms three active sites at the interfaces of the three subunits (colored in blue, pink and yellow) that form a pseudo-α/β-barrel. (B) The non-allosteric secreted CMs are α-helical dimers with one active site per protomer. The prototypic example for the unregulated CMs is *MtCM from the AroQγ subclass (PDB ID 2FP2). (C) The allosterically regulated monofunctional enzyme from S. cerevisiae represents the AroQδ subclass. It forms a homodimer with one active site per single subunit and undergoes large conformational changes upon Trp (black spheres) binding to adopt the activated form (PDB ID 1CSM). (D) The CM domain of the E. coli CM-prephenate dehydratase has been assigned to subclass AroQα. It comprises a homodimeric α-helical structure with two active sites composed by the two subunits. (E) The CM from M. tuberculosis belongs to the AroQδ subclass and occurs as a helical homodimer with two shared active sites (MtCM alone from PDB ID 2W1A). It forms a non-covalent complex with MtDS (PDB ID 2W1A), and is allosterically regulated by Phe, Tyr, and Trp, which bind to MtDS (See Figure 1.3E). In (A), (B), (C), (D), and (E) the different colors represent one subunit of the tertiary structure. The TSA (see Figure 1.4) in (A), (B), (D), and (E) is shown in black sticks; the structures are not shown to scale.

Interestingly, allosteric trifunctional enzymes have been identified in Nanoarchaeum equitans,[44] and Archeoglobus fulgidus,[45] which exhibit CM, PDT, and PDH activities. PDT activity was strongly inhibited by Phe in both organisms, while Tyr did not affect PDH in N. equitans, but weakly inhibited PDH and activated PDT in A. fulgidus. However, not much is known yet about allosteric regulation of CM activity in these proteins.

A unique case of inter-enzyme allosteric control was identified for monofunctional CMs from certain Actinobacteria.[46, 47, 48, 28] The prototype of these CMs is the intracellular CM from M. tuberculosis (MtCM) belonging to the AroQδ subclass enzymes (Figure 1.5E). It exhibits one of the simplest AroQ-type architectures that lacks a C-terminal helical extension, characteristic for structurally similar AroQα enzymes. The extended C-terminus of AroQα enzymes encodes a conserved active site residue (Gln88 in EcCM and its homologue in MjCM), which is absent in the truncated AroQδ enzymes. Initially, this was assumed to be the reason why catalytic efficiency of MtCM (k_{cat}/K_m =1.75×10^3 \text{ M}^{-1} \text{s}^{-1})[48] is two orders of magnitude lower than typical efficiencies demonstrated by natural CMs (k_{cat}/K_m in the range of 10^5 to 10^6 \text{ M}^{-1} \text{s}^{-1}).[37, 49, 36] However, it was shown later, that MtCM requires
MtDS as an interaction partner to enhance its catalytic activity to a \( k_{\text{cat}}/K_m \) of \( 2.4 \times 10^5 \) M\(^{-1}\) s\(^{-1}\).[48] Moreover, this transient protein-protein interaction (PPI) renders MtCM responsive to the pathway’s end products, the allosteric inhibitors of MtDS - Phe, Tyr, and Trp (see section 1.1.1 and Figure 1.3E). The mechanism by which aromatic amino acids affect the CM activity of the MtCM-MtDS complex and the closely related CM-DAHP synthase complex from Corynebacterium glutamicum will be the subject of Chapter 2 of this thesis.

Taken together, the variety of structural organizations exhibited by CMs implies that they have emerged at least twice during the course of evolution and converged to adopt similarly functionalized active sites.[50, 51] The modes of allosteric regulation, however, vary greatly throughout the structural classes and reflect the specific physiological needs of the host organisms.

1.2 Shikimate pathway enzymes as targets against tuberculosis

1.2.1 The need of novel drugs against tuberculosis

Tuberculosis (TB) is an ancient disease that emerged some 70,000 years ago, accompanying the migration of modern humans out of Africa.[52] Despite prominent successes to eradicate many infectious diseases, TB alongside with AIDS is still a leading cause of death worldwide. According to the World Health Organization (WHO), 0.4 million HIV-positive and 1.1 million HIV-negative TB cases were lethal in 2014 while up to one third of the world’s population is latently infected and can develop the active form of the disease at any time.

TB is a chronic, granulomatous infection that typically affects lungs resulting in pulmonary TB. However, it can also cause the extrapulmonary TB by affecting the skin, joints, bones, lymph nodes, meninges, pleura, abdomen, and genitourinary tract.

Adaptive immunity rarely eradicates the \( M. \) tuberculosis pathogen, allowing a long-term latent TB infection.[53] The innate immune cells, such as alveolar macrophages and granulocytes, immediately engulf the pathogen upon infection and travel to regional lymph nodes to initiate the immune response. The specific T cells then return to the
site of infection and participate in formation of granulomas (tubercules).\cite{54} Granulomas are organized as compact aggregates of infected macrophages and epithelioid cells that are surrounded by T cells.\cite{55} For decades granulomas have been considered to be a barrier to bacterial proliferation; however, recent evidence suggests that granulomas not only allow persistence of the pathogen, but can also permit its growth.\cite{55}

To prevent the disease the Bacille Calmette-Guérin (BCG) vaccine was developed in 1921 using an attenuated strain of *Mycobacterium bovis*, the causative agent of TB in cattle.\cite{56} BCG provides partial protection against pediatric TB, but neither BCG nor prior infections are sufficient against adult pulmonary TB. The ineffectiveness of the vaccine can be addressed by enhancing the Th1 response that is pre-induced by BCG.\cite{57} However, the most recent effort to boost the Th1 response by implementing the novel Modified-Vaccinia-Ankara MVA85A did not provide satisfactory results.\cite{58} Therefore, the effort for improving the vaccination continues by attempts to supplement BCG with virulence factors and latency/resuscitation antigens to enhance immune responses that are insufficiently promoted or lacking.\cite{57}

Once the disease has developed, antibiotics come into play. The discovery of streptomycin in the early 1940s marks the beginning of the development of anti TB drugs. However, despite streptomycin and several other antibiotics, such as amino salicylic acid, isoniazid, pyrazinamide, cycloserine, and kanamycin, the length of treatment and the poor efficiency of these drugs rendered the therapy largely ineffective. The major breakthrough came in the 1960s when rifampicin and the weaker antibiotic ethambutol were discovered and applied together with isoniazid and pyrazinamide.\cite{59}

*M. tuberculosis* is difficult to eradicate due to the protection provided by the granulomas and the cell wall rich in mycolic and fatty acids, compromising penetration of antibiotics. Moreover, life cycle studies show that the pathogen develops a dormancy phenotype due to the anaerobiosis and nutrient deficiency experienced in granulomas. These non-replicating persisters are tolerant to anti TB drugs; therefore, a lengthy treatment is necessary.\cite{60} The current TB treatment regiment includes administration of the first-line oral drugs
rifampicin, isoniazid, pyrazinamide, streptomycin, and ethambutol during the first two months, and the treatment during the next four months continues using rifampicin, ethambutol, and isoniazid.$^{[59]}$

Due to poor adherence of the patients to the treatment as well as the insufficient administration and control in the developing countries, multi-drug resistant TB (MDR-TB) strains have become tolerant to the first-line antibiotics isoniazid and rifampicin. Extensively-drug resistant (XDR-TB) strains are even insensitive to isoniazid, rifampicin, any fluoroquinolone, and at least one of the second-line drugs: capreomycin, amikacin, or kanamycin. Moreover, reports about totally-drug resistant (TDR-TB) strains are growing making the surgical treatment to be routinely considered again.$^{[61,59]}$

While treatment of the sensitive strains takes 6-9 months, treatment of MDR-TB and XDR-TB strains takes 18-24 months or longer. Moreover, the first-line anti TB-regiment drugs are currently incompatible with antiretroviral therapies used to treat HIV whereas the second-line drugs are toxic and expensive. Therefore, any novel anti-TB drugs should be safer and more potent to shorten the treatment duration. They should be effective against MDR-TB, XDR-TB, and TDR-TB strains and safe for patients undergoing antiretroviral therapies. Finally, according to WHO, about one-third of the world's population has latent TB; therefore, the novel antibiotics should target *M tuberculosis* in any of its replication and physiological state, including the persisters.

### 1.2.2 Chorismate mutases as anti-TB drug targets

It is generally accepted that a good drug target against a pathogen should be absent in the host cell. Therefore, several novel TB antibiotics target the proteins that are only present in mycobacteria, such as delamanid (OPC-67683), which inhibits mycolic acid synthesis,$^{[62]}$ and bedaquiline, which is highly selective for the mycobacterial ATP synthase.$^{[63]}$

The enzymes from the shikimate pathway could also serve as potential anti TB drug targets due to their absence in mammalian cells. In fact, the shikimate pathway has already been widely targeted by novel fungicides and herbicides. A potent herbicide glyphosate,
inhibiting the sixth enzyme of the pathway, is a prominent example.\cite{2, 64}

*M. tuberculosis*, like several other members of *Corynebacteriales*, exhibits a rare enzymatic system that leads to the biosynthesis of the essential aromatic amino acids Phe and Tyr. The distinct structural and functional features of the MtCM-MtDS complex (Figure 1.5E) allows for development of mycobacteria-specific inhibitors that would be safe to human cells and most of the general gut microbiota.

The thorough validation of MtDS as a potential drug target is still lacking, as there are no reliable deletions of its gene. However, being the first enzyme of the vital biosynthetic pathway for many aromatic compounds, it would most probably serve as a good drug target.

There are also no reports yet of an *aroQ\*\_S* deletion in *M. tuberculosis* nor in its model organism *Mycobacterium marinum*.\cite{65} Assuming that procedural issues of constructing the deletion stains are ruled out, the failure in obtaining reliable knock-outs might indicate that bacteria cannot survive even on rich medium and be retrieved without a functional *aroQ\*\_S* gene. Moreover, even though the gene is assigned as a non-essential gene on the TARGET database,\cite{66} one could assume that hitting the $\approx 270$ bp long MtCM gene by *Himar1* transposon insertion might be difficult due to its small size.

The strongest indication so far that the MtCM-MtDS complex is essential for mycobacterial survival can be deduced from the fact, that the transcription level of MtCM and MtDS genes is similar while growing the bacteria within macrophages and on minimal medium.\cite{67} This implies that these proteins are required for intracellular production of aromatic amino acids and cannot rely on the uptake from the outside.

Taken together, the most potent inhibitor of the aromatic amino acid biosynthesis might be one that simultaneously acts on both MtDS and MtCM. Provided the unique transient interaction of the two enzymes, one could assume, that disrupting the MtDS tetramer will also prevent binding and activating of MtCM thus “locking” it in its catalytically inactive form.

In addition to its house-keeping intracellular MtCM, *M. tuberculosis* presents the secreted *MtCM* as a potential drug target.
There is no indication that *MtCM is essential for bacterial survival, and its gene has been hit by several transposon insertions (TARGET database). Nonetheless, evidence obtained from studying secreted CMs of other organisms indicates their importance in host-pathogen interactions.

Secreted CMs have an N- terminal signal sequence that directs them to the periplasmic space in Gram-negative or to the extracellular space in Gram-positive bacteria, and were also found in fungi and nematodes. The secreted CM from maize pathogen *Ustilago maydis can illustrate their role in host-pathogen interactions. This CM reduces the available chorismate pool needed for pathogen-induced biosynthesis of salicylic acid, which plays a role in the plant’s resistance to the pathogenic fungus. By mediating the metabolic status of its host cell *U. maydis can reduce the plant’s resistance.

However, the highly active CMs that are exported from pathogenic bacteria such as *M. tuberculosis and *Salmonella into their mammalian host cells, which do not have a shikimate pathway, are likely to have a function as virulence factors. Therefore, finding potent inhibitors for *MtCM would not only serve as potential drugs but also as tools for elucidation of its function.

1.2.3 Small molecules or peptides?

When talking of enzyme inhibitors, one typically refers to small molecules that may compete with a substrate in the active site. The field of small molecule identification is constantly expanding and shows great advances with powerful methods for lead identification, such as ligand-based virtual screening, screening of fragment libraries and high-throughput-screens (HTS) of small molecules.

Virtual screening offers the advantage of process automation for testing compounds from chemical databases, such as ZINC, however, the detailed structural information of the target protein is required. In contrast, screening of fragment libraries and small molecule HTSs are better suited for testing the compounds on a greater number of target proteins or even whole cells. The compound libraries for fragment screening are composed of low-molecular mass molecules that follow the so-called rule of three. It states that the
molecular mass of the compounds should be less than 300 Da and the polar surface area less than 60 Å². Also, the number of hydrogen bond donors and acceptors, rotatable bonds and the octanol-water partition coefficient log P should not exceed the number of three.\textsuperscript{78} Due to these requirements the fragments are typically weak binders. However, they serve as interesting starting points for chemical optimization, enabling the development of tailor-made enzyme inhibitors.

The compound libraries for HTS contain bigger, drug-like molecules that follow Lipinski’s rule of five.\textsuperscript{79} The compounds from these libraries, however, are difficult to optimize further since any increase in molecular mass can result in issues with absorption, distribution, metabolism, and excretion.\textsuperscript{80}

In addition to the inhibition of the active sites, disruption of PPIs is also emerging as a promising approach to modulate essential processes in the cell. PPIs constitute the interactome of a cell and exist in a variety of modes. Depending on their composition they can be homo- and hetero-oligomeric. Affinity defines the non-obligate or obligate nature of PPIs, and their lifetime classifies them into transient or permanent complexes.\textsuperscript{81} Transient PPIs are involved in a variety of biological functions such as protein folding assisted by chaperones, protease inhibition, signal transduction, receptor binding, and enzyme allostery (e.g. regulation of AroQ\textsubscript{δ} CMs). Therefore, inhibition of PPIs has gathered a significant interest in pharmaceutical research, due to their importance in oncogenic signaling pathways,\textsuperscript{82} or in diseases such as Huntington’s disease\textsuperscript{83} and Alzheimer’s disease,\textsuperscript{84} where protein misfolding and aggregation occurs.

Despite the rapid development of novel approaches to identify PPI modulators, such as rationally mimicking the regions essential for PPI (hot spot residues), screening combinatorial chemical libraries, and virtual screening of compound databases,\textsuperscript{85,86,87} PPIs remain difficult targets for drug design. One of the reasons for the low success rate is the fact that small molecules cannot offer enough functional groups to efficiently bind to large protein surfaces, which span from 1,500 to 3,000 Å² \textsuperscript{88,89} in comparison to those of protein–small-molecule interactions (\textasciitilde300 – 1,000 Å² \textsuperscript{90}).
Researchers have thus resorted to larger molecules with more functional groups that could target a bigger area of the protein surface.\cite{86} Peptidic modulators can be interesting starting points for developing such molecules,\cite{91,92} since they offer considerable structural diversity and a larger surface for high-affinity binding to the target protein. Moreover, peptides exhibit less toxicity due to their fast degradation into non-toxic amino acids.\cite{93}

Despite the fact that many peptides are clinically used (e.g. the antibiotics vancomycin, gramicidin and tyrocidine, the immunosuppressant cyclosporin A, peptide hormones and analogs, HIV protease inhibitors, or peptide receptor agonists and antagonists \cite{94}) they impose challenges such as poor oral absorption and membrane permeability as well as a rapid eradication by proteases. These challenges can be overcome by a variety of strategies to introduce conformational constraints, such as stapling of helical peptides or generating cycles.\cite{95}

Taken together, the choice of an inhibitor library largely depends on the specifics of the target protein. In this thesis we aimed to find inhibitors for *MtCM* and the MtCM-MtDS complex that offer both the active sites and, in the case of the latter, the PPI for inhibition. Therefore, we employed small molecule (Chapter 3) and peptide libraries (Chapter 4) to explore their potential to inhibit these enzymes.

### 1.3 Chorismate mutases as models to investigate enzyme catalysis

Enzymes are indispensible constituents of the living world. Faced with seemingly limitless resources of natural enzymes, humankind has always sought to employ their exquisite features for improvement of life. The fundamental understanding of how these biological molecules accelerate and fine-tune chemical reactions to physiologically relevant magnitudes, lays the groundwork for creation of novel catalysts and the development of medicines.

It took nearly a century from 1878, when Friedrich Wilhelm Kühne realized that a substance within the yeast cells mediates fermentation reactions, until 1965 when the first crystal structure of
an enzyme (in Greek: en (in) zyme (yeast)) was solved by X-ray crystallography.\textsuperscript{[96]} It was quickly realized that only a small fraction of an enzyme – the active site – is used for catalysis. These surface clefts or pockets are equipped with specific residues that accommodate the substrate and convert it into a product by decreasing the activation energy of the reaction. As early as 1896, Emil Fischer hypothesized that the active site ("the lock") is pre-organized to bind a substrate ("the key"), thus explaining the specificity of the enzymes. However, it was observed that the active sites could also accommodate slightly different substrates and were complementary to the transition state of the reaction rather than the substrate. This resulted in Daniel Koshland’s \textit{induced fit} theory stating that the active sites are solvated until the substrate binds simultaneously inducing the catalytically competent shape of the binding pocket.\textsuperscript{[96]}

The binding process is responsible for substrate affinity and specificity and thus the chemo-, regio-, and stereoselectivity of enzymes. For the enzymes that exhibit Michaelis-Menten-type kinetics this phenomenon is described by the affinity constant $K_m$, representing the substrate concentration at half maximal velocity. The remarkable rates achieved by enzymes, is described by the catalytic rate constant $k_{\text{cat}}$, also called turnover number. The overall effectiveness of an enzyme is thus measured by the $k_{\text{cat}}/K_m$ value, or the enzyme’s catalytic efficiency, which effectively combines the effectiveness of substrate binding (low $K_m$) and the high turnover number (high $k_{\text{cat}}$).

The $k_{\text{cat}}/K_m$ parameter is widely used to describe how efficient enzymes are. In contrary, to the prominent example of TIM, which operates at diffusion-controlled limits ($k_{\text{cat}}/K_m \approx 10^8$) and is, therefore, regarded as a ‘perfect enzyme’,\textsuperscript{[97]} most of the natural enzymes catalyze their reactions with much lower efficiencies.\textsuperscript{[98]} This is likely due to the fact that natural enzymes have evolved to serve the overall fitness of an organism by operating at biologically relevant concentrations of metabolites. In fact, it is observed that $K_m$, is usually within the range of intracellular substrate concentrations.\textsuperscript{[99]} On the other hand, the factors that influence the $k_{\text{cat}}$ are less clear and catalytic effects coming from electrostatics, hydrogen tunneling, strain or distortion, and overall enzyme dynamics have been recognized.\textsuperscript{[100, 101, 102, 103]}
Regulation, effectors, and evolvability of chorismate mutases from Corynebacteriales

Upon substrate binding, the binding energy can be deployed for rearrangements in the enzyme that facilitate the chemical reaction. Thereby, the protein exploits its energy landscape to adopt configurations that suitably position active site residues for the chemical catalysis. The source of these motions might not only be the active site, but can also be regions far away from it, explaining why effector binding, mutations or post-translational modifications at a distant locations can have a big impact on an enzyme’s activity.

Finally, protein function cannot be maintained without a substantial stability of the fold. However, it was observed that most proteins have evolved to be only marginally stable.\textsuperscript{[104]} It was rationalized that it is the function, but not the stability, which is the primary driving force in natural evolution. However, protein stability might be a prerequisite for introducing new mutations that enhance catalytic efficiency or introduce new functionalities in existing proteins.\textsuperscript{[105]}

The two fundamental features of enzymes – the catalytic activity and high stability – are also the most relevant for their application. Understanding the mechanistics of enzyme catalysis will allow tailoring them for greater stereo- and regioselectivity, open up possibilities to expand the substrate scope for the existing enzymes, as well as create novel enzymes with abiological functions. In parallel, enhancing enzyme stability will provide us with easily producible and robust catalysts that operate under a desired set of conditions.

Along with the efforts to understand and tailor their intriguing mechanisms, researchers have tried to create new enzymes from scratch. Such \textit{de novo} enzyme design goes hand-in-hand with fundamental studies of enzymatic catalysis, since our knowledge of how enzymes actually work is still too limited to create them anew. We can only claim to have truly understood how these molecular machines work, if we are able to design them. Until then, the efforts to elucidate the relationship between protein structure and function must continue.
Chapter 1: Introduction

1.3.1 Directed evolution

Protein engineering allows altering enzymes, which in turn results in deeper understanding of the factors governing their catalysis. The development of recombinant DNA techniques and pioneering work by Michael Smith marked the dawn of protein engineering. Smith and colleagues implemented an earlier concept that gene sequences could be altered by using synthetic oligonucleotides,\textsuperscript{[106]} which earned him the Nobel Prize for Chemistry in 1993. Since then, together with the development of X-ray crystallography, the relationship between a protein’s primary sequence, its structure, and its function has been probed by a variety of methods.

Rational approaches of protein engineering, empowered by extensive knowledge about structure, reaction mechanism, and homologous proteins have sometimes been successful in enhancing protein stability and activity.\textsuperscript{[107,108]} However, these successes typically exploited only a small fraction of the possibilities that could be tackled to modulate proteins.

To also sample an enzyme’s sequence space that is less well characterized or not known at all, an approach inspired by the insights outlined in Darwin’s \textit{On the Origin of Species}, is implemented.\textsuperscript{[109]} The so-called \textit{directed evolution} (or \textit{laboratory evolution}) involves the generation of a library of mutagenized genes followed by screening or selection for the biological entity exhibiting a desirable feature, such as higher activity, selectivity or stability. Several cycles of diversification and selection can be performed in an iterative fashion to accumulate beneficial mutations until the desired level of improvement is reached.

Directed evolution involving random mutagenesis is widely used as a comprehensive and unbiased approach to sample the vast sequence space of proteins. The concept of \textit{sequence space} describes the number of possible combinations of amino acids that a given polypeptide stretch can contain. For instance, if every position of a rather small 100-residue protein (N=100) is probed by any of the 20 natural amino acids, the number of possible combinations is $20^N$ or $\approx 1.3 \times 10^{130}$. This diversity, which alone is greater than the number of atoms in the known universe, is even further increased by non-canonical amino acids, disulfides, cofactors, posttranslational
modifications, or multi-subunit organization. Moreover, most randomly assembled sequences will result in inactive proteins and one needs to ‘hit’ the right regions in sequence space to obtain active enzymes. Therefore, due to the limited possibilities to sample all combinations, the key to a successful directed evolution experiment lies within the ability to simultaneously optimize the strategies of gene diversification and screening/selection.

Gene diversification is the first step of generating large libraries of genes. Depending on the experiment it can target the entire gene or pre-selected sites using random or targeted mutagenesis strategies, respectively. Random mutagenesis is perhaps the most common method used to probe the entire gene, particularly if structural information is missing. Even though there are a variety of methods to introduce point mutations (e.g. chemical or physical agents, such as alkylating or deaminating compounds, base analogs, UV irradiation, or mutator strains\(^\text{[110]}\)), the single gene mutagenesis via an error-prone polymerase chain reaction (epPCR)\(^\text{[111]}\) has often emerged as the method of choice due to higher mutation rates and better control.

Furthermore, single mutations generated by epPCR can be recombined in a random fashion using methodologies that mimic natural recombination. The original method of recombining artificial mutations, known under the term DNA shuffling, was described by Willem P. Stemmer:\(^\text{[112]}\) It employs PCR to re-assemble small DNase-generated DNA fragments of the target gene thereof additionally introducing point mutations. This allows producing more diverse pools of variants, potentially recombining interesting mutations, than possible by epPCR alone.

Although epPCR and DNA shuffling with a variety of improvements exist\(^\text{[110]}\) and are indispensable for testing large portions of sequence space, they impose extensive screening efforts to sample the entire library of variants. Therefore, a gene diversification strategy called saturation (or cassette) mutagenesis is often implemented, that focuses randomizing mutagenesis to predetermined sites in the sequence. This approach, however, requires structural or homology information, but allows sampling of all the possible amino acid substitutions per codon and can reduce the library size to be screened. It introduces mutations into the target area via a mutagenic cassette\(^\text{[113]}\)
that contains degenerate codons allowing for all or a subset of the natural amino acids to be encoded per position. The most popular degenerate codon is NNK (N encodes any of the nucleotides A, T, G or C, while K encodes G or T only), which admits all 20 amino acids and only one of the three possible stop codons, thus reducing the probability of generating truncated library members.

Sampling the gene library members imposes the next challenge in directed evolution. The general concept for identification of the winner gene among the thousands of variants is spatial separation and the linkage of the coding gene to the property of interest. Often analytical tools to monitor the desired property include UV/vis absorbance, fluorescence, high-performance liquid chromatography (HPLC), gas chromatography (GC), mass spectrometry (MS) or nuclear magnetic resonance (NMR).

Individual library members are most commonly assayed in 96-well microtiter plates. To increase throughput, methodologies such as fluorescence-activated cell sorting (FACS)\textsuperscript{114} or the newly emerging droplet-based microfluidics\textsuperscript{115} can be used. The microfluidics technology is built on the \textit{in vitro} compartmentalization (IVC) approach pioneered by Dan S. Tawfik and Andrew D. Griffiths\textsuperscript{116} and enables encapsulation of a single cell or individual DNA library members using water-in-oil emulsions. It allows for ultra-high-throughput, which approaches the efficiency of \textit{selection systems}.

Selection systems enjoy unparalleled high throughput. They link the phenotype to its genotype via tight physical interaction (display methods) or in space (\textit{in vivo} selection). Basic protein display strategies rely on tight and specific binding to the immobilized target molecule by the library members that are in turn associated with their encoding genes. Cell-surface (yeast or \textit{E. coli}) and phage displays utilize the cell wall and the bacteriophage capsid as a compartment to preserve the encoding DNA, respectively.

Generation of large libraries using these methods is generally limited by the efficiency of transformation. Nevertheless, this can be overcome by ribosome display, which relies on \textit{in vitro} transcription and translation\textsuperscript{117} or mRNA display, which involves tethering of the target protein to its mRNA.\textsuperscript{118}
Display methods have proven themselves useful for evolving high-affinity binders,[119] however, cases where catalytic activity was evolved using this technology are rare.[120] The alternative approach to evolve enzyme activity is in vivo selection, which links the desired activity to bacterial survival under selective conditions. For instance, an evolved functional enzyme complements the metabolic defect of an auxotroph and allows cell survival on minimal medium lacking the otherwise required metabolite. Interestingly, in vivo selection systems can also be modified to link the desired properties to antibiotic resistance or shielded exposure from toxic molecules. These approaches have been successfully implemented to evolve a highly active recombinase[121] and aminoacyl-tRNA synthetases,[122] or even encapsulation properties of a protein container.[123]

CMs are essential metabolic enzymes allowing the development of a powerful in vivo selection system for directed evolution.[51] The CM selection system is based on E. coli strain KA12, which has the CM-encoding pheA and tyrA genes deleted.[51] To make the strain dependent on the CM reaction only, the missing genes for a monofunctional PDT and a PDH from Pseudomonas aeruginosa and Erwinia herbicola, respectively, were supplemented on the plasmid pKIMP-UAUC.

In this way, the survival of KA12/pKIMP-UAUC on minimal medium lacking Phe and Tyr becomes solely dependent on a functional CM gene, which may be provided on a plasmid carrying gene library members. Typically, this in vivo selection allows assessment of up to 10⁷ independent CM variants and is only limited by the transformation efficiency of KA12/pKIMP-UAUC.

Due to the availability of a powerful selection system, CMs have served as model enzymes to elucidate the secrets of enzymatic structure-function relationships in a number of studies,[51, 124, 125, 126] In Chapter 5 of this work, we apply the CM selection system to elucidate the activation mechanism of MtCM.
1.4 Aims of this thesis

The AroQ₈ CMs present a unique system for elucidation of how enzymes work. Their minimalistic active site, ability to switch between active and inactive states and the unique response mechanism to allosteric effectors has intrigued our research group for nearly a decade. Furthermore, CMs of this subclass are found in pathogenic and in industrially important organisms presenting an attractive area for specific drug design and improvement of commercial biosynthesis of aromatic compounds.

In this study, we set out to explore (i) how AroQ₈ enzymes are feedback regulated, (ii) whether *M. tuberculosis* CMs could be exploited for finding novel anti-TB drugs, and (iii) factors governing the activity switch of MtCM to foster the fundamental understanding of AroQ₈ CM catalysis.

In Chapter 2 we test the response of the AroQ₈ CM from *C. glutamicum* to the cognate DAHP synthase to investigate whether the same activation phenomenon described previously for the closely related MtCM would also exist in *C. glutamicum*. We aimed to answer how general the observation is that AroQ₈ CMs, which lack otherwise conserved active site residues and which are weak catalysts on their own, can be activated by the first enzyme of the shikimate pathway. Also, we aimed to understand how such an activation process via this transient PPI permits the regulation of CM activity by the end products of the shikimate pathway.

In Chapters 3 and 4 we designed an *in vitro* screen for small molecules and an *in vivo* selection system for constrained peptides to identify potential inhibitors of the MtCM-MtDS complex. By this we attempted to evaluate the potential of small molecule compounds and peptides to modulate AroQ₈ CMs and PPIs. Moreover, we also included the secreted *MtCM* as a target for the small-molecule screen as it would be intriguing to find inhibitors of this potential pathogenicity factor.

Finally, Chapter 5 describes directed evolution experiments to elucidate the mechanism used by MtCM to achieve catalytic proficiency. We asked whether AroQ₈ CMs would in principle possess the inherent ability to evolve into highly active CMs without the help of
the interaction partner. We hope that our results will give some insight into what sequence and structural factors are necessary for catalytic activity of these enzymes.
2 Inter-enzyme allosteric regulation of AroQδ chorismate mutases

Parts of this chapter have already been published or will soon appear in print:


2.1 Introduction

As detailed in sections 1.1.1 and 1.1.2, the organisms have developed diverse strategies to regulate DAHP synthases and CMs from the shikimate pathway.

In the organisms of the order Actinobacteria, as in many others, several gene copies of DAHP synthases can be present. For example, C. glutamicum ATCC 13032 has two genes – encoding for a Type II (CgDS, cg2391, aroG) and a Type I (cg1129, aroF) enzyme – whereas M. tuberculosis H37Rv has only a Type II DAHP synthase (MtDS) gene (Rv2178c, aroG). Even though C. glutamicum has two DAHP synthase genes, CgDS might be the dominant one. It was shown, that deletion of aroF in C. glutamicum RES167 did not alter the growth phenotype on minimal medium, whereas deleting aroG resulted in aromatic amino acid auxotrophy.\(^{[127]}\) The aroF gene occurs in the mca operon\(^{[128, 129]}\) under control of the aroR attenuator.\(^{[130]}\) The corresponding leader peptide contains an F-Y-F motif suggesting attenuation by Phe and Tyr; however, only the repression by Phe was analyzed and confirmed so far.\(^{[130]}\)

Tight control of DAHP synthase activity is achieved not only by regulation of gene expression but also by allostERIC control. For instance, the Type I DAHP synthase from C. glutamicum is strongly inhibited by Tyr alone,\(^{[127]}\) whereas MtDS and CgDS have developed a more sophisticated regulatory mechanism. Tyr and Phe synergistically inhibit MtDS and CgDS. The strongest inhibition of these enzymes occurs when Tyr, Phe, and Trp are present simultaneously;\(^{[131, 132, 26, 28]}\) however, Trp alone can have a weak stimulatory effect.\(^{[132]}\)

In addition to the tight regulation of DAHP synthases, intracellular CMs are also subject to allostERIC control. MtCM and the CM from C. glutamicum (CgCM) belong to the structural AroQ\(_8\) subclass and are only mediocre catalysts by themselves.\(^{[133, 48]}\) They form a non-covalent complex with MtDS and CgDS, respectively, to become responsive to increasing concentrations of aromatic amino acids.\(^{[46, 134, 48]}\)

MtCM and CgCM share 59 %, while MtDS and CgDS share 66 % of sequence identity (according to a BLAST comparison) and form essentially identical non-covalent complexes (Figure 1.5E and 2.3).\(^{[135, 48]}\) The MtCM activity boost by more than a 100-fold following complex
formation with MtDS was confirmed in several independent studies.\cite{48,136,28} However, the reports about the magnitude of CgCM activation by CgDS present contradictory results. In a recent study no activation was observed,\cite{133} which is in stark contrast to earlier reports where the activity of CM in enzyme preparations from \textit{Brevibacterium flavum} (later re-classified as \textit{C. glutamicum})\cite{137} was only detectable in the presence of DAHP synthase.\cite{46,138}

This chapter investigates the effect of CgDS on the activity of CgCM as well as the allosteric regulation of the complexed MtCM and CgCM at biologically relevant effector concentrations. We also demonstrate that the allosteric regulation of complexed MtCM is achieved via the dissociation of the MtCM-MtDS complex and suggest that the same principle applies for the CgCM-CgDS complex as well.

\section*{2.2 Results}

\subsection*{2.2.1 CgCM is activated by complex formation with CgDS}

The kinetic parameters of native CgCM in the presence or absence of the His-tagged CgDS were determined to elucidate the activity changes upon complex formation. It has been shown previously that the His tag had little influence on the activity of MtDS\cite{26,28} and it is reasonable to assume that the His tag would not have much of an effect on the structurally very similar CgDS.

Much like MtCM, CgCM is a mediocre enzyme on its own. The activity of CgCM without the complex partner follows Michaelis-Menten kinetics with a catalytic efficiency $k_{\text{cat}}/K_m$ of $1.1 \times 10^2$ M$^{-1}$ s$^{-1}$, which is within a factor of three of previously published values (Table 2.1).\cite{133,48} Notably, the CgCM activity is one order of magnitude lower than that of MtCM (Table 2.1) and at least three orders of magnitude lower than the catalytic efficiencies ($k_{\text{cat}}/K_m \approx 10^5$-10$^6$ M$^{-1}$ s$^{-1}$) determined for typical AroQ CMs.\cite{139,36}

The CM activity of the CgCM-CgDS complex was assayed in the presence of the DAHP synthase-stabilizing agents Mn$^{2+}$, TCEP, and PEP as described previously for activity assays of MtCM-MtDS\cite{48} and MtDS.\cite{140} The catalytic parameters of CgCM in the presence of a 100-fold excess of CgDS exhibited a 180-fold efficiency boost to a $k_{\text{cat}}/K_m$ of
2.0 × 10^4 M⁻¹ s⁻¹ (Table 2.1), which clearly contradicts the data from Li and co-workers,[133] but is in agreement with the earlier studies.[46, 138]

The overall 180-fold activation of CgCM when in complex with CgDS is slightly higher than the 140-fold activation of MtCM in the presence of MtDS. However, the apparent catalytic efficiencies of these enzyme complexes critically depend on the apparent $K_{d,app}$ of the CM-DAHP synthase complex, which combines the $K_d$ values of DAHP synthase tetramer, CM dimer, and other assay components; therefore, the total protein concentration plays an important role. The $K_{d,app}$ for the MtCM-MtDS complex could be determined as 140 nM,[48] since MtCM could be titrated with MtDS. However, the apparent $k_{cat}/K_m$ of the CgCM-CgDS complex showed a continuous linear increase when the CgDS concentration was varied from 1 to 4 μM (data not shown) indicating that the $K_{d,app}$ of the CgCM-CgDS complex must be above 4 μM. Because the CgDS concentration in the present study had to be limited to 1 μM, the overall activation of CgCM is expected to be even higher at CgDS concentration above this value.

### Table 2.1 Kinetic parameters of investigated chorismate mutases

<table>
<thead>
<tr>
<th>Protein</th>
<th>$k_{cat}$, (s⁻¹)</th>
<th>$K_m$, (μM)</th>
<th>$k_{cat}/K_m$, (M⁻¹ s⁻¹)</th>
<th>Fold activation by DAHP synthase</th>
</tr>
</thead>
<tbody>
<tr>
<td>CgCMᵃ</td>
<td>1.05 ± 0.19</td>
<td>2830 ± 530</td>
<td>317</td>
<td></td>
</tr>
<tr>
<td>CgCMᵇ</td>
<td>ND</td>
<td>ND</td>
<td>111 ± 7</td>
<td></td>
</tr>
<tr>
<td>CgCM + CgDSᶜᵈ</td>
<td>ND</td>
<td>&gt; 2000</td>
<td>(2.00 ± 0.01) × 10⁴</td>
<td>182 ± 12</td>
</tr>
<tr>
<td>MtCMᵉ</td>
<td>2.0 ± 0.1</td>
<td>1140 ± 20</td>
<td>1750 ± 90</td>
<td></td>
</tr>
<tr>
<td>MtCM + MtDSᶠ</td>
<td>8.1 ± 1.9</td>
<td>34 ± 3</td>
<td>(2.4 ± 0.6) × 10⁵</td>
<td>140 ± 35</td>
</tr>
</tbody>
</table>

ᵃ Data from[133] measured in 50 mM Tris/HCl buffer, pH 7.5 containing 1 mM DTT.
ᵇ Measured in 50 mM BTP, pH 7.5, using 950 nM CgCM.
ᶜ Measured in 50 mM BTP+ (50 mM BTP, pH 7.5, additionally containing 1 mM TCEP, 0.2 mM PEP, and 0.1 mM MnCl₂) using 50 nM CgCM and 1000 nM CgDS. Because the $K_{d,app}$ is > 4 μM, the values must be regarded as apparent $k_{cat}/K_m$ and apparent activation factors due to the fact CgCM cannot be saturated with CgDS in the assay used here.
ᵈ The crystal structure revealed some Trp tightly bound in CgDS; therefore, the activation by CgDS alone might be somewhat overestimated.
ᵉ ND, not determined, since substrate saturation could not be achieved using up to 3 mM chorismate. Measuring at higher substrate concentrations was not possible due to the high chorismate absorption of the substrate at 310 nm and the relatively high background reaction.
ᶠ Data from[48] measured in 50 mM BTP+ pH 7.5, using 10 nM MtCM and 2000 nM MtDS ($K_{d,app}$ = 140 nM).
Chapter 2: Inter-enzyme allosteric regulation of AroQδ chorismate mutases

Moreover, it is important to note that even minor contamination with Trp might also affect the extent of CgCM activation in the complex with CgDS (see section 2.2.2 below). In fact, the crystal structure of CgDS revealed some tightly bound Trp even though it was not provided in the crystallization solution.\textsuperscript{[135]} Since the same CgDS protein preparation was used in the presented kinetic experiments (Table 2.1), it is reasonable to assume that a part of the CgCM activation observed might be due to Trp effect.

Interestingly, in contrast to the MtCM-MtDS system, the $K_m$ of the activated CgCM remains high ($K_m > 2 \text{ mM}$) consistent with the earlier studies showing a $K_m$ in the range of 2.9 to 5.5 mM.\textsuperscript{[141, 134]} This indicates that the activated CgCM has a much weaker substrate affinity than activated MtCM ($K_m = 34 \text{ µM}$); however, $K_m$ may improve at enzyme concentrations above the $K_{d,\text{app}}$ of the CgCM-CgDS complex as well as in the presence of Trp.

2.2.2 Impact of the feedback inhibitors Phe, Tyr, and Trp on CM-DS complex activity.

The end products of the shikimate pathway Phe, Tyr, and Trp are allosteric modulators of MtDS and CgDS, but do not affect MtCM or CgCM in the absence of their complex partner.\textsuperscript{[133, 48]} However, complex formation with DAHP synthase renders these CMs sensitive to the pathway end products.\textsuperscript{[141, 138, 48]}

The effects of aromatic amino acids on MtCM-MtDS and CgCM-CgDS were previously determined at rather high (0.1 - 0.2 mM) concentrations.\textsuperscript{[138, 48]} These concentrations probably do not represent the conditions in the cell, since, for instance, the concentrations of aromatic amino acids in \textit{E. coli} had been determined to be 29 µM, 18 µM, and 12 µM for Tyr, Phe, and Trp, respectively.\textsuperscript{[142]} Therefore, we examined the impact of aromatic amino acids at 25 µM on both the MtCM-MtDS and CgCM-CgDS complexes and assume that these conditions are physiologically more relevant.

The state assembly of a transient protein complex is governed by the $K_{d,\text{app}}$ and thus the total protein concentration in the assay. We assumed, that the shifts in the population of the individual protein ensembles might play an important role in their response to the
Regulation, effectors, and evolvability of chorismate mutases from *Corynebacteriales*

Aromatic amino acids. Therefore, the feedback inhibition response of the MtCM-MtDS complex was investigated at concentrations that are below and above the $K_{d,app}$ of 140 μM. For the CgCM-CgDS complex only the conditions below the assumed $K_{d,app}$ (> 4 μM) could be tested due to protein precipitation issues at high enzyme concentration.

**Figure 2.1 Modulation of CM activity by Phe, Tyr, and Trp.** The initial velocity was monitored at 30°C using 100 μM and 23 μM chorismate for CgDS-CgCM (blue) and MtCM-MtDS (green, brown), respectively. The concentration of each aromatic amino acid was 25 μM. The specific initial velocities of chorismate consumption monitored at 274 nm representing 100% were 1.5 ± 0.06 s⁻¹, 4 ± 0.2 s⁻¹ and 0.4 ± 0.1 s⁻¹ in the presence of 100 nM CgCM (1000 nM CgDS), 30 nM MtCM (300 nM MtDS) and 25 nM MtCM (25 nM MtDS), respectively. The error bars show standard deviations of several measurements. For the CgCM-CgDS system, four measurements with two independently purified CgCM batches were plotted. For the assays with the higher MtCM-MtDS concentrations four data points were averaged. Two data points were averaged for the 25 nM MtCM (25 nM MtDS) kinetic measurements using two independently purified MtCM batches.

As shown in Figure 2.1, at protein concentrations above the $K_{d,app}$ of the MtCM-MtDS complex (30 nM MtCM and 300 nM MtDS), the impact of the single amino acids is weaker than for the CgCM-CgDS complex. Tyr and Trp had essentially no effect on MtCM-MtDS,
whereas Phe exhibited the largest inhibitory effect reducing the MtCM activity to the remaining 83 %. However, when Phe was present in combination with Tyr, the MtCM activity dropped to 30 % indicating synergistic inhibition by these two amino acids. Trp had no effect on the MtCM-MtDS complex, neither alone nor in combination with Phe, Tyr, or both Phe and Tyr.

A similar response to Phe and Tyr was exhibited by the MtCM-MtDS complex when the protein concentration was below the $K_{d,app}$ (25 nM MtCM and 25 nM MtDS). Phe alone was inhibiting more strongly than the other amino acids, but to the same 83 % residual activity as seen before. The strongest inhibition was achieved when Phe was added in combination with Tyr, inhibiting the CM activity to remaining 15 %, which is 2-fold stronger than at the conditions above the $K_{d,app}$. Interestingly, at 25 nM of MtCM and MtDS, Trp caused a substantial 2.5-fold activation while Tyr activated only marginally, probably within the experimental error. Trp could also partially revert the inhibitory effect of Phe (to 97 % residual CM activity) and retain a 2-fold activating effect even in combination with Tyr. When all three aromatic amino acids were tested on the MtCM-MtDS complex at low protein concentration, the inhibitory effect was identical to the one observed at high protein concentration, probably due to a slight tendency of Trp to counteract the strong inhibition by Phe and Tyr.

The inhibitory effects of aromatic amino acids on the CgCM-CgDS complex (100 nM CgCM and 1000 nM CgDS) at the protein concentration below the assumed $K_{d,app}$ (> 4 μM), followed a similar pattern as for the MtCM-MtDS complex, however, were more pronounced (Figure 2.1). The strongest inhibitor of the CgCM-CgDS complex was Phe; it reduced CM activity to 10 % whereas Tyr inhibited CgCM to 77 % residual activity. Similarly to the MtCM-MtDS complex, the combination of Tyr and Phe caused the strongest inhibition of CgCM-CgDS. These effectors inhibited CgCM to 4.6 % residual activity, which is 6-fold stronger than the inhibition seen for the complexed MtCM, and brought down the CM activity level close to that of the uncomplexed CgCM (Figure 2.1). In line with Trp activation reported earlier,$^{[141]}$ a 2.5-fold Trp activation was observed for the CgCM-CgDS complex in parallel to that for MtCM-MtDS at protein concentrations below the complex $K_{d,app}$. Trp could also partially revert the inhibition
by Phe alone (increasing the remaining activity from 10 % to 22 %) and by Phe in combination with Tyr (from 4.6 % to 7.5 %). The weak Tyr inhibition was also nearly completely reverted by Trp – to 240 % remaining activity, which is within the experimental error of Trp activation alone.

Taken together, our data suggests that AroQ₈ CMs of certain Actinomycetales are feedback regulated by Phe, Tyr, and Trp via the transient CM-DAHP synthase complex. Phe and Tyr are synergistic inhibitors, while the exact response to Trp depends on the $K_{d,app}$ of the complex and thus the intracellular enzyme concentration.

### 2.2.3 Assembly of the MtCM-MtDS complex in solution

CgCM and MtCM undergo significant conformational changes upon complex formation with the corresponding DAHP synthase.$^{[135, 48]}$ These changes induce a catalytically more efficient orientation of active site residues and thus are the basis of the activity increase of these enzymes.

We assumed that feedback regulation by the pathway end products is achieved via the reversible dissociation of the complex rather than by binding of the aromatic amino acids to the complexed CMs, based on the following observations: (i) the kinetic data suggest that aromatic amino acids do not affect CgCM and MtCM when the interaction partner is absent.$^{[133, 48]}$ (ii) no binding sites of Phe, Tyr or Trp were identified next to MtCM or CgCM in the crystal structure of the corresponding complex and (iii) no significant conformational changes in MtCM were observed in the crystal structure of the MtCM-MtDS complex with or without the allosteric inhibitors Phe and Tyr.$^{[28]}$

Here, we investigate the behavior of the MtCM-MtDS complex in solution when aromatic amino acids are present. Previous attempts to investigate MtCM-MtDS complex dissociation by SAXS (small-angle X-ray scattering) and preliminary SEC (size-exclusion chromatography) experiments did not yield conclusive results.$^{[26]}$ The SAXS experiments failed to show the dissociation of the complex in the presence of aromatic amino acids probably due to the experimentally required high protein concentrations that shift the equilibrium towards complex formation.
Therefore, we have performed HPLC SEC experiments at enzyme concentrations that allowed monitoring complex formation and dissociation. The concentrations at approximately equimolar ratio of MtCM and MtDS were chosen to be close to the determined $K_{d,\text{app}}$ of 140 nM\cite{48} on the SEC column. Collected HPLC fractions were analyzed by SDS-PAGE since the strong absorption of required ligand chorismate (see below) at 280 nm precluded analysis by the standard optical methods.

**Figure 2.2 Interaction of MtCM with MtDS in the presence of chorismate and FYW.** The samples containing 15μM MtDS (A, D and G), 10 μM MtCM (B, E and H) and 15μM MtDS together with 10 μM MtCM (C, F and I) were subjected to SEC and the eluted fractions were analyzed by SDS-PAGE. SEC experiments were performed in a buffer containing 20 mM BTP, pH 7.5, 150 mM NaCl, 1 mM TCEP, 0.2 mM PEP, 0.1 mM MnCl$_2$, and 100 μM chorismate without additives (A, B and C) with 100 μM Tyr and 100 μM Phe (D, E and F) or with 100 μM Tyr, 100 μM Phe, and 100 μM Trp (G, H and I). In all cases, the numbers on the top of the gels represent consecutive fractions, M stands for the Amersham Low Molecular Weight marker (GE Healthcare), and the molecular mass values are in kDa.
As depicted in Figure 2.2 the elution volume of the individual proteins by themselves was not affected by the presence of the effectors indicating that they had no significant effect on the enzymes’ quaternary structure at the protein concentrations tested (compare panels A, D and G or panels B, E and H in Figure 2.2).

Moreover, clear co-elution of MtCM with MtDS was observed in the absence of effectors (Figure 2.2, panel C), which is counteracted by the addition of Phe and Tyr (Figure 2.2, panel F). At the enzyme concentrations employed here, the addition of Trp had no additional effect on complex dissociation apparent from similar MtCM elution patterns with and without Trp (Figure 2.2, panel F vs. I).

The SEC experiments shown here and those published recently confirm that the MtCM-MtDS complex is highly dynamic and can be detected only when the substrate chorismate is present.[48, 27, 28] This suggests that the reduction of CM activity of MtCM-MtDS in the presence of Tyr and Phe is achieved via the reversible dissociation of the complex. The greatest inhibition and least co-elution occur when Tyr and Phe are both present and the addition of Trp, at least at these protein concentrations, does not influence co-elution behavior.

Due to structural and functional similarity of MtCM-MtDS and CgCM-CgDS complexes we assume that the regulation of CgCM activity proceeds via shifting the complex formation equilibrium in a similar fashion as in the MtCM-MtDS system.

2.3 Discussion

*C. glutamicum* ATCC 13032 is a non-pathogenic, free-living bacterium whereas the phylogenetically related *M. tuberculosis* H37Rv is a human pathogen. Thus, the different mechanisms how these organisms control the aromatic amino acid biosynthesis may well reflect the biological needs in very different natural habitats.

*M. tuberculosis* possesses only one gene for a Type II DAHP synthase, whereas *C. glutamicum* encodes both a Type I and Type II enzymes. The presence of two genes in *C. glutamicum* might provide higher robustness to environmental changes, while the disappearance of a second DAHP synthase gene in *M. tuberculosis* might have occurred by gene deletion to (i) prevent energetically costly expression of an
additional gene,[143] (ii) remove additional regulation machinery, or (iii) adapt to a reduced need for environmental robustness due to strict tissue and host specificity of the pathogen. In fact, the pathogenic Corynebacterium diphtheriae, a close relative of C. glutamicum, has also lost one of the two DAHP synthase genes still present in the common ancestor of Corynebacteria.[144]

Interestingly, both C. glutamicum ATCC 13032 and M. tuberculosis H37Rv, use Type II DAHP synthases to activate intracellular CMs via non-covalent complex formation. In this study, we show that similarly to MtCM, the mediocre CgCM activity is boosted upon interaction with CgDS by 180-fold. Our results are in sharp contrast to the recent study of Li et al.[145] which reported no activation. It is conceivable that the lack of activation observed by Li and co-workers, might be due to low DAHP synthase concentration in the assay and/or the absence of assay components such as PEP, reducing agent (TCEP) and metal ions. Previous studies have shown that metal ion cofactors are important for the activity and stability of Type II DAHP synthases.[140,127] Moreover, the binding site of Mn$^{2+}$ in CgDS is relatively close to the interface with CgCM; therefore small perturbations in this area might preclude the complex formation.

In addition to activation of AroQδ CMs, it has been reported that Type II DAHP synthases are also slightly activated upon interaction with CMs in both organisms. Li et al observed that CgCM not only activates CgDS by 3-fold but also subjects it to weak inhibition by chorismate and prephenate.[133] However, these data should be considered with caution, as they contradict the earlier studies, which indicate a much weaker CgDS activity increase of only 30 %[134] that is similar to the marginal activation of MtDS seen upon complex formation with MtCM.[136]

Complex formation with Type II DAHP synthase establishes allosteric regulation of the intracellular CMs in C. glutamicum ATCC 13032 and M. tuberculosis H37Rv. Thereby, Phe is the stronger, Tyr is the weaker inhibitor, and a combination of Tyr and Phe or all three aromatic amino acids severely inhibits the CM activity of both MtCM-MtDS and CgCM-CgDS complexes. Since hardly any conformational changes were observed in the DAHP synthase x-ray structures upon
Regulation, effectors, and evolvability of chorismate mutases from *Corynebacteriales*.

Trp, Phe, or Tyr binding,[28] it is possible that regulation of CM activity proceeds solely by complex dissociation dynamics.

Trp activation, which is also common to both systems, adds another variable into the complex picture drawn by inter-enzyme allosteric regulation of AroQ<sub>δ</sub> CMs. However, it only seems to appear under conditions that favor complex dissociation (i.e. below $K_{d,app}$). For the CM-DS complex, the DS tetramer must be pre-formed prior to binding CM, since it binds at the DS tetramer interface (Figure 2.3). Therefore, the individual $K_d$ values, of both the DS tetramer and the CM-DS complex might play a role in the complex formation equilibrium at any given enzyme concentration.

**Figure 2.3 Crystal structure of the MtCM-MtDS complex with bound Phe and Tyr (PDB ID: 5CKX).** The two dimers of MtCM (subunits are shown in cyan and wheat) bind to the MtDS homotetramer (two subunits of a tight MtDS homodimer are shown in two shades of blue). Phe and Tyr represented in red and green spheres, respectively, bind at the dimer interface. Trp binding sites (black circles at the tetramer interface) are shown according to the MtDS structure with all three aromatic amino acids bound (PDB ID: 5CKV, see Figure 1.3E). Mn<sup>2+</sup> in the active site of MtDS is represented by purple spheres. The CgCM-CgDS complex adopts a very similar structure (D. Burschowsky, unpublished results[135]).
The fact that Trp activation is only observed if the protein concentration is below the $K_{d,\text{app}}$ of the CM-DS assembly, might suggest that Trp drives complex formation and thereby increases the total CM activity. Alternatively, the protein concentrations tested might have been below the $K_d$ of the DS tetramers. In this case, Trp might drive DS tetramer formation, which then results in more CM-DS assemblies. Either way, the assumption that Trp is required for a stable CM-DS complex at low protein concentration needs further experimental evidence.

Intriguingly, however, the response to Phe and Tyr does not seem to depend on $K_d$ of the CM-DS complex and the relative enzyme concentrations. This might indicate, that these amino acids do not affect the assembly of DS tetramer, but only of the CM-DS complex (as seen from the SEC experiments). If Phe and Tyr would affect the DS tetramer one could expect that, similarly as for Trp activation, their effects at low protein concentrations would be stronger. However, this was not observed, suggesting that changes in the DS quaternary state might not be the factor governing the disassembly of the CM-DS complex by Phe and Tyr.

The assumptions that Trp binding affects DS tetramer formation, while Phe and Tyr only affect the CM-DS complex, is further supported by the location where the individual effectors bind. The crystal structure of MtDS (Figure 1.3E and 2.3) reveals that Trp binds at the tetramer interface. Therefore, one could infer that conformational changes in this region can influence tetramer formation. On the other hand, Phe and Tyr bind at the dimer interface, far from the DS tetramer interface. It is thus conceivable, that binding of these effectors does not have a significant effect on the DS tetramer. However, Phe and Tyr binding could induce changes at the CM-DS interface and foster its dissociation. In fact, a slight rotation by 1-2° was observed in the MtDS structure with Phe and Tyr bound [28], indicating that even a small change might be enough to make the DS surface unsuitable for productive CM binding.

An important aspect of this study was to test the biologically most relevant conditions that would reflect the regulatory patterns in vivo. The intracellular DAHP synthase concentration was shown to be in the range of 3 µM for both M. tuberculosis H37Rv and the closely related
Mycobacterium bovis BCG,[146] as well as for B. flavum,[147] whereas a CM concentration of 300 nM was reliably determined only in M. bovis.[146] Also, the intracellular concentration of chorismate, which has been shown to stabilize both protein complexes,[46,48,28] was estimated to be in the range of 40 - 70 μM in B. flavum.[147, 148] Finally, the aromatic amino acid concentrations were chosen to reflect typical intracellular concentrations in bacteria.[148, 142] Under the assumption that the assay conditions used in vitro (see legend of Figure 2.1) resemble the biologically relevant substrate and effector concentrations, it is reasonable to speculate that activation by Trp will occur only in C. glutamicum ATCC 13032 but not in M. tuberculosis H37Rv, since the enzyme concentrations in these organisms would be below and above the $K_{d,\text{app}}$ of the respective CM-DS complex (Figure 2.4).

Finally, as seen in Table 2.1, chorismate affinity reflected by the $K_m$ values of the activated CMs is also significantly different for the two protein complexes. While the $K_m$ of MtCM is strongly decreased upon complex formation, the $K_m$ of CgCM remains high. This situation might reflect a preferential synthesis of Trp in C. glutamicum. Since the $K_m$ for chorismate of anthranilate synthase (48 μM[149]) is much lower than the $K_m$ of CgCM-CgDS (> 2 mM) the metabolic flow should favor Trp biosynthesis at 70 μM intracellular chorismate.[148] As soon as the Trp pool is full, Trp feedback inhibits anthranilate synthase and simultaneously activates CgCM in the complex with CgDS effectively shifting the metabolic flow towards the Tyr and Phe branch (Figure 2.4). The anabolic pathway to the aromatic amino acids in C. glutamicum is, therefore, a perfect example of how enzymes exploit differential $K_m$ parameter to intrinsically adjust the production according to the needs of the organism.

Taken together, our data demonstrate how mechanistically intriguing and functionally diverse the two essential shikimate pathway enzymes can regulate the metabolic flow by exploiting their transient interaction and inherent substrate affinity. It is clear that subtle changes of this interaction result in a profoundly altered regulation of the synthesis of aromatic amino acids and other essential compounds. Therefore, modulation of these enzymatic complexes could be exploited for increased productivity of amino acid producing C. glutamicum as well as weakening the pathogen M. tuberculosis.
Chapter 2: Inter-enzyme allosteric regulation of AroQδ chorismate mutases

Figure 2.3 Proposed regulation of aromatic amino acid biosynthesis in (A) C. glutamicum ATCC 13032 and (B) M. tuberculosis H37Rv. The C. glutamicum Type I DAHP synthase (aroF) is inhibited by Tyr.\[^{150, 127}\] The Type II DAHP synthase (CgDS) is inhibited by Tyr and Phe, whereas Trp has a weak activating effect.\[^{132}\] Interaction with CgCM weakly activates CgDS as well as subjects it to a weak inhibition by chorismate and prephenate.\[^{133}\] Complex formation with CgDS activates CgCM. The complex is stabilized by chorismate\[^{46}\] and inhibited by Tyr and Phe (this work, Figure 2.1). Trp strongly activates CM activity of the CgCM-CgDS complex and shifts the metabolic flow towards the Phe/Tyr branch.

The Type II DAHP synthase of M. tuberculosis (MtDS) is inhibited by the simultaneous presence of Trp, Tyr, and Phe (this work, Figure 2.1). Other genes encoding DAHP synthases have not been identified in this strain. MtCM is activated upon interaction with MtDS and the complex is stabilized by chorismate.\[^{48, 28}\] Simultaneous presence of Tyr and Phe inhibits complex formation whereas Trp has no influence under physiological concentrations of MtCM and MtDS.\[^{28}\] Weak activation of MtDS by MtCM has also been reported.\[^{136}\]
2.4 Materials and Methods

2.4.1 Materials and general procedures

Chorismate was purified as described before.[151] All other chemicals were purchased from Sigma-Aldrich/Fluka.

2.4.2 Protein production and purification

*E. coli* strain KA13 (genotype: Δ(srl-recA)306::Tn10 Δ(pheA-tyrA-aroF) endA1 hsdR17(rK-, mK+) λ(DE3) [P*lacUV5*-expressed T7 RNA pol gene]; TetR)[152, 153] used for protein overproduction and plasmid pKTDS-HN[48] have been described before. Native MtCM was produced and purified by Severin Sasso or in this study according to a published protocol.[154] Native CgCM and His-tagged CgDS were produced and purified by Helen Vikdal Thorbjørnsrud at the University of Oslo.

The His-tagged MtDS was produced following the previously established protocol [26] with minor modifications. A single colony of KA13/pKTDS-HN was inoculated into 5 mL LB-Amp (150 µg/mL of the ampicillin sodium salt) and grown overnight at 30°C. The resulting pre-culture was inoculated in M9c minimal medium consisting of M9 salts[155] (6 mg/ml Na₂HPO₄, 3 mg/ml KH₂PO₄, 1 mg/ml NH₄Cl, and 0.5 mg/ml NaCl) containing US* trace elements,[156] 1 mM MgSO₄, 0.1 mM CaCl₂, 0.2 % (w/v) D-(+)-glucose, 5 µg/ml thiamine-HCl, 5 µg/ml 4-hydroxybenzoic acid, 5 µg/ml 4-aminobenzoic acid, 1.6 µg/ml 2,3-dihydroxybenzoic acid, 10 µg/ml Trp, 10 µg/ml Tyr, 10 µg/ml Phe, and 150 µg/ml sodium ampicillin. The cultures were grown at 30°C until OD₆₀₀ of 0.3-0.5 and protein production was induced with 0.1 mM salicylate. After incubation at 30°C for 16 h the cells were harvested and resuspended in BTP++ buffer consisting of 20 mM 1,3-bis(tris(hydroxymethyl)methylamino)propane (BTP) pH 7.5, 1 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 0.2 mM phosphoenolpyruvate (PEP), 0.1 mM MnCl₂, and 150 mM NaCl[140, 48].

The cell pellets were treated with 1 mg/mL lysozyme for 30 min on ice and ruptured by sonication. The insoluble cell debris was removed by centrifugation for 20 min at 13,000 rpm (Sorvall rotor SA600) at 4 °C and if there was a significant amount of cell pellet left, the sonication was repeated. The crude soluble cell extract was subjected to Ni-NTA affinity chromatography (Qiagen AG). The bound
protein was first washed with BTP++ containing 40 mM imidazole and then with BTP++, containing 40 mM imidazole and 2 M NaCl. MtDS was eluted in BTP++ containing 250 mM imidazole and dialyzed overnight in Buffer A (20 mM BTP, pH 7.5 containing 0.1 mM MnCl₂, 0.2 mM PEP, 1 mM TCEP). The sample was further purified by FPLC on a MonoQ HR 10/10 column with Buffer A as the running buffer eluting with linear gradient of Buffer B (Buffer A, containing 500 mM NaCl). For storage a protease inhibitor cocktail without EDTA (catalogue No. P-8849; Sigma–Aldrich) was added. On average protein yield was around 1 – 5 mg/L cell culture. The electrophoretic homogeneity of the protein preparations was assessed by SDS–PAGE using the PhastSystem (GE Healthcare).

2.4.3 Enzymatic assays

The enzymatic assays were performed as described before.[48] The disappearance of the substrate chorismate was followed at 310 nm ($\varepsilon_{310} = 370$ M⁻¹·cm⁻¹) or 274 nm ($\varepsilon_{274} = 2630$ M⁻¹·cm⁻¹).

Whereas the activity of CgCM alone was measured in 50 mM BTP, pH 7.5, the activity of CgCM in complex with CgDS was measured in BTP+ (50 mM BTP, pH 7.5, additionally containing 1 mM TCEP, 0.2 mM PEP, and 0.1 mM MnCl₂).

To determine the catalytic parameters of CgCM and the CgCM-CgDS complex the CM activity was measured by varying chorismate concentrations from 0.5 mM to 2.5 mM, and the data were fitted to the Michaelis-Menten equation$^{[157]}$ using the program KaleidaGraph (Synergy Software).

The effects of aromatic amino acids on the CM activity of MtCM-MtDS and CgCM-CgDS were measured in BTP+ buffer in the presence or absence of 25 μM of each aromatic amino acid. The MtCM-MtDS assay contained 23 μM chorismate, 30 nM MtCM, and 300 nM MtDS, or 25 nM MtCM, and 25 nM MtDS. The assay for CgCM-CgDS contained 100 μM chorismate, 100 nM CgCM, and 1000 nM CgDS.

Chorismate was produced following a published protocol$^{[158]}$ and protein concentrations were determined using the Bradford assay, which was previously calibrated for absolute MtCM and MtDS concentrations using the results from quantitative amino acid
For determination of CgCM and CgDS concentrations, BSA was used as a standard.

The correct molecular mass of the purified proteins was confirmed by electrospray ionization mass spectrometry (ESI) at the Mass Spectrometry (MS) Service at the Laboratory of Organic Chemistry, ETH Zurich.

### 2.4.4 Size-exclusion chromatography

Analytical SEC was performed using a Yarra™ 3u SEC-3000, 300 x 4.6 mm (Phenomenex) HPLC column. The buffer consisting of 20 mM BTP, pH 7.5, 150 mM NaCl, 1 mM TCEP, 0.2 mM phosphoenolpyruvate, 0.1 mM MnCl₂, and 100 μM chorismate was used as the mobile phase. If necessary, 100 μM Tyr, 100 μM Phe, or 100 μM Trp was added. For column loading, 40-μL samples containing 10 μM MtCM and 15 μM MtDS were injected and run for 20 min at a flow rate of 0.35 mL/min.

Successive fractions of 550 μL were collected and the proteins were concentrated for analysis by precipitation upon adding an equal volume of 20 % TCA. After incubating for 1 h at room temperature the proteins were pelleted by centrifugation for 20 min at 12,000 rpm. The pellets were washed twice with acetone and dissolved in 15 μL SDS loading buffer (50 mM Tris-HCl, pH 6.8, 0.1 M dithiothreitol, 2 % sodiumdodecylsulfate, 0.1 % Bromophenol Blue, and 10 % glycerol). Samples were analyzed by 20 % SDS-PAGE using the PhastSystem (GE Healthcare). Elution volumes (fraction number) and the presence of individual protein species turned out to correlate very reproducibly over several repetitions of the same experiment.
3 IN SEARCH FOR SMALL MOLECULE INHIBITORS OF THE CHORISMATE MUTASES FROM M. TUBERCULOSIS

3.1 Introduction

TB remains one of the deadliest diseases today (see section 1.2.1). The ever-growing threat of MDR-TB strains and the fact that one third of the world’s population is latently infected with M. tuberculosis urge humanity to develop novel antibiotics.

The enzymes from the shikimate pathway are generally considered to be good targets for drug development due to their absence in mammalian cells. Moreover, the MtCM-MtDS complex presents a unique system that is not only essential for the biosynthesis of Tyr and Phe, but also is only found in a few taxonomic orders of Actinobacteria, making any potential drug against it safe for humans and most gut microbiota. In addition, M. tuberculosis features a second, highly active secreted CM (*MtCM), proposed to play a role in virulence of M. tuberculosis (see section 1.2.2). While inhibition of the intracellular MtCM or disruption of the MtCM-MtDS complex would preclude biosynthesis of essential aromatic amino acids for the producer organism, inhibitors of *MtCM would be useful both for
functional studies and potential interference with the host-pathogen interaction.

Some efforts have been made to identify inhibitors for *MtCM*[159, 160]; however, there have been no successful discoveries reported neither for *MtCM* nor for MtCM-MtDS so far. To date, Bartlett’s transition state analog (TSA, see Figure 1.4)[33] remains the best-known inhibitor for most CMs. However, it cannot be considered a good drug candidate due to its expensive synthesis and hydrophilic properties, precluding efficient cell penetration.

**Figure 3.1 Scheme of the direct and coupled CM assays.** During the direct UV absorbance CM assay the conversion of chorismate to prephenate is monitored by measuring the absorbance decrease at 274 nm or 310 nm. The coupled CM assay also involves the subsequent step in the shikimate pathway where prephenate is converted into 4-hydroxyphenylpyruvate. The concomitant formation of NADH is monitored by measuring the fluorescence at 460 nm upon excitation at 340 nm. *MtCM* secreted CM of *M. tuberculosis*; MtCM-MtDS non-covalent complex between the intracellular MtCM and the MtDS from *M. tuberculosis*; EhPDH *E. herbicola* prephenate dehydrogenase.

In order to identify novel inhibitors for both MtCM-MtDS and *MtCM* an HTS of a 32,032 compound library was initiated in collaboration with Dr. J. P. von Kries at the Screening Unit of the Leibniz-Institute of Molecular Pharmacology in Berlin. The core of this library was designed for a broad range of targets and consisted of a core of 16,671 bioactive compounds identified using the World Drug Index.[161] These library members fulfilled the following the
requirements: (i) enrichment of putative bioactive structures, (ii) large chemical diversity, (iii) clusters of related compounds enabling reliability of hits, (iv) exclusion of reactive and unstable compounds, and (v) commercial availability. In addition to the 16,671 compounds, the library contained another set of compounds that had been derived as mentioned before, but with an emphasis on solubility. Furthermore, it included a set that contained carboxylate, ketone, and amine fragments. Also, compounds from pharmaceutically active compound libraries from LOPAC (Sigma-Aldrich) and Art-Chem (Berlin, Germany) as well as donations from research laboratories were included. For more instrumentation on the FMP compound library collection, see: http://www.leibniz-fmp.de/core-facilities/screening-unit/screening-unit/downloads.html

To facilitate the HTS, a suitable CM assay had to be developed.\textsuperscript{[162]} The continuous CM assay\textsuperscript{[36]} monitoring the depletion of the substrate chorismate at 274 nm or 310 nm (Figure 3.1) cannot be used in the HTS setting due to interference by strong UV absorption of many of the small molecule compounds and the need for costly UV transparent microtiter plates. The new coupled CM assay included the engineered monofunctional PDH from \textit{Erwinia herbicola} (EhPDH) as a coupling enzyme (Figure 3.1). EhPDH converts the CM reaction product prephenate to 4-hydroxyphenylpyruvate by reducing the non-fluorescent NAD\textsuperscript{+} to NADH. The formation of NADH provides a real-time fluorescence readout at 480 nm circumventing the issues associated with UV absorbance of the individual library compounds.

However, the coupled CM assay is also prone to artifacts giving false positives including (i) abnormal signal due to auto-fluorescent compounds, (ii) quenching of the fluorescence due to strong light absorbance, and (iii) inhibition of the coupling enzyme by the compounds. To eliminate the two latter sources of false positives, a prephenate counter-screen was developed.\textsuperscript{[26, 160]} It included addition of a defined amount of prephenate after the HTS reaction was completed and determination of NADH fluorescence before and after the addition. Any samples that deviated from a threshold value around the fluorescence change of the control without inhibitor were excluded from further analysis.
Figure 3.2 Identified compounds for *MtCM. HTS hit compounds identified previously for *MtCM using the coupled CM assay[26, 160]. IC<sub>50</sub> defines compound concentration at which half maximal inhibition is achieved.

The coupled CM assay and the prephenate counter-screen were successfully implemented to screen the FMP compound library for *MtCM.[26, 160] The 348 best inhibitors were validated using various inhibitor concentrations to determine the IC<sub>50</sub> values. Several compounds that then exhibited inhibitory effects are shown in Figure 3.2.

In this chapter, we describe adaptation of the coupled assay, the primary HTS, and the validation of the best inhibitors from the FMP compound library for the MtCM-MtDS complex. Furthermore, the direct CM assay was miniaturized and employed it to confirm the inhibition of the best candidate compounds identified in the HTSs of both the *MtCM and the MtCM-MtDS complex.
3.2 RESULTS

3.2.1 Optimization of the coupled MtCM-MtDS assay for the HTS format

To transfer the coupled CM assay for MtCM-MtDS to an HTS format, suitable concentrations needed to be established for (i) the substrate chorismate, (ii) the target enzymes MtCM and MtDS, (iii) the coupling enzyme EhPDH, and (iv) the stabilizing agents BSA/Tween 20. We aimed for a chorismate concentration sufficiently above the $K_m$ to ensure constant substrate saturation and appropriate linearity of the reaction over the required measurement period. The concentrations of the target proteins MtCM and MtDS were chosen to be close to $K_{d,app}$ of the complex to provide conditions where the complex is just formed, in the hope of also finding complex-disrupting compounds. Varying the MtDS concentrations between 70 and 280 nM in the presence of 15 nM MtCM showed that the maximum initial velocity is achieved as soon as $>100$ nM MtDS is added. Therefore, the chosen 15 nM and 140 nM of MtCM and MtDS, respectively, were assumed to be sufficient for transient complex formation. The coupling enzyme EhPDH and its cofactor NAD$^+$ were provided in excess (550 nM and 500 μM, respectively) to ensure that the PDH reaction was not the rate-limiting step. The addition of 0.1 mg/mL BSA and 0.005 % Tween 20 allowed for more reproducible results by preventing enzyme adhesion to the walls in the microtiter plate.

$$Z' = 1 - \frac{(3\sigma_{c+} + 3\sigma_{c-})}{|\mu_{c+} - \mu_{c-}|}$$

**Equation 3.1 Z’-factor or screening window coefficient.** A parameter for evaluation of the assay quality, judged by the signal dynamic range and the data variation. In the case of $1 > Z’ > 0.5$ the separation band or “signal window” is large indicating excellent assay conditions.$^{[163]} \sigma_{c+}/\sigma_{c-}$ are the standard deviations of the positive/negative control, $\mu_{c+}/\mu_{c-}$ are the means of the positive/negative controls.
The newly developed assay was evaluated for a 384-well plate format using the sample volumes (20 μL) suitable for the Tecan Freedom Evo 200 liquid handling platform (Tecan AG) at the FMP Berlin. To determine the validity and robustness of the assay, the Z'-factor,\(^\text{[163]}\) which allows evaluation of the assay window (Equation 3.1), was determined based on 192 positive (with MtCM) and 192 negative (without MtCM) controls. The Z'-factor was determined to be 0.67 for our MtCM-MtDS screen, exceeding the threshold value of 0.5. This confirmed the suitability of the assay for identification of reliable hit compounds using the desired HTS setting.\(^\text{[163]}\)

### 3.2.2 The HTS campaign and hit validation for the MtCM-MtDS complex using the coupled assay

The primary screen of the FMP compound library was performed using a compound concentration of 10 μM per assay. The NADH fluorescence increase was monitored over 5.5 minutes and the initial velocity of the reaction was determined based on ten data points. Every 384-well plate included 352 test compounds plus 16 positive and 16 negative controls to evaluate the inhibitory effects of the compounds and to simultaneously monitor the assay performance by determining the Z'-factor (Equation 3.1) for each individual plate. The samples that showed strong auto-fluorescence causing detector saturation or those failing the prephenate counter-screen by exhibiting a fluorescence difference outside of the range of 0.5 to 1.5-fold the positive control value were excluded from further analysis.

The screen identified 31,587 samples that showed valid numeric values. These samples were evaluated employing the Z-score method\(^\text{[164]}\) that allows identification of the outliers among the measured samples. The method assumes that the effects of all the investigated compounds have a normal distribution around the mean value and that the hits behave very different from the majority of the compounds (are the “outliers”) (Equation 3.2).

After the primary screen, 1,034 samples that showed a significant decrease in the initial velocity and exhibited Z-score below -4 were considered as hits. However, out of these samples, the ones that showed an initial velocity ≥ 20 % lower than the negative control were discarded, since they probably exhibited photobleaching effects. Out of
Chapter 3: In search for small molecule inhibitors of the chorismate mutases from M. tuberculosis

\[ Z = \frac{(x_i - \bar{x})}{s_x} \]

**Equation 3.2 Z-score.** The Z-score provides information on the strength of each measurement relative to the distribution of the other measurements on a particular plate. Here, \( x_i \) is the measurement of the \( i^{th} \) compound, \( \bar{x} \) is the mean, and \( s_x \) is the standard deviation of all measurements within the plate.\[^{164}\]

The remaining 851 compounds the best 348 were selected for further validation.

The validation screen for the aforementioned 348 compounds was performed by varying the concentration of the test compounds from 50 \( \mu \text{M} \) to 0.195 \( \mu \text{M} \). Subsequently, dose-response curves were generated to obtain the degree of inhibition, IC\(_{50}\), and the Hill coefficient. In parallel, the dose response for the TSA\[^{33}\] (Figure 1.4) inhibitor was tested in quadruplicate to validate the assay. The average IC\(_{50}\) for TSA obtained from four independent measurements was 2.03 ± 4.8 \( \mu \text{M} \). This was comparable with IC\(_{50}\) of 9.84 ± 2.76 \( \mu \text{M} \) (the calculated corresponding \( K_i \) is 7.4 ± 2.1 \( \mu \text{M} \)) obtained for *MtCM using identical HTS setting,\[^{26,160}\] which correlated well with the \( K_i \) value of 3.7 \( \mu \text{M} \) obtained with an absorption assay for *MtCM.\[^{36}\]

The validation of hit compounds for the MtCM-MtDS complex yielded a reliable data set for 222 compounds. The ones that exhibited an IC\(_{50}\) in the low micromolar range, a Hill coefficient ranging between -0.5 and -1.5, a rare appearance in other screens, as well as little or no fluorescence increase associated with compound concentration were considered the best candidates (Figure 3.3). Among those, the most promising hits found in the validation screen (the first two compounds in Figure 3.3) distantly resemble TSA. However, the fact that the two compounds are enantiomers, but inhibit the CM reaction with the same IC\(_{50}\) makes them doubtful candidates. The other hit molecules were mostly aromatic compounds often having ester, nitro or sulfoxide-containing groups.
Regulation, effectors, and evolvability of chorismate mutases from *Corynebacteriales*

---

**Figure 3.3 Identified compounds for the MtCM-MtDS complex.** HTS hit compounds identified in this work for the MtCM-MtDS complex using the coupled CM assay. IC$_{50}$ defines compound concentration at which half maximal inhibition is achieved.

Finally, 49 compounds were identified in the MtCM-MtDS screen that also showed inhibitory effects in the *MtCM* screen. This was surprising since the two proteins have differently arranged active sites and a significantly distinct overall structure. Therefore, these compounds were most probably just fluorescence quenchers or might have inhibited the coupling enzyme EhPDH.

### 3.2.3 Validation of the hit compounds for MtCM-MtDS and *MtCM using the direct assay*

The coupled CM assay proved to be a robust tool for identification of potential inhibitors using the automated setting based on the fact that TSA was identified routinely. However, it did not exclude the possibility of finding false positives due to inhibition of the coupling enzyme EhPDH. Therefore, a direct CM assay (Figure 3.1) was implemented to validate the hit compounds obtained from the primary HTSs for both MtCM-MtDS complex and *MtCM.*
The existing direct UV absorption assay in cuvettes for the MtCM-MtDS complex and *MtCM was adapted to the 384-well plate format to allow following linear progress of the reaction over the first 5-10 min for several samples simultaneously. Enzyme concentrations were 10 nM and 140 nM for MtCM and MtDS, respectively, and 2.5 nM for *MtCM. The determined signal to background ratio was approximately 10:1 for both the MtCM-MtDS and the *MtCM assays.

All 348 compounds identified after the primary HTS screen for MtCM-MtDS were tested at a single concentration of 12.5 μM. The 37 compounds that showed 50 % slower initial velocities when compared to the uninhibited control were subjected to the validation screen using concentrations ranging from 0 to 25 μM.

Similarly, the 348 compounds identified after the previously conducted primary HTS screen for *MtCM[26,160] were subjected to the miniaturized direct UV assay at a single concentration of 25 μM. 50 compounds that still showed inhibitory properties were subsequently tested by the concentration-dependent direct UV assay using concentrations ranging from 0 to 25 μM.

As a control inhibitor, TSA was tested in parallel at concentrations ranging from 0 to 25 μM. It showed a clear concentration-dependent inhibition pattern for both MtCM-MtDS and *MtCM (Figure 3.4). However, none of the other tested compounds showed a similar data set, including the ones that were identified as the best inhibitors in the automated coupled assay (Figures 3.2 and 3.3).

Nevertheless, we obtained five compounds from the *MtCM and one from the MtCM-MtCM screen that showed some inhibition in the microtiter plate assay and evaluated them in a more accurate cuvette assay. It was observed that precipitation of the compounds became a severe issue at concentrations above 10 μM precluding a reliable UV readout. When the compounds were measured below 10 μM, they did not show any sign of inhibition neither for *MtCM nor for the MtCM-MtDS complex.
Regulation, effectors, and evolvability of chorismate mutases from *Corynebacteriales*

3.3 Discussion

We have tested the compound library provided by the Leibniz Institute of Molecular Pharmacology for the MtCM-MtDS complex employing the previously developed coupled CM assay. The assay had to be optimized, miniaturized and adapted for the robotic HTS setup to give linear fluorescence readout over the first 10 min of the reaction. The screen resulted in a number of compounds that showed a reliable concentration-dependent inhibition, with the best inhibitors exhibiting an apparent IC$_{50}$ between 0.4 and 1 μM. A similar screen performed previously for *MtCM* identified inhibitors with IC$_{50}$ values within the same range.

However, since the coupled CM assay involved the additional enzyme EhPDH, the possibility that any identified compound would act instead on the coupling enzyme could not be ruled out. Therefore, we tested the 348 best inhibitors from the primary screens for both *MtCM* and MtCM-MtDS using a newly established miniaturized direct UV absorbance assay. Even though the established inhibitor TSA was readily singled out, we could not identify any new inhibitors. The measurements with the library compounds were prone to artifacts such as too high absorbance and precipitation.
When the most promising candidates were tested at larger volumes in the more reliable cuvette assay, it became clear that the compounds precipitate heavily at concentrations above 10 μM. The precipitation caused an apparent absorbance increase compensating the absorption decrease expected for the uninhibited CM reaction. Therefore, the measurements in the microtiter plate using a higher than 10 μM compound concentration might have appeared as false positives with reduced initial velocities. Moreover, many of these compounds exhibited a strong absorbance at 274 nm or photobleaching effects causing a perturbed signal or a faster decrease in absorbance, respectively. These measurements were difficult to interpret and were usually discarded. It is, therefore, possible that true inhibitors were lost due to their absorption properties artifacts and testing these compounds at a concentration lower than 10 μM could be necessary.

On the other hand, judging from the chemical structures of the inhibitors (Figures 3.2 and 3.3), none of them are likely to bind to the small and highly charged active site of a CM. Therefore, it could well be that the compounds identified using the coupled assay are all inhibitors of EhPDH, photobleachers or cause other artifacts. If the former is true, one could consider testing the same library of compounds for inhibitors of PDH from *M. tuberculosis* since this enzyme is also an essential component in the biosynthesis of Tyr.

The composition of the starting compound library is crucial and needs to be revised for any given target. In our case, the structure and catalytic mechanism of both target CMs are well investigated and one could take advantage of this instrumentation when selecting a compound library. For example, libraries enriched in compounds that are small and possess various negatively charged groups could yield better hits in this case. In fact, prof U. Krengel group at the University of Oslo have identified isocitric acid derivatives as potential inhibitors of *MtCM* using virtual screening of the ZINC database.[77] The crystal structure of (1S,2S)-1-hydroxypropane-1,2,3-tricarboxylic acid-bound *MtCM* showed that the carboxylic acid groups bind to the active site residues in a similar manner as the ones of TSA.[160] This suggests, that
compound libraries that contain carboxyl moieties could be better suited for finding inhibitors directed to the CM active site.

Finally, targeting the MtCM-MtDS complex assembly poses a particular challenge since small molecules are usually considered to be less potent for disrupting PPIs. Larger molecules presenting more atoms for contacts are better suited to interact with large protein surfaces, and peptides have recently gained more attention as potential candidates. In the next chapter of this thesis we devise a strategy that could allow testing constrained peptides for disrupting the MtCM-MtDS complex.

### 3.4 Materials and methods

#### 3.4.1 Materials and general procedures

Chorismate was purified as described before. The compounds for HTS and the subsequent in-house UV-based validation experiment were provided by the Screening Unit at the Leibniz-Institute of Molecular Pharmacology in Berlin (FMP) with at least 90 % purity. The selected compounds that showed the strongest inhibitory effects were purchased from either the Vitas-M laboratory (Apeldoorn, Netherlands) MolPort (Riga, Latvia). The purity was determined by the provider and was at least 90 %. Bartlett's transition state analog was prepared by R. Pulido. All other chemicals needed for the assays were purchased from Sigma-Aldrich/Fluka.

#### 3.4.2 Protein production and purification

EhPDH was purified by Kathrin Roderer and MtCM was purified by Severin Sasso as described before. MtDS was produced and purified as described in section 2.4.1 and *MtCM was purified following the previously published protocol.

#### 3.4.3 HTS for MtCM-MtDS

Initial tests needed for establishing the suitable assay conditions were performed in a 96-well plate by monitoring the formation of fluorescent NADH ($\lambda_{ex} = 340$ nm $\lambda_{em} = 460$ nm) with the plate reader Spectramax M2 (Molecular Devices).
Chapter 3: In search for small molecule inhibitors of the chorismate mutases from M. tuberculosis

HTS was performed on the Tecan Freedom Evo 200 liquid handling platform at the FMP, Berlin. The final concentrations of the assay components were 0.1 mg/mL BSA, 0.005 % Tween 20, 500 μM NAD\(^+\), 550 nM EhPDH, 100 μM chorismate, 140 nM MtDS, and 15 nM MtCM in BTP + buffer (50 mM BTP, pH 7.5, containing 0.5 mM TCEP, 0.2 mM PEP, and 0.1 mM MnCl\(_2\)).

Firstly, 0.2 μL of compound solution in DMSO was added to 15 μL of the reaction mix, containing all the components except for MtCM and MtDS. After a short mixing the reaction was initiated by adding 5 μL of the enzyme mix. The fluorescence of NADH (excitation and emission wavelengths as described before) was followed on a Safire II plate reader (Tecan) at 30 °C. The data was analyzed using the Accelrys® software package following the described routines\(^{[163, 164]}\).

### 3.4.4 Direct UV absorbance assay

The validation experiment of the inhibitors identified after HTS was performed in UV-transparent 384-well plates by monitoring chorismate depletion at 274 nm.

The assay conditions for MtCM-MtDS were BTP + buffer, pH 7.5, containing 0.1 mg/mL BSA, 0.005 % Tween 20, 100 μM chorismate, 140 nM MtDS and 10 nM MtCM.

The assay condition for *MtCM were 50 mM sodium phosphate buffer, pH 7.5 containing 0.1 mg/mL BSA, 0.005 % Tween 20, 100 μM chorismate, and 2.5 nM *MtCM.

20 μL of the enzyme mix (containing either *MtCM or MtCM-MtDS) was pipetted into each well and 0.4 μL of compound solution in DMSO was added. This mixture was incubated at 30 °C in the plate reader for 10-15 min, and the reaction was initiated by addition of 20 μL pre-incubated reaction mix. The final assay mixture was pipetted once up and down to ensure proper mixing and UV absorbance was monitored at 274 nm on a Varioskan (Thermo Scientific) plate reader at 30 °C.

The enzymatic assays in cuvette were performed as described before.\(^{[48]}\)
4 In vivo selection for cyclic peptide modulators of the MtCM-MtDS complex

4.1 Introduction

As outlined in the section 1.2.3, peptides have gathered attention as promising PPI inhibitors. However, fast degradation, poor membrane permeability, and not sufficient target binding are the main issues while developing peptides as potential drugs.

These shortcomings can be circumvented by introducing conformational constraints into a peptide chain. For example, conformational rigidity achieved via cyclization was shown to improve the target binding by decreasing the entropy term of the Gibbs free energy.\textsuperscript{[168]} Moreover, cyclic peptides (CPs) are resistant to exopeptidases and endopeptidases since they lack free N- and C-termini and exhibit less flexibility. Certain CPs exhibit better membrane permeability than their linear counterparts due to specific interactions within the peptide that prevent exposure of their hydrophilic groups.\textsuperscript{[169]}

Due to the promising features of peptides as drug candidates, recent progress has been made in developing efficient screening and
selection systems to improve the size and the diversity of the peptide libraries. For example, combinatorial chemistry can generate libraries of $10^8$ unique compounds,\cite{170} while phage or mRNA displays can expand this number to $10^{13}$,\cite{171}

The *in vivo*-based methods have also gained significant attention, due to possibility of simultaneous testing against other cellular components and thus elimination of CPs that affect the off-targets or are toxic to the living cells. These methods allow testing large libraries of $10^6$-$10^8$ variants, since they are only limited by transformation efficiency and the ability to form a cycle *in vivo*.

One way to create CPs is employing the split intein-catalyzed ligation of proteins and peptides system (SICLOPPS). It utilizes a circularly permuted intein, which catalyzes the cyclization of a peptide inserted between the split intein fragments (Figure 4.1),\cite{172} The cyclization of the target peptide or protein occurs spontaneously in the cell, does not require a specific peptide sequence and the recent innovations allow incorporation of the unnatural amino acids.\cite{173} By utilizing SICLOPPS, a wide range of CPs can efficiently be produced within the cell; therefore, this technology has been successfully applied to identify inhibitors of proteases, methyl transferases and other proteins including PPIs.\cite{174}

In this chapter, we describe the development of an *in vivo* selection system for CP modulators of the MtCM-MtDS complex using SICLOPPS and the directed evolution approach outlined in section 1.3. We expect that appropriate selection conditions will allow identification of CPs that positively or negatively affect the CM activity *in vivo*.

In the long run we aim to retrieve potential CP activators and inhibitors of the MtCM-MtDS complex that can be valuable drug candidates. MtCM-MtDS inhibitors would prevent the biosynthesis of essential aromatic amino acids, while activators would deplete the chorismate pool necessary for other aromatic compounds thus weakening the *M. tuberculosis* pathogen.
Figure 4.1 The split intein-catalyzed ligation of proteins and peptides system (SICLOPPS). The SICLOPPS construct[^175] has the C-terminal (I_c) and N-terminal (I_N) fragments of the intein at the N- and C- termini of the target peptide, respectively. The expressed fusion protein folds to form an active intein. An N-to-S acyl shift at the N-terminal intein produces a thioester, which undergoes trans-thioesterification with the side-chain nucleophile (X; representing Cys or Ser) at the C-terminal intein moiety. The action of an Asn side chain of the intein and an S-to-N acyl shift generates the cyclic product.
4.2 Results

4.2.1 Genetic selection system for CP modulators of the MtCM-MtDS complex

For identification of CP modulators of the MtCM-MtDS complex, an in vivo selection system was established based on the previously described CM-deficient E. coli strain KA12.[51] Two sets of plasmids shown in Figure 4.2A were newly constructed. The plasmids pKIMP-ACGQ and pKIMP-ACGQ-RBS6 carried aroQδ and aroG genes encoding the target proteins MtCM and MtDS, respectively. The plasmids pKINT and pCyP1 carried the SICLOPPS construct for CP libraries.

4.2.1.1 Construction of the plasmids pKIMP-ACGQ and pKIMP-ACGQ-RBS6

The plasmid pKIMP-ACGQ was based on the previously described plasmid pKIMP-ACG.[26] It contained the moderate-copy (p15A) origin of replication, the chloramphenicol resistance gene (cat), and the genes tyrA* and pheC for the CM complementation system. The genes aroQδ and aroG were cloned under inducible Psal and Ptet promoters, respectively, allowing independent induction of the target genes. Thus, the selection stringency could be adjusted by variation of intracellular MtCM concentration and by providing a sufficient amount of interaction partner.

The KA12/pKIMP-ACGQ transformant could complement the cell growth on minimal medium without the aroQδ and aroG inducers salicylate (Sal) and tetracycline (Tet) (Table 4.1). Basal activity of Psal and Ptet promoters indicated that the selection system was not stringent enough for subsequent selection experiments.

The selection stringency was intended to be increased by leveling down the intracellular MtCM concentration. For that, the ribosomal binding site (RBS) modification approach was pursued to impair its translation efficiency (Figure 4.2B). A gene library was created where the –21 to –4 region upstream of the start codon of the aroQδ gene was randomized by using the NNK codon pattern. A small fraction of 4×10⁶ independent members in KA12 was plated on non-selective M9c+FY (M9c supplemented with Tyr and Phe) plates. The remaining library was subjected to selection on minimal M9c+Tet500 ng/mL+ Sal0.5 mM (M9c medium with 0.5 mM Sal and 500 ng/mL Tet) plates. These moderately
Figure 4.2 Genetic selection system for CP modulators of the MtCM-MtDS complex. (A) The plasmids used for genetic selection system based on the CM-deficient *E. coli* strain KA12. The plasmids pKIMP-ACGQ or pKIMP-ACGQ-RBS6 carry the independently regulated genes of the target proteins MtCM (encoded by *aroQ₈* gene and controlled by *P₎* promoter) and MtDS (encoded by *aroG* gene and controlled by *P₄* promoter). The plasmids pKINT or pCyP1 carry SICLOPPS constructs controlled by either *P₇₇* or *P₆₆* promoters, respectively. (B) The RBS regions (boxed) of the original *P₅* promoter on pKIMP-ACGQ plasmid and the mutagenized RBS on pKIMP-ACGQ-RBS6 are aligned. The original RBS has the TCACGAG (shown in blue) nucleotide sequence six nucleotides upstream of the *aroQ₈* start codon ATG. The newly generated purine-rich sequence GGGCC and the two rare start codons on the plasmid pKIMP-ACGQ-RBS6 are shown in blue and red, respectively.
selective conditions allowed retrieving the clones that have a weak, but functional \textit{aroQ\textsubscript{8}} expression. After 96 hours of incubation on M9c+Tet\textsuperscript{500} ng/mL + Sal\textsuperscript{0.5} mM, 3 \% complementing clones could be selected.

102 clones from non-selective M9c+Fy and 306 from selective M9c+Tet\textsuperscript{500} ng/mL + Sal\textsuperscript{0.5} mM plates were picked and grown under more stringent conditions when the Sal concentration was varied between 0.1 mM and 0.005 mM in the presence of 200 ng/mL Tet as well as the M9c plates without inducers.

Only 4 \% of the 102 clones picked from non-selective medium could complement at selective conditions, indicating that most of the RBS alterations completely destroyed \textit{aroQ\textsubscript{8}} expression. In contrast, 95 \% of the 306 clones from selective plates also complemented on higher stringency conditions. This could suggest that the initial selection conditions on M9c+Tet\textsuperscript{500} ng/mL + Sal\textsuperscript{0.5} mM were high enough to select for fully functional RBS variants.

Nevertheless, 5 \% of the 306 tested clones exhibited the desired slow growth rate at high stringencies and were analyzed further. The promising variants were re-transformed into strain KA12 and plated on minimal medium varying the concentrations of inducers Sal and Tet. Most of the clones exhibited slow growth only when inducers were excluded; however, variant RBS6 (plasmid pKIMP-ACGQ-RBS6) was also growing weaker than the parent pKIMP-ACGQ in the presence of 0.05 mM Sal and 10 ng/mL Tet (Table 4.1), providing a wider selection window for further experiments.

Sequencing analysis of the newly constructed pKIMP-ACGQ-RBS6 variant revealed that a purine rich sequence (GGGCC) is shifted further upstream from the ATG codon and thus is suboptimally positioned for putative ribosome binding. Also, two rare start codons GTG and TTG were generated during mutagenesis; however, they occur out-of-frame for the MtCM gene (Figure 4.2B).

4.2.1.2 Construction plasmids pKINT and pCyP1

The C-terminally His\textsubscript{6}-tagged split intein construct I\textsubscript{C}-peptide-I\textsubscript{N} based on DnaE from \textit{Synechocystis} sp. PCC6803 intein\textsuperscript{[176]} was designed according to the previously described SICLOPPS
construct.\textsuperscript{175} It included the Ala35His mutation, which was shown to facilitate peptide cyclization, and the tyrosinase inhibitor pseudostellarin F (PsF) sequence SGGYPPLL between the originally C-terminal (I\(_C\)) and N-terminal (I\(_N\)) intein fragments. The splicing product of the resulting I\(_C\)-PsF-I\(_N\) construct is cyclic PsF (cyclo-[SGGYLPPLL]), which was successfully used to test the general utility of the split intein for protein splicing by Scott et al.\textsuperscript{175} The gene encoding I\(_C\)-PsF-I\(_N\) was optimized for \textit{E. coli} and cloned under the P\(_{T7}\) inducible promoter resulting in plasmid pKINT-PsF. The backbone of plasmid pKINT-PsF was based on the high-copy number plasmid pMG214 (M. Gamper, unpublished results) where the ampicillin resistance gene (\textit{bla}) was replaced with kanamycin resistance gene (\textit{kanR2}) allowing for the subsequent penicillin enrichment experiments.

To test the intein production and peptide cyclization efficiency, we used the I\(_C\)-PsF-I\(_N\) construct and a newly designed I\(_C\)-CMi-I\(_N\) construct, which carried the MtCM-based peptide sequence (CMi, SGRLGHPL). The first (Ser-Gly) and the last (Pro-Leu) two residues of the PsF peptide were included into the CMi peptide to facilitate cyclization. It was previously shown, that a Cys, Ser, or Thr residues at the first position of the target peptide is mechanismically essential to serve as a nucleophile in the split intein catalyzed transesterification reaction\textsuperscript{177} while Gly at the second position provides flexibility for the forming cycle.

After β-D-1-thiogalactopyranoside (IPTG)-induced gene expression in \textit{E. coli} KA13/pKINT-PsF and KA13/pKINT-CMi, SDS-PAGE analysis of the crude cell lysate and elution fractions from Ni-NTA column allowed for the visualization of bands corresponding to the splicing products (Figure 4.3). The resolution was insufficient to separate the fusion protein and thioester intermediate, or the lariat intermediate and I\(_C\); therefore, only three bands, including the one corresponding to I\(_N\) were visible. The fusion proteins, the thioester intermediates and the I\(_C\) and I\(_N\) fragments were confirmed by ESI mass spectroscopy (data not shown).

The cyclic products were too small to be identified by SDS-PAGE. Therefore, the cyclized PsF and CMi were extracted with \textit{n}-butanol from crude cell lysate and Ni-NTA elution. As a negative control, an empty vector pKINT-0 was used. It contained a stuffer fragment
Chapter 4: In vivo selection for cyclic peptide modulators of the MtCM-MtDS complex

A

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>CMi 0.05 mM IPTG</th>
<th>CMi 0.1 mM IPTG</th>
<th>PsF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>27</td>
<td>27</td>
<td>27</td>
</tr>
</tbody>
</table>

Figure 4.3 SDS-PAGE analysis of intein splicing products. (A) Cell lysate of KA13/pKINT-CMi was analyzed before (0 h) and after 2, 4, and 27 hours of induction with 0.05 mM or 0.1 mM IPTG. As a control, the cell lysate of KA13/pKINT-PsF was analyzed after 27 hours of induction with 0.1 mM IPTG. (B) Cell lysates after 27 h induction (see above) were applied onto Ni-NTA columns. The flow-through (FT), wash (W) and elution (E) fractions were analyzed by SDS-PAGE. M stands for the Amersham Low Molecular Weight marker (GE Healthcare); values are in kDa.
Regulation, effectors, and evolvability of chorismate mutases from Corynebacterales

instead of a peptide and a stop codon downstream from \( l_C \) resulting in a prematurely terminated intein. The mass spectra of the elution fractions measured by LC-MS showed the existence of a peak that corresponded to the CMi cyclic peptide (817.3 g/mol), which was absent in an analogously prepared elution sample from the pKINT-0 negative control (data not shown). However, we could not confirm the existence of cyclic PsF in any of the samples, even though, the cyclization of PsF using an identical split intein construct has been shown in the literature.\footnote{175} Nevertheless, the evidence for the splicing reaction from both \( l_C\text{-PsF}\text{-I}_N \) and \( l_C\text{-CMi}\text{-I}_N \) constructs provided by SDS-PAGE, ESI MS and LC-MS indicates that the SICLOPPS construct assembled in this thesis is suitable to produce CPs \textit{in vivo}.

The \( P_{T7} \) promoter system on the pKINT plasmid proved to be highly efficient for split intein production (Figure 4.3). However, it coupled the intein gene expression with expression of \( pheC \) and \( tyrA^* \) on the pKIMP-ACGQ and pKIMP-ACGQ-RBS6 plasmids since the \( lacO \) sequence was upstream of these genes. This precluded the possibility to independently regulate the intein production, because reducing or even excluding inducer IPTG also caused impaired complementation ability of the KA12/pKIMP-ACGQ and KA13/pKIMP-ACGQ constructs (data not shown).

To decouple the expression of the intein construct from \( pheC \) and \( tyrA^* \), a new plasmid pCyP1 was constructed including the \( P_{prpB} \) promoter and the \( prpR \) repressor from the \( prpBCDE \) promoter expression system (Figure 4.2A).\footnote{178} Inducing the gene expression with 10 mM propionate showed that protein production was comparable to that on the pKINT plasmid (data not shown), indicating that pCyP1-based CP libraries can be used for \textit{in vivo} selection of the MtCM-MtDS complex modulators.

**4.2.2 Construction of the cyclic peptide libraries**

The CP libraries cpL6, cpL7, cpL8, and cpL9 encoded peptides of 6 to 9 randomized residues specified by multiple NNK codons (Figure 4.4A). All libraries allowed Cys or Ser at the first position to facilitate the intein-mediated transesterification reaction.\footnote{177} It was achieved via the TSC codon (S represents C or G), which excluded the stop and Trp codons and only allowed frequently used codons in \textit{E. coli}. The second
position of the cycle encoded Gly to provide flexibility for cycle formation and to avoid racemization during the subsequent chemical synthesis of selected CPs. The 9-residue peptide library cpL9 was additionally provided with the chromophore Trp at the third position to facilitate easy purification of the peptides.

Figure 4.4 Designed CP libraries. The inserted sequence between the I_C and I_N parts of the split intein is shown in brackets. The first position of the peptide was randomized via the TSC codon to allow Cys or Ser. The residues that are indicated by X, were randomized via the NNK codon. The cpLCMi library encoded the last four residues of MtCM (green sticks in the figure on the right) that interact with MtDS (blue spheres). The yellow dashed line indicates the 6 Å distance between the two Cα atoms of residues Arg87 and His90.

The cpL6, cpL7, cpL8, and cpL9 libraries could be used for selection of both activators and inhibitors of the MtCM-MtDS complex. The cpLCMi library was exclusively designed for inhibitors and encoded the last four C-terminal MtCM residues as described for the pKINT-CMi construct. It was hypothesized that the cyclic product from this library would compete with the C-terminus of MtCM for binding at the same cleft in MtDS and thus disrupt the MtCM-MtDS complex (Figure 4.4B). The length of the peptide was set to 8 residues to retain the cycle big enough for mimicking the hook formed by MtCM residues ^87RLGH^90 when they bind to MtDS. The distance of 6 Å was determined between the backbone Cα atoms of Arg87 and His90 residues in the crystal structure of MtDS-bound MtCM. This assured that a four-residue connecting loop would certainly permit the correct fold of the RLGH sequence within the CP (Figure 4.4B).
The individual libraries cpL6, cpL7, cpL8, cpL9, and cpLCMi were cloned under $P_{prpB}$ promoter on a pCyP1 plasmid and prepared for selection experiments in the reporter strain KA12/pKIMP-ACGQ-RBS6. However, the transformation efficiency was relatively low if the library plasmids were transformed directly after ligation. Moreover, preliminary growth tests of KA12/pKIMP-ACGQ-RBS6 carrying pCyP1-PsF, which encodes I$_C$-PsF-I$_N$ construct, showed that the cells were sensitive to temperature and produced exopolysaccharide when grown on standard M9c agar plates. Therefore, the carbon source in the growth medium was changed from glucose to glycerol to maintain pH stability during growth, and a step-wise library construction approach was adopted to increase the number of transformants of the CP libraries.

First, the gene libraries were individually ligated into the pCyP1 vector. The ligation products were transformed into the electrocompetent KA12 strain, since the transformation efficiency was better when the helper plasmid was absent. A portion of the electroporated cells was plated on non-selective M9c*+FY (M9c medium containing glycerol as a carbon source and supplemented with Phe and Tyr) to determine the library sizes. The determined library sizes for cpL6, cpL7, cpL8, cpL9 and cpLCMi at this step were 4.99×10$^7$, 1.64×10$^7$, 1.80×10$^8$, 1.10×10$^8$, and 1×10$^7$, respectively. The remaining cells were plated on M9c*+FY agar plates and the plasmid DNA was isolated from the grown cells.

Second, the supercoiled plasmid DNA from the previous step was transformed into the reporter strain KA12/pKIMP-ACGQ-RBS6. Again, a portion of the electroporated cells were plated on M9c*+FY agar plates to determine the number of transformants, which were 1.96×10$^9$, 2.15×10$^9$, 6.04×10$^8$, 3.21×10$^8$, and 5×10$^5$ for libraries cpL6, cpL7, cpL8, cpL9, and cpLCMi respectively. The remaining cell culture was incubated overnight in non-selective liquid M9c*+FY+Prp$^{10\text{ mM}}$ additionally containing 10 mM propionate. This step allowed removal of the untransformed cells and the clones expressing toxic CPs. Moreover, early production of intein allowed accumulation of sufficient amounts of CPs for the cells to acquire the required phenotype prior to selection for activators (see section 4.2.3) or inhibitors (see section 4.2.4).
4.2.3 Selection for CP activators of the MtCM-MtDS complex

The reporter strain KA12/pKIMP-ACGQ-RBS6 showed slower growth than the original KA12/pKIMP-ACGQ on selective conditions excluding the CM inducer salicylate (Table 4.1). However, to enable selection for CP activators of the MtCM-MtDS complex, the stringency needed to be further increased adding toxic DL-para-fluorophenylalanine (pFPhe).\textsuperscript{[179,180,181]}

The bacterial KA12/pKIMP-ACGQ-RBS6 cultures transformed with libraries cpL6, cpL7, cpL8, cpL9, and cpLCMi in the reporter strain KA12/pKIMP-ACGQ-RBS6 were grown in liquid M9c*+FY+Prp\textsuperscript{10} mM prior to selection experiments (see section 4.2.2). After the growth, a fraction was plated on non-selective M9c*+FY agar plates to determine the number of viable cells that will be used for selection. This number was regarded as complementation frequency of 100 % and was typically adjusted to 10\textsuperscript{9}. This number allowed covering the entire number of transformants that were obtained after KA12/pKIMP-ACGQ-RBS6 transformation with supercoiled DNA from individual CP library.

The libraries were plated on selective M9c*+Prp\textsuperscript{10} mM+pFPhe\textsuperscript{40} or \textsuperscript{100} μM agar plates containing 40 μM or 100 μM of toxic pFPhe. All the individual libraries exhibited complementation frequency of less than 10\textsuperscript{-4} %. Approximately the 100 fastest growers from each library were picked and tested on the highest available stringency M9c*+Prp\textsuperscript{10} mM+pFPhe\textsuperscript{100} μM plates. The 30 best growing clones from each library were individually pooled, re-transformed into fresh KA12/pKIMP-ACGQ-RBS6 and plated on selective plates M9c*+Prp\textsuperscript{10} mM4+pFPhe\textsuperscript{100} μM to re-confirm that the observed phenotype was linked to plasmids. The clones that complemented after re-transformation were picked and sequenced.

Sequence analysis of several clones from libraries cpL6, cpL7, and cpL8 revealed various cloning artifacts. A few clones had a foreign insert between the NdeI and KpnI cloning sites used for construction of the libraries. The insert carried a part of the aroF gene coding DAHP synthase from \textit{E. coli} chromosomal DNA. It was assumed that the entire \textit{aroF-tyrA} operon\textsuperscript{[182,183]} (carrying CM-PDH) was extracted during vector preparation from Xl1-Blue cells. The \textit{tyrA} gene was now on the
high-copy number plasmid pCyP1; therefore, the clones carrying this insert had an advantage on the selective conditions due to the increased intracellular CM activity.

Figure 4.5 Selected peptides from the cpL9 library. (A) Sequencing results of clones complementing the growth of the reporter strain KA12/pKIMP-ACGQ-RBS6 on high stringency M9c*+Prp10 mM+pFPhe100 μM plates. (B) CM activity of the MtCM-MtDS complex in the presence of chemically synthesized CPs. The initial velocity was monitored at 30°C using 23 μM chorismate in 50 mM BTP+ pH 7.5. The final concentrations of the components were: (i) 25 nM MtCM (25 nM MtDS), (ii) 5 μM of each Phe and Tyr (iii) 1.25 μM of peptide added from the stock in N,N-dimethylformamide (DMF). The initial velocities of chorismate consumption in the presence of 0.125 % DMF representing 100% were 0.25 ± 0.2 s⁻¹ and 0.24 ± 0.04 s⁻¹ when Phe and Tyr were present or absent, respectively. The peptides L9s1 (cyclo-[CGWQQADAM]), L9s4 (cyclo-[SGWKLCRR]), L9s6 (cyclo-[CGWTSPSPY]), L9s11 (cyclo-[SGWVKTRNY]), L9s12 (cyclo-[CGWLYHK]), L9s20 (cyclo-[CGWRNKRL]), and L9s21 (cyclo-[SGWPGIAE]) were tested.

Nevertheless, selected clones from the cpL9 library showed the presence of intact intein sequences coding for various peptides, despite a few clones that contained a stop codon within the peptide (variants...
L9s17 and L9s21), (Figure 4.5A). The plasmid integrity was analyzed by restriction digestion and showed no evidence of a foreign insert elsewhere on the plasmid. The in vivo complementation tests in KA12/pKIMP-ACGQ-RBS6 showed that at 30 °C these clones could complement the FY deficiency better than the cells carrying plasmid pCyP1-PsF or the empty vector pCyP1-0. However, at 25 °C the controls grew equally well as the selected clones from cpL9 and no response to inducer propionate was observed neither at 30 °C nor at 25 °C (data not shown).

To test if the selected peptides were MtCM-MtDS activators, several chemically synthesised CPs (see legend of Figure 4.5) were purchased and tested by the direct CM assay. The initial experiments showed no CM reaction acceleration of the MtCM-MtDS complex at the protein concentration above the complex $K_{d,app}$ (data not shown). Therefore, following the example of Trp activation, which is only observed at lower protein concentration (see section 2.2.2), the peptides were tested in the presence of equimolar MtCM and MtDS (25 nM each). In addition, the CPs were tested in the assay containing 5 μM of each Phe and Tyr that have been shown to disrupt the complex$^{[28]}$ (see section 2.2.2) to see whether the complex could be stabilized and the CM reaction rate restored. Unfortunately, as shown in Figure 4.5, none of the chemically synthesized CPs could accelerate the CM reaction of the MtCM-MtDS complex.

### 4.2.4 Genetic selection system for CP inhibitors of the MtCM-MtDS complex

Penicillin enrichment$^{[184]}$ was a common method to isolate various auxotrophic E. coli mutants for elucidation of the metabolic pathways. This antibiotic prevents the transpeptidation reaction during the formation of cross linkages in the cell wall$^{[185, 186]}$. Upon penicillin addition to the mixed cultures of prototrophic and auxotrophic cells in minimal medium without the required metabolite only the non-growing auxotrophs can survive while the well-growing prototrophic cells become osmotically fragile and lyse.

In this study, the clones that produce CP inhibitors of the MtCM-MtDS complex become Phe and Tyr (FY) auxotrophs due to reduced CM activity and should survive penicillin treatment in minimal medium.
lacking FY. The auxotrophs can afterwards be retrieved by removing penicillin and re-growing the cells on a FY supplemented medium. The detailed protocol for penicillin enrichment was developed following the previously described procedures that were adapted to specifics of the reporter strains KA12/pKIMP-ACGQ-RBS6 or KA12/pKIMP-ACGQ. One penicillin enrichment experiment included three iterative cycles of (i) starvation, (ii) penicillin treatment, and (iii) re-growth.

(i) The starvation step was performed in M9c* minimal medium lacking Phe and Tyr but containing inducers for MtCM, MtDS, and the intein as well as 20 % sucrose. During this step, the FY auxotrophs use up the intracellular nutrients and stop dividing. The presence of sucrose prevents cells from lysis during the subsequent penicillin treatment by fixing them in the spheroplast stage. It is important to prevent the release of nutrients from the lysing cells into the growth medium to avoid the cross-feeding. During this process auxotrophs start growing due to the presence of the required metabolites and thus become sensitive to penicillin.

(ii) The penicillin treatment in the starvation medium was initiated after the cells underwent up to three doublings but did not exceed the density of $10^7-10^8$ cells/mL. The relatively high concentration of 2000 U/mL penicillin was chosen to shorten the incubation time required for elimination of prototrophic cells. Keeping the cell density low, applying high penicillin concentration, and reducing the shaking frequency in sucrose-supplemented medium minimized the risk of cross-feeding. A fraction of the cells were plated on supplemented M9c*+FY agar plates to determine the total number of viable cells in the culture before and after the penicillin treatment. This allowed determination of the survival rate, which should increase after each cycle indicating the accumulation of auxotrophic cells in the culture.

(iii) After the penicillin treatment, the cells were washed and re-grown in M9c*+FY medium to multiply the enriched auxotrophs.

As mentioned before, a cycle of starvation, penicillin treatment, and re-growth was repeated three times per one penicillin enrichment experiment. It is not recommended to perform more cycles to prevent accumulation of spontaneous mutations that result in penicillin-
resistant cells. For the same reason, the plasmid DNA from the re-grown cells after one enrichment experiment was isolated and re-transformed into a fresh reporter strain prior to subjecting them to the next penicillin enrichment.

The CP libraries cpL6, cpL7, cpL8, cpL9, and cpLCMi in KA12/pKIMP-ACGQ-RBS6 strain that have been grown in liquid M9c*+FY+Prp10 mM medium (see section 4.2.2) were subjected to penicillin enrichment. The cpL6, cpL7, cpL8 and cpL9 libraries were pooled together (cpL6-9), since we were interested in the best inhibitors regardless of the peptide length, while the cpLCMi library was treated separately. We assumed that cpLCMi might contain more MtCM-MtDS complex inhibitors and the best clones from this library would dominate the culture after the penicillin enrichment, precluding the discovery of other types of peptides.

For the first penicillin enrichment, approximately 10^7 viable cells from the cpLCMi library and 2x10^7 from each individual library cpL6, cpL7, cpL8, and cpL9 were used. The survival rates after the third cycle of penicillin treatment were ≈ 3 % and ≈ 1 % for the cpLCMi and cpL6-9 libraries, respectively.

The plasmid DNA from individually enriched cpLCMi and cpL6-9 library members were re-transformed into fresh KA12/pKIMP-ACGQ-RBS6 cells and subjected to the next penicillin enrichment. The survival rate for cpL6-9 further increased to 48 %, indicating strong enrichment for penicillin-resistant cells.

To increase the stringency for the next penicillin enrichment, the isolated plasmids from cpL6-9 and cpLCMi were individually re-transformed into KA12/pKIMP-ACGQ strain. It was assumed that increasing the intracellular concentration of the target MtCM would allow enrichment for stronger CP inhibitors.

After the first enrichment of cpLCMi and cpL6-9 libraries in KA12/pKIMP-ACGQ strain the survival rates were ≈ 0.3 % and ≈ 32 %, respectively, indicating that fewer cells can survive under increased stringency conditions. The enrichment for cpL6-9 was repeated again after re-transformation and resulted in > 140 % survival rate.

The high percentage of survivors in cpL6-9 was suspicious and indicated that the cells not only survive the penicillin treatment, but
also grow in minimal medium lacking Phe and Tyr. The growth was confirmed when several clones from the enriched cpL6-9 library were tested on minimal M9c* medium verifying, that there were no FY auxotrophs among them. Sequencing and restriction digest analysis of several clones from the enriched cpLCMi and cpL6-9 libraries revealed that the enriched plasmids contained foreign DNA inserts carrying β-lactamase genes.

The enriched clones from cpLCMi library were the artifacts that most probably originated during the cloning. The sequencing analysis of the DNA insert between the cloning sites NdeI/KpnI in the clones of the cpLCMi library identified the *E. coli* chromosomal *frdA* gene encoding fumarate reductase.\[188\] The gene is part of an operon that also carries the *E. coli* ampC gene\[189\] encoding a β-lactamase that causes resistance to many β-lactam antibiotics, including penicillin.\[190\] The digestion of the genomic DNA containing the *frdA* and *ampC* genes with NdeI and KpnI yields a 4311 bp fragment, which is similar in size with the accordingly digested 4164 bp pCyp1 fragment required for library construction. It is, therefore, possible, that the contamination with genomic DNA occurred during the gel purification of the digested vectors and the unwanted fragment was subsequently ligated with the pCyp1 fragment.

An even more peculiar DNA insert was identified in the enriched plasmids from the cpL6-9 library. Analytical restriction analysis showed the presence of a large ≈ 3000 bp insert in the vicinity of the origin of replication and the sequencing analysis revealed that it originated from a pKSS-based plasmid,\[191\] which is widely used in our laboratory. The insert carried the genes for AmpR and the secreted CM from *Salmonella enterica* explaining why the cells not only were resistant to penicillin but also grew well in the minimal medium lacking FY. It remains unclear how this insert got cloned into the pCyp1-based library plasmid. The most probable mechanism could be a homologous recombination-like process since the ori regions of pCyp1 and pKSS are identical.

Nevertheless, even though the penicillin enrichment of cpLCMi and cpL6-9 libraries did not yield the desired CP inhibitors, it clearly showed the power of this method to enrich for rare penicillin-resistant clones.
4.2.4.1 Counter-selection system for elimination of CP inhibitors of other shikimate pathway enzymes

As described earlier in section 4.2.4, penicillin enrichment can result in identification of a variety of artifacts. While the major issue is penicillin resistant cells emerging due to β-lactamase gene on a plasmid, the target FY auxotrophic phenotype can also occur for reasons other than CM inhibition. Therefore, it is important to develop methods that ensure a robust elimination of these artifacts.

One of the issues that can cause FY auxotrophy is inhibition of other shikimate pathway enzymes essential for biosynthesis of Phe or Tyr in the reporter strain. For example, the clones producing CPs that inhibit *E. coli* shikimate pathway enzymes or PDH and PDT encoded by the *tyrA* and *pheC* genes on plasmids pKIMP-ACGQ and pKIMP-ACGQ-RBS6 plasmids will be enriched as FY auxotrophs after penicillin enrichment.

To eliminate these false positives an additional counter-selection system was developed that involves chorismate mutase from *M. jannschii* (MjCM). The MtCM gene on pKIMP-ACGQ-RBS6 was replaced by the MjCM gene to obtain plasmid pKIMP-ACGQ-RBS6-MjCM. The enriched CP library can then be transformed into KA12/pKIMP-ACGQ-RBS6-MjCM and grown on minimal medium. It was assumed that clones producing CPs that inhibit the off-target enzymes from the shikimate pathway will arrest the growth of KA12/pKIMP-ACGQ-RBS6-MjCM and thus be eliminated. At the same time, the CPs specific for MtCM or the MtCM-MtDS complex, but not to the other enzymes, will not affect the growth of KA12/pKIMP-ACGQ-RBS6-MjCM and will be retrieved.

MjCM, which has similar catalytic activity as the MtCM-MtDS complex, was cloned under the impaired RBS of *P. sal* to obtain a low intracellular CM concentration. This should prevent the drain of the products of any weakly inhibited shikimate pathway enzymes upstream from the CM. Moreover, MjCM is structurally similar to MtCM; therefore, the system would allow elimination of the CP inhibitors that are not species specific.

The complementation of the KA12/pKIMP-ACGQ-RBS6-MjCM was compared to KA12/pKIMP-ACGQ-RBS6 and KA12/pKIMP-ACGQ to
assessed the in vivo CM activity of MjCM. The bacteria were grown at 30 °C or 37 °C on minimal plates containing various combinations of the inducers salicylate and tetracycline.

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Medium</th>
<th>KA12/pKIMP-ACGQ</th>
<th>KA12/pKIMP-ACGQ-RBS6</th>
<th>KA12/pKIMP-ACGQ-RBS6-MjCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 °C</td>
<td>FY</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Tet\textsuperscript{100} Sal\textsuperscript{0.1}</td>
<td>9</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Tet\textsuperscript{100} Sal\textsuperscript{0}</td>
<td>9</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Tet\textsuperscript{0} Sal\textsuperscript{0.1}</td>
<td>8</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Tet\textsuperscript{0} Sal\textsuperscript{0}</td>
<td>8</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>37 °C</td>
<td>FY</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Tet\textsuperscript{100} Sal\textsuperscript{0.1}</td>
<td>8</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Tet\textsuperscript{100} Sal\textsuperscript{0}</td>
<td>9</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Tet\textsuperscript{0} Sal\textsuperscript{0.1}</td>
<td>7</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Tet\textsuperscript{0} Sal\textsuperscript{0}</td>
<td>7</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

\(a\) The medium was M9c\textsuperscript{*} with chloramphenicol and the indicated supplements. Sal, salicylate, value in superscript indicates the concentration in mM and Tet, tetracycline, value in superscript indicates the concentration in ng/mL.

\(b\) The single colony size was evaluated in the scale from 0 (no growth) to 9 (larger that 1.5 mm in diameter) after incubation at 30 °C for 72 h.

No temperature sensitivity was detected and all constructs were growing equally well on non-selective M9c\textsuperscript{*}+FY plates suggesting that MjCM was non-toxic in vivo (Table 4.1). However, the KA12/pKIMP-ACGQ-RBS6-MjCM transformation was growing slower than KA12/pKIMP-ACGQ-RBS6 on all minimal plates, indicating that in vivo CM activity of MjCM is weaker than that of the MtCM-MtDS complex.

These results show that that KA12/pKIMP-ACGQ-RBS6-MjCM can be used for the elimination of weak CP inhibitors of other shikimate pathway enzymes since the weak intracellular CM activity will prevent the leaky production of Phe and Tyr.

4.3 Discussion

In this chapter we describe the steps towards the development of a novel phenotype-based selection system for PPI modulators. Since constrained peptides are gaining more attention as promising compounds to affect PPIs, the previously described SICLOPPS system was employed to generate intein-mediated peptide cyclization within
Chapter 4: In vivo selection for cyclic peptide modulators of the MtCM-MtDS complex

the cell. We have cloned the required construct under the $P_{T7}$ and $P_{prpB}$ promoters on plasmids pKINT and pCyP1, respectively, and tested the production of split inteins with model peptides PsF (cyclo-[SGGYLPPL]) and CMi (cyclo-[SGRLGHPL]). The visualization of the protein fragments resulting from the intein splicing by SDS-PAGE as well as detection of the cyclized CMi by LC-MS indicated that a functional SICLOPPS construct was created.

The SICLOPPS system has proven itself as an efficient way to find PPI inhibitors when a reverse two-hybrid system (the disruption of a PPI is reported by an antibiotic-resistance phenotype of a bacterial clone\cite{192, 193}) is used. However, the two-hybrid system requires modification of the target proteins by fusing them to transcription regulation factors, which might, in certain cases, be not possible due to the specifics of the interacting proteins. Moreover, the fusion strategy increases the chances of finding inhibitors that are specific to the modified version of a target protein.

Here we describe an approach that does not require any target protein modification and relies on the activity switch of the partner enzyme(s). This system, however, is limited to metabolic enzymes that exhibit this unique property.

MtCM’s activity is increased by the complex formation with MtDS and this fact can be used as a readout for modulation of this transient PPI. We base the reporting assay on the powerful CM complementation system, which, upon suitable selective conditions, can report the increase and decrease of the metabolically essential CM activity.

In this work, we re-engineered the CM selection system by creating the helper plasmid pKIMP-ACGQ that, in addition to the monofunctional $tyrA^*$ and $pheC$ genes, carries the independently controlled target genes $aroG$ and $aroQ_\delta$ under the $P_{tet}$ and $P_{sal}$ promoters, respectively. This configuration allowed adjusting the intracellular levels of MtDS and MtCM and thus the growth of the reporter strain on minimal medium lacking Phe and Tyr. However, the relatively fast growth of KA12/pKIMP-ACGQ indicated that the stringency for the subsequent selection experiments might not be high enough. The approach of leveling down the intracellular concentration of MtCM by introducing a protein degradation tag was considered not
suitable since a modification of the target protein might result in selection for CPs specific to it and e.g. the C-terminal SsrA tag would modify the part of MtCM that is essential for complex formation with MtDS. Moreover, the tagged MtCM could serve as a reporter for clones that produce CPs inhibiting the protein degradation machinery. The restored pool of MtCM enzymes would again permit the growth on selective medium thus increasing the abundance of these artifacts during the selection for CP activators of the MtCM-MtDS complex. In fact, CP inhibitors for the ClpXP protease have been identified using the reporter GFP fused to the ClpXP degradation tag. Therefore, we resorted to an approach where RBS of the P<sub>sal</sub> promoter was impaired by mutations on plasmid pKIMP-ACGQ-RBS6 to dial down the production of MtCM.

Unfortunately, the pilot selection experiments did not result in successful identification of CP activators or inhibitors of the MtCM-MtDS complex; therefore the system will require further optimization.

The biggest setback for both selections was the identification of the artifacts associated with undesirable DNA insertions in the library plasmid pCyP1. The insertion of <i>E. coli</i> chromosomal DNA carrying the <i>pheA</i> and <i>tyrA</i> genes resulted in increased intracellular CM activity that granted the survival of the cells on minimal medium. While this is not an issue for the penicillin selection, since the fast growing cells will be eliminated, it precluded the selection for activators. However, preparing the cloning vector from a CM-deficient strain and performing a restriction cleanup prior to transformation using restriction endonucleases specific to <i>E. coli</i> <i>pheA</i> and <i>tyrA</i> genes could completely eliminate such artifacts.

Moreover, the low number of complementers (<10<sup>-4</sup> %) under the stringent conditions indicates that the abundance of activating peptides in the tested libraries might be too low. Therefore, it is worth testing different lengths of peptides or designing a SICLOPPS-based system where unnatural amino acids<sup>173</sup> are introduced. The better-suited libraries would thereby be enriched with activating CP candidates that would also diminish the possibility of selecting the rare artifacts such as the aforementioned gene inserts or CPs that affect other cellular machineries and, therefore, cause faster growth on minimal medium.
Chapter 4: In vivo selection for cyclic peptide modulators of the MtCM-MtDS complex

The chemically synthesized CPs corresponding to the ones in the selected clones of the cpL9 library did not activate the MtCM-MtDS complex in the *in vitro* CM assay. This could suggest that splicing intermediates are better activators than the cyclized products. This assumption can be tested using an intein splicing-disabled construct containing Thr69Ala and His72Ala mutations that have been shown to block the first step of the splicing mechanism.\textsuperscript{[195, 196, 197]} Since the splicing intermediates are not obtained from such an intein-disabled construct, the clones expressing the CP of interest should not grow on the minimal medium and thus be eliminated.

Moreover, since the medium was supplemented with the toxic pFPhe, the peptides that had this unnatural amino acid incorporated could have been better activators than the ones synthesized chemically and containing the proteinogenic Phe. It might, therefore, be necessary to avoid using pFPhe as in the selection for CP activators.

Bacterial cells that constantly produce an active CM might not survive due to a depleted chorismate pool necessary as a precursor for other aromatic compounds. It could even be that the very best activators were lost during the selection procedure due to their bactericidal effects. Therefore, further reduction of MtCM levels might in fact be essential and can be achieved by utilization of different promoter systems, low copy number plasmids or even constructing a strain that carries the target genes in its chromosomal DNA.

Similarly to the selection for the activators, the penicillin enrichment for inhibitors yielded mostly artifacts. These included the β-lactamase genes *ampC* or *ampR*, originating from chromosomal *E. coli* DNA or pKSS-based plasmids, respectively. These artifacts can also be removed by a restriction digestion using *SalI* or *Scal* (and *XbaI*) restriction endonucleases that are specific for *ampC* or *ampR*, respectively, but do not cut elsewhere in the pCyP1 vector.

The appearance of the genomic *ampC* gene on the library plasmid might signal the general requirement for deleting this gene from the chromosomal DNA of the reporter strain KA12. However, we observed that most of the clones in the cpLCMi library still lysed even though the clones producing β-lactamase already dominated the culture after several rounds of penicillin enrichment. This indicates that even when
expressed on a high copy number plasmid, the \textit{ampC} gene product is not strong enough to cause sufficient resistance. Therefore, a high number of true CP-dependent auxotrophs may dominate over these rare artifacts. It is, therefore, worth testing alternative libraries that produce CPs of different lengths or composition to ensure that desirable clones dominate the cultures.

Additionally, to prevent the accumulation of β-lactamase-carrying cells during the enrichment experiment a combination of two antibiotics and a selection strategy in solid medium\cite{198} can be used. By applying a mixture of penicillin and D-cycloserine\cite{199} any well-growing penicillin resistant cells are eliminated by D-cycloserine-induced lysis. Conversely, D-cycloserine resistance could also appear due to foreign DNA inserts but simultaneously added penicillin could eliminate these artifacts. Furthermore, a selection strategy in solid medium,\cite{198} where single antibiotic-resistant clones are entrapped in the agar and cannot overtake the culture due to their faster growth, could be used as an alternative to liquid selection. However, this method allows screening of only up to $10^2$-$10^3$ clones per plate (whereas in liquid culture $10^8$-$10^9$ clones can be screened at once) and it would thus only be beneficial when the population is already enriched for the desired best candidates.

Despite the appearance of penicillin-resistant artifacts, too low selection stringency might be another issue precluding penicillin enrichment for CP inhibitors. It is conceivable that even when the MtCM-MtDS complex is disrupted by the CP variant, the remaining CM activity of uncomplexed MtCM is high enough to enable cell growth in minimal medium. This causes the normal cell division in the starvation medium and thus lysis during the subsequent penicillin treatment. This assumption can be tested by constructing a pKIMP-ACGQ (or pKIMP-ACGQ-RBS6) variant where the MtDS gene has been deleted. In the case of residual growth of the cells producing MtCM alone, the selection stringency must be substantially increased. The most straightforward way to do so would probably be using a strain, which contains single copies of MtCM and MtDS genes on its chromosomal DNA.

Finally, it might be interesting to replace the currently used \textit{pheC} and \textit{tyrA*} from \textit{P. aeruginosa} and \textit{E. herbicola} on the helper plasmid
pKIMP-ACGQ(-RBS6) with the corresponding genes from *M. tuberculosis*. As discussed in section 4.2.4.1, inhibition of these enzymes will also cause an FY auxotrophic phenotype and allow selection by penicillin. This strategy should not only increase the abundance of the auxotrophic cells in the culture, but also increase the chances of finding inhibitors for at least one enzyme needed for the mycobacterial biosynthesis of aromatic amino acids.

### 4.4 Materials and Methods

#### 4.4.1 Materials and general procedures

Plasmid DNA purification from *E. coli* cultures was performed using the ZR-MiniPrep Classic kit (Zymo research). DNA from PCR, restriction digestions and ligations was purified either by agarose gel electrophoresis in TAE buffer followed by Zymoclean™ Gel DNA Recovery Kit (Zymo research) or directly from reaction mixtures using the DNA Clean and Concentrator Kit-5 (Zymo Research). The DNA concentration was determined spectrometrically using NanoDrop (Thermo Scientific). DNA manipulations were performed using standard procedures[^155] or according to the manufacturer’s recommendations. Sanger DNA sequencing and oligonucleotide synthesis was performed by Microsynth AG. Oligonucleotides were purified by desalting for primers used in routine PCR reactions and by HPLC for degenerate primers. Restriction endonucleases, Phusion DNA-polymerase and T4 DNA ligase were purchased from New England Biolabs. DL-para-fluoro-phenylalanine was from Bachem. Chemically synthesized CPs L9s6 and L9s20 were purchased from ABX (Radeberg, Germany), the other CPs (L9s1, L9s4, L9s11, L9s12, and L9s21) were purchased from China Peptides (Shanghai, China). Penicillin G (Sodium salt) and other chemicals were purchased from Sigma-Aldrich/Fluka.

#### 4.4.2 Bacterial strains and plasmids

*E. coli* strain XL1-Blue (Stratagene, La Jolla, CA, USA) was used for cloning, the strains KA12 (genotype: F', Δ(srlR-recA)306::Tn10, Δ(pheATyrA-aroF), thi-1, endA1, hsdR17(ρK, mK'), Δ(argF-lac)205(U169), supE44, Tet^R)[^51, 200] and KA13 (genotype: F-, Δ(srlR-recA)306::Tn10
Δ(pheA-tyrA-aroF) thi-1, endA1, hsdR17(ρK, mK+), λ(DE3) [P<sub>lacUV5</sub>-expressed T7 RNA pol gene], Tet<sup>8</sup>) that carries a chromosomal IPTG-inducible T7 RNA polymerase gene<sup>152, 153</sup> were used for cloning, protein production, and in vivo assays.

Plasmids pKIMP-ACG,<sup>[26]</sup> pKTCMM-H,<sup>[48]</sup> pMG-P<sub>tet-GFP</sub>,<sup>[201]</sup> and pKMCMT-W<sup>[202]</sup> were described previously. Plasmids pET-29b(+) (Novagen), pGH_Split_Ssp_DnaE (containing the designed split intein gene, from ATG biosynthetics), pPro18-kan (from AddGene) were purchased. Plasmid pMG214 was obtained from M. Gamper (unpublished).

4.4.3 Construction of the split intein gene

The wild-type intein gene sequence was used for the split intein construct (BAA18870.1 <i>Synechocystis sp. PCC 6803</i> DNA polymerase III alpha subunit: Location:1..2694). The circularly permuted variant was generated according to the previously published construct.<sup>[203]</sup> The I<sub>C</sub> part contained an <i>NheI</i> restriction site and an Aal35His mutation at the 3’ end. The I<sub>N</sub> part of the intein gene contained <i>KpnI</i> restriction site at the 5’ end and the C-terminal His<sub>6</sub>-tag followed by a <i>HindIII</i> restriction site at the 3’ end. The sequence encoding Pseudostellarin F<sup>[203]</sup> (SGGYLPPL) was inserted between the I<sub>C</sub> and I<sub>N</sub> parts of the gene. The gene sequence starting from the 5’ end of the I<sub>C</sub> fragment to the <i>HindIII</i> site was optimized for <i>E. coli</i> using GeneOptimizer from GeneArt.<sup>[204]</sup> Subsequently, the <i>P<sub>T7</sub></i> promoter and the RBS region starting from the <i>XbaI</i> restriction site to the start codon on the pET22 plasmid was added upstream of the optimized split intein gene. The entire construct (from <i>XbaI</i> to <i>HindIII</i>) was synthesized by ATG biosynthetics (Germany) and delivered on plasmid pGH_Split_Ssp_DnaE.

4.4.4 Construction of plasmids

4.4.4.1 Plasmid pKIMP-ACGQ

The plasmid pKIMP-ACG was used as the basis for the plasmid pKIMP-ACGQ. pKIMP-ACG contains a chloramphenicol resistance gene, a p15A origin of replication, <i>tyrA</i>* and <i>pheC</i> genes with upstream <i>lacO</i> operator sequences and <i>aroG</i> encoding MtDS under the tandem <i>P<sub>sal</sub></i> / <i>P<sub>T7</sub></i> promoter. To construct plasmid pKIMP-ACGQ with <i>aroG</i> and <i>aroQ</i>,
under the $P_{\text{tet}}$ and $P_{\text{sal}}$ promoters, respectively, a cassette carrying $aroQ_\delta$, $tetR$ and the $P_{\text{tet}}$ promoter was inserted between the $P_{\text{sal}}$ promoter region and start codon for $aroG$, eliminating the $P_{T7}$ promoter sequence.

The $aroQ_\delta$-$tetR$-$P_{\text{tet}}$ cassette was generated by cloning two PCR fragments. The first fragment (PCR I) containing the $aroQ_\delta$ gene was obtained by using plasmid pKTCMM-H as a template. The forward primer 564-KTCMM-Fw (5’-GCCATCAGGACCTGGAATG) introduced the RBS belonging to $P_{\text{sal}}$ and the required BssSI restriction site. The reverse primer 568-pKTCMM-Rev (5’-CATGGGCAGCCAAAAAAAGCCCGCTCATTAGCGGCTCTAGCTGACTAGTTATTAG) introduced the strong trpA terminator sequence down from the stop codon of the $aroQ_\delta$ gene. The second (PCR II) fragment containing the $tetR$ gene and the $P_{\text{tet}}$ region was obtained by using plasmid pMG-P$_{\text{tet}}$-GFP[201] plasmid as a template and the forward 569-pMGPtet-tetR-Fw (5’-ATCGGCGCGCCCGCAAGGAATGGTGCA) and reverse 567-pMGPtet-tetR-Rev (5’-GCTAGCCATATGTATATCTCCTTC) primers that introduced Ascl and Ndel restriction sites, respectively.

The acceptor plasmid pKIMP-ACG was digested with restriction enzymes BssSI and EcoRV and subsequently with Kpnl and EcoRI in order to isolate the required fragment from the pool of other similarly sized fragments. The 2198 bp and 5274 bp fragments obtained from BssSI/EcoRV/Kpnl/EcoRI digested pKIMP-ACG were used for ligation with the 329 bp fragment obtained after digestion of PCR I with BssSI and Ascl and the 784 bp fragment obtained from Ascl/Ndel digested PCR II.

The resulting 8585 bp plasmid pKIMP-ACGQ was transformed into chemically competent E. coli KA13 cells. The correct DNA sequence of the $aroQ_\delta$-$tetR$-$P_{\text{tet}}$ cassette was confirmed by sequencing using the 336-nahRs5 (5’-ACCACAAGGACCTAGCCAATTT) primer to validate the nahR region and the trpA terminator, the 337-Dss5 (5’-TCTCCCCGTGGAGGATAATT) to validated aroG region and 570-Psal-MtCM (5’-CCGCAGTGATAGAGACCCGT) primer to validate the $P_{\text{sal}}$ promoter and the $aroQ_\delta$ region. Since the four-fragment ligation required restriction digestion within the ori site on the pKIMP-ACG
vector, the validated \textit{aroQs-tetR-P_{tet}} cassette was re-cloned into fresh pKIMP-ACG by using \textit{NdeI} and \textit{KpnI} restriction sites.

4.4.4.2 Plasmid pKIMP-ACGQ-RBS6-MjCM

The gene of the previously described truncated version of MjCM\textsuperscript{[202]} was produced via overlap extension PCR. The PCR I fragment was obtained by using oligonucleotides 570-Psal-MtCM (5'-CGCAGTGATGAGACGGC) and 630-RBS6-rev (5'-GCTACAATTACCAGAGGCC) with pKIMP-ACGQ-RBS6 (for construction of this plasmid, see 4.4.5) as a template. The PCR II fragment was obtained using pKMCMT-W as a template with oligonucleotides 631-pKMCMT-W-Fw (5'-GGCCTCGTGGTAATTGTAGCATGATCGAGAACTTGCT) and 632-pKMCMT-W-rev (5'-CATGCGCCGCCACAAAAAAGCCCGCTCATTAGCGGGCTTCATTTCCCTAAGATATTGC), which carry the \textit{trpA} terminator and the \textit{Ascl} restriction site. The PCR I and PCR II fragments were assembled by using oligonucleotides 570-Psal-MtCM and 632-pKMCMT-W-Rev. The resulting DNA fragment encoding the MjCM gene followed by the \textit{trpA} terminator was digested with \textit{MluI} and \textit{Ascl} to obtain the required 526 bp insert. The 8562 bp pKIMP-ACGQ-RBS6-MjCM plasmid was obtained by ligating the 526 bp insert with the 8036 bp fragment obtained from the appropriately digested pKIMP-ACGQ-RBS6 plasmid. The ligation product was transformed into \textit{E. coli} KA12 and selected on LB+Cam\textsuperscript{30} plates containing 30 \textmu g/mL chloramphenicol. The correct insert was confirmed by sequencing.

4.4.4.3 pKINT-based plasmids

To replace the \textit{ampR} with \textit{kanR2} gene, the 4790 bp pMG214 plasmid was restriction digested with \textit{MluI} and \textit{BamHI}. The 800 bp and 3990 bp fragments were gel-purified and the 3990 bp fragment was additionally digested with \textit{BspHI}. The resulting 1421 bp and 1561 bp fragments were isolated while the 1008 bp \textit{ampR} gene was discarded. The 800 bp, 1421 bp, and 1561 bp fragments from pMG214 and the 875 bp fragment encoding the \textit{kanR2} gene from the \textit{BspHI}-digested pET29b(+) were ligated to obtain the 4657 bp pMG214-kan plasmid.

To test the orientation of the newly inserted \textit{kanR2} gene, the 4657 bp pMG214-kan clone C1.1 was digested with \textit{BamHI} and \textit{ClaI}. The
identified 2123 bp and 2534 bp fragments indicated that the orientation of the *kanR2* gene was clockwise in the resulting plasmid pMG214-kan_C1.1.

To construct the pKINT-PsF plasmid carrying the split intein gene with PsF as a target peptide, the 573 bp fragments from pGH_Split_Ssp_DnaE (see section 4.4.3) digested with *Xba*I and *Hind*III was ligated with the 3438 bp fragment from the accordingly digested pMG214-kan-C1.1. The correct DNA sequence of the insert in the newly constructed 4011 bp pKINT-PsF plasmid was confirmed by sequencing with oligonucleotides 60-T7Pro (5’-TAATACGACTCACTATAGGG) and 131-TERM (5’-CCCTCAAGACCCGTTTAGA).

To construct the acceptor vector for the CP libraries, the PsF sequence in the pKINT-PsF was replaced by a DNA fragment encoding a part of the *pheS* gene. The fragment was amplified via PCR using oligonucleotides 582-Rv_KpnI_pheC (5’-CATGGTACCAAAATGGCAAGTAAAATAGCC) and 581-Fw_NheI_pheS (5’-ACTGCTAGCGCCATAGTGATTTGATTTGCC) and pMG214 as template. The *NheI/KpnI*-digested 1172 bp insert was ligated with the fragment from the accordingly digested pKINT-PsF to obtain the 5106 bp pKINT-0 vector. The sequence of the insert was confirmed by sequencing using oligonucleotides 60-T7Pro and 131-TERM.

The 4002 bp pKINT-CMi plasmid was constructed by assembling a SICLOPPS construct encoding the CMi (SGRLGHPL) peptide sequence. The required gene fragment was generated by assembly PCR. The PCR I fragment was obtained with pKINT-PsF as a template and oligonucleotides 60-T7Pro and 587-Rev-RLGH pep (5’-GTGACCGAGGCGACCGCTATTGTGGGCAAA). The PCR II fragment was generated using pKINT-PsF as a template and oligonucleotides 131-TERM and 586-Fw-RLGH pep (5’-CGCCTCGGTCACCCGCTGTCTGAGCCTT). The full-length PCR fragment was assembled using PCR fragments I and II and oligonucleotides 60-T7Pro and 131-TERM. The assembly PCR product was restriction digested with *Xba*I and *Hind*III and the resulting 563 bp insert was ligated with the 3439 bp fragment of the accordingly digested pKINT-0 vector. The insert was verified by sequencing using oligonucleotide 131-TERM.
4.4.4.4 pCyP1-based plasmids

Plasmid pCyP1 was based on the pKINT-PsF vector where the lacI and \( P_{T7} \) promoter regions were replaced by the \( prpBCDE \) promoter expression system consisting of the \( prpR \) repressor gene and the \( P_{prpB} \) promoter. The \( P_{prpB} \) promoter region upstream from the split intein gene was designed based on the corresponding region on plasmid pPro24\[[206]\] with few modifications. First, the \( NheI \) recognition site was replaced by the corresponding sequence from the wild-type \( E. coli \) \( prpBCDE \) promoter region (accession AM946981, region 316480...318066). Second, an \( NdeI \) restriction site was introduced overlapping the start codon of the intein gene. The RBS was left intact and its distance to the start codon of the intein was kept similar to the one in pPro24.

In order to remove the lacI gene and the \( P_{T7} \) promoter region, the pKINT-PsF vector was digested with \( BssSI \) and \( NdeI \), which also removed part of the ori sequence. To re-introduce the ori sequence, PCR I fragment was generated using pKIMP-PsF as a template and oligonucleotides 600–Fw Ori (5’-AGCGCAGGAGGAGCTT, carrying a \( BssSI \) site) and 60-RevOri (5’-CTGAGGCGCGCCTTCCGCATGAAGAC, carrying an \( Ascl \) site). A PCR II carrying the \( prpR \) repressor gene and the \( P_{prpB} \) promoter was generated using pPro18-kan\[[206]\] (purchased from AddGene) as a template and oligonucleotides 605-FwprpR (5’-AAGGCCGCCTCTTCAGCTTTCCGCGC, carrying an \( Ascl \) site) and 606–RevpCyp1 (5’-CCATATGGAATTCCTCCTCATTCTTGTATCAACTTGT, carrying the designed promoter region with an \( NdeI \) site).

The 2160 bp fragment resulting from the \( NdeI/BssSI \)-digested pKINT-PsF plasmid was ligated with two accordingly digested PCR fragments: the 313 bp fragment from the \( BssSI/Ascl \)-digested PCR I, and the 1841 bp fragment from the \( Ascl/NdeI \) digested PCR II. This resulted in the 4314 bp pCyP1-PsF plasmid, which was sequenced using 608-pCyP-Ori (5’-TGAGATACCTACAGCGTGA, for validation of the ori region), 609-pCyP-prpR (5’-GCACATCCACCGGAACAG, for validation of the \( prpR \) region), 610-pCyP-Ppro (5’-TTGTGGATGTAGGTCAACGC, for validation of the \( P_{prpB} \) region).
To construct the acceptor vector for cloning the CP libraries, the PsF sequence of plasmid pCyP1-PsF was exchanged with a stuffer fragment. For that, the 2236 bp fragment from the BssSI/Nhel-digested pCyP1-PsF was ligated with the 3182 bp fragment of accordingly digested pKINT-0 to generate the 5418 bp plasmid pCyP-0.

### 4.4.5 Identification of pKIMP-ACGQ-RBS6

The RBS region of the $P_{sal}$ promoter on plasmid pKIMP-ACGQ was randomized by cassette mutagenesis. Two PCR products (PCR I and PCR II) were generated by using pKIMP-ACGQ as a template. The PCR I fragment was obtained using primers 621-ACGQ (5'-TTGATGACTTGGTTAATAACG) and 570-Psal-MtCM, while the PCR II fragment was generated by using primers 622-ACGQ-RBS-Lib (5'-TTTATTAACAGTCATCAAANNKNNNK NNKNNKNKKAGCATGAACCTGGAAAATGC) and 620-ACGQ (5'-CTTGCGGGGCGCGCCAAAAAAAAG). The PCR I and PCR II fragments were assembled via PCR by using primers 620-ACGQ and 570-Psal-MtCM. The $Mlu$I/Ascl-digested 532 bp assembly PCR product was ligated with the accordingly digested 8036 bp pKIMP-ACGQ fragment.

The ligation product was transformed into electrocompetent KA12 cells and plated on M9c [M9 salts (6 mg/mL $\text{Na}_2\text{HPO}_4$, 3 mg/mL $\text{KH}_2\text{PO}_4$, 1 mg/mL $\text{NH}_4\text{Cl}$, and 0.5 mg/mL $\text{NaCl}$, pH 7) supplemented with 1 mM $\text{MgSO}_4$, 0.1 mM $\text{CaCl}_2$, 5 $\mu$g/mL thiamine-$\text{HCl}$, 5 $\mu$g/mL 4-hydroxybenzoic acid, 5 $\mu$g/mL 4-aminobenzoic acid, 1.6 $\mu$g/mL 2,3-dihydroxybenzoic acid, 20 $\mu$g/mL L-tryptophan, and 0.2 % (w/v) D- (+)-glucose as a carbon source] minimal medium 1.5 % agar plates containing 30 $\mu$g/mL chloramphenicol, 500 ng/mL tetracycline, and 0.5 mM salicylate. For determination of the library sizes a fraction of the cell culture was plated on supplemented M9c+FY (containing 20 $\mu$g/mL of each L-tyrosine and L-phenylalanine in addition to M9c medium provided with 30 $\mu$g/mL chloramphenicol). The clones that grew on these plates were sequenced to determine the quality and randomness for the library. The cells were incubated at 30°C and analyzed after two (M9c+FY) or four days (M9c+Tet$^{500 \text{ng/mL} + \text{Sal}^{0.5 \text{mM}}}$). The single colonies appearing on selective plates and non-selective plates were picked and drop-tested on M9c+Tet$^{500 \text{ng/mL} + \text{Sal}^{0.5 \text{mM}}}$, M9c+Tet$^{200 \text{ng/mL} + \text{Sal}^{0.1 \text{mM}}}$, M9c+Tet$^{200 \text{ng/mL} + \text{Sal}^{0.05 \text{mM}}}$, M9c+Tet$^{200 \text{ng/mL}}$.
Regulation, effectors, and evolvability of chorismate mutases from Corynebacteriales

+Sal$^{0.01}$ mM, M9c+Tet$^{200}$ ng/mL +Sal$^{0.005}$ mM, and M9c plates containing 30 μg/mL chloramphenicol. The clones that exhibited slower growth after three days at 30°C on the high stringency plates M9c+Tet$^{200}$ ng/mL +Sal$^{0.005}$ mM, and M9c were sequenced using oligonucleotide 570-Psal-MtCM. The most promising variants were re-transformed into KA12 and plated on M9c+Tet$^{10}$ ng/mL +Sal$^{0.05}$ mM and M9c+Tet$^0$ ng/mL +Sal$^0$ mM to identify the slowest grower after incubation for three days at 30°C. The clone that exhibited the slowest growth was named pKIMP-ACGQ-RBS6.

4.4.6 Construction of the cyclic peptide libraries

The five randomized gene libraries cpL6, cpL7, cpL8, cpL9, and cpLCMi were constructed by PCR amplification using pCyP1-CMi as a template and the sense primer 614-Ppro-pCyP1 combined with either 623-CPlib6 (5'-TTTCGGTACCAAAGCTCAGACAMNNMNNMNNACCGSAGTTGTTGGGCAATTGCACCATTGGCTAG for cpL6, S here and below represents C or G), 624-CPlib7 (5'-TTTCGGTACCAAAGCTCAGACAMNNMNNMNNMNNACCGSAGTTGTTGGGCAATTGCACCATTGGCTAG for cpL7), 628-CPlib8 (5'-TTTCGGTACCAAAGCTCAGACAMNNMNNMNNMNNMNNACCGSAGTTGTTGGGCAATTGCACCATTGGCTAG for cpL8), 626-CPlib9 (5'-TTTCGGTACCAAAGCTCAGACAMNNMNNMNNMNNMNNACCGSAGTTGTTGGGCAATTGCACCATTGGCTAG for cpL9), and 627-CPlibCM (5'-TTTCGGTACCAAAGCTCAGACAMNNMNNATGACCTAAACGACCGSAGTTGTTGGGCAATTGCACCATTGGCTAG for cpLCMi). An additional “zipper” PCR [207] to remove mismatches was conducted using the appropriate PCR product from the first round and the primers 614-Ppro-pCyP1 (5'-AAGGATGAGGAGGAATTCACA) and 617-Zipper (5'-TTTCGGTACCAAAGCTCAGAC). The zipper PCR products were restriction digested with NdeI and KpnI-HF and ligated into accordingly digested pCyP1-0.

The ligated library plasmids were transformed into electrocompetent KA12 and the entire sample was plated on M9c*+FY+Kan$^{30}$ agar plates. The components of M9c* medium were the same as for M9c medium except that 0.2 % glycerol was used as a
carbon source instead of glucose. The agar plates also contained 20 μg/mL of each L-tyrosine and L-phenylalanine, and 30 μg/mL kanamycin). A diluted sample of the transformed cells was plated on the same medium to determine the library size. After incubation at 30 °C for two days, the colonies were scraped off the agar and plasmid DNA was isolated.

The plasmid DNA was used to transform electrocompetent KA12/pKIMP-ACGQ-RBS6 cells. A portion of the electroporated cells was plated on M9c*+FY+Prp10 mM+Kan30+Cam30 (M9c* containing 20 μg/mL of each L-tyrosine and L-phenylalanine, 10 mM propionate, 30 μg/mL kanamycin, and 30 μg/mL chloramphenicol) agar plates to estimate the number of transformants and evaluate the quality of the libraries. Sequencing analysis of several clones from these plates revealed that up to 50 % of the clones in each library contained plasmid mixes. The remaining cell culture was inoculated in 100 mL of liquid M9c*+FY+Prp10 mM+Kan30+Cam30. After overnight growth at 30 °C the cells were collected, diluted to OD600=1 and a fraction of them was used for selection of CP activators (see section 4.4.7). The rest of the culture was supplemented with glycerol and stored at -80 °C for subsequent penicillin enrichment experiments (see section 4.4.8).

4.4.7 Selection for the cyclic peptide activators

The CP libraries in KA12/pKIMP-ACGQ-RBS6 prepared as described in section 4.4.6 were used for selection directly after overnight growth in liquid M9c*+FY+Prp10 mM+Kan30+Cam30 medium. The cells were washed in ice-cold M9 salts and a fraction was plated on M9c*+FY+Kan30+Cam30 to determine the number of viable cells, which would correspond to a complementation frequency of 100 %. For selection, the washed cells were plated on M9c*+ Prp10 mM+Kan30+Cam30+pFPhe40 μM (40 μM DL-para-fluoro-phenylalanine (pFPhe), a toxic analog of L-phenylalanine) or M9c*+Prp10 mM+Kan30+Cam30+pFPhe100 μM (100 μM pFPhe).

The colonies that appeared after a 3-4 day incubation at 30 °C were picked and resuspended in M9 salts in a 96-well plate. Approximately 5 μL of cell suspension was drop-spotted on M9c*+Prp10 mM+Kan30+Cam30+pFPhe100 μM agar plates and incubated at 30 °C for three days.
The 30 best growers from each library were inoculated into 100 μL liquid M9c*+FY+Kan30+Cam30 medium in a 96-well plate in order to prevent the fast growers to outnumber the slow ones. After overnight growth at 30 °C the cultures of the 30 clones were pooled for each library separately and the plasmid DNA was isolated. These plasmid pools were transformed into electrocompetent KA12/ pKIMP-ACGQ-RBS6 cells and plated on M9c*+Prp10 mM +Kan30+Cam30+pFPhe100 μM agar plates. The plasmid DNA of the clones that grew at 30 °C after a three-day incubation was sequenced using oligonucleotide 610-pCyP-Ppro (5’-TTGTGGATGTAGGTCACC).

### 4.4.8 Penicillin enrichment for the cyclic peptide inhibitors

The CP libraries in KA12/pKIMP-ACGQ-RBS6 prepared as described in section 4.4.6 and stored at -80 °C after overnight growth in liquid M9c*+FY+Prp10 mM +Kan30+Cam30 medium were thawed and washed in M9 salts. A fraction of the washed cells was plated on M9c*+FY+Kan30+Cam30 to determine the number of viable cells. The rest of the culture was subjected to starvation in a 2 L Erlenmeyer flask containing 300 mL minimal medium M9c*+Prp10 mM +Kan30+Cam30+20 % sucrose to starve the cells. The cultures were started at OD₆₀₀ < 0.02 and grown by shaking (180-200 rpm) at 30 °C until the OD₆₀₀ increased 2 to 4 times (approximately after 16 hours), but did not exceed OD₆₀₀ = 0.1.

After this growth phase in minimal medium, a fraction of the cells was plated on M9c*+FY+Kan30+Cam30 to determine the number of viable cells prior to penicillin treatment. 2000 U/mL penicillin G (1 U is equivalent to 0.6 μg of penicillin G, sodium salt) was added to the main 300 mL cell culture and the treatment was continued at slow (50-75 rpm) shaking at 30 °C for 4 hours. After the treatment a fraction of the cell culture was collected, washed in M9 salts and plated on M9c*+FY+Kan30+Cam30 agar plates to determine the number of viable cells after the penicillin treatment. The comparison of the number of viable cells before and after penicillin treatment allowed the determination of cell survival rates.

After the penicillin treatment phase the 300 mL cell culture was spun (10 min, 4000 rpm (Sorvall rotor SA600) at 4 °C), the supernatant was filtered through 0.2 mm filters, and the pellet was collected. Both
the cell pellet and the filter were washed extensively with M9 salts and
the resulting cell suspensions were added to 100 mL of liquid
M9c*+FY+Kan30+Cam30 medium.

After overnight growth at 30 °C, the cells were collected, washed
and subjected to the next cycle of starvation and penicillin treatment. A
total of three iterative cycles of starvation, penicillin treatment, and
growth in supplemented medium were performed per one penicillin
enrichment. After the third cycle, the cells that were grown in
M9c*+FY+Prp10 mM +Kan30+Cam30 medium were collected and their
plasmid DNA was isolated. Then, the plasmid DNA was transformed
into fresh electrocompetent KA12/pKIMP-ACGQ-RBS6 or
KA12/pKIMP-ACGQ cells for the next penicillin enrichment.

4.4.9 Intein production and purification

The plasmids pKINT-CMi, pKINT-PsF, and pKINT-0 were
transformed into the E. coli strain KA13 and the pCyP1-based plasmids
were transformed into the E. coli strain KA12. Production of C-
terminally His6-tagged intein variants was carried out in LB medium
containing 30 µg/mL kanamycin. The cultures were grown at 37°C
until they reached an OD600 of 0.3 – 0.6 and the induction of the gene
expression was initiated with 0.05 – 0.1 mM IPTG (for the variants on
pKINT-based plasmids) or 5 - 10 mM propionate (for the variants on
pCyP1-based plasmids). After overnight induction and growth at 37°C,
the cells were collected and the crude lysate was obtained following
published protocols[208] (omitting the RNaseA and DNase I treatment).

The crude cell lysate was supplemented with 10 mM imidazole
and applied onto a column with equilibrated Ni-NTA agarose beads
(Sigma). Intein variants were either eluted with 250 mM imidazole
immediately, or after incubation on the column for 16 h at room
temperature to allow protein splicing reaction to occur prior to elution.
Eluted proteins were analyzed by SDS-PAGE using the PhastSystem
(20 % homogeneous gels, GE Healthcare).

4.4.10 Cyclic peptide isolation and detection by LC-MS

The CPs were directly extracted from the crude cell lysates or
elution fractions after Ni-NTA purification using n-butanol following
the published procedures.[175, 207] Due to poor phase separation at room
Regulation, effectors, and evolvability of chorismate mutases from *Corynebacteriales*

temperature, the samples were centrifuged for 10 min at 7000 rpm at 4°C. The extracts were dried under high vacuum at 40 °C and resuspended in 0.1 % formic acid. High-pressure liquid chromatography (HPLC) coupled to mass spectrometry (LC-MS) was performed on an HPLC (Thermo Separation Products) system connected to a photodiode array detector (UV6000LB, TSP) and to an ion-trap mass-spectrometry detector (LCQdeca, Finnigan). Linear gradients of 5 - 85 % acetonirile in H₂O/0.1 % formic acid (0.2 ml/min) were applied on a Waters Atlantis T3-3u 2 x 10 mm column to elute the peptides.

4.4.11 Enzymatic assay

The CM assay was performed as described in section 2.4.3 in BTP+ buffer (50 mM BTP, pH 7.5, additionally containing 1 mM TCEP, 0.2 mM PEP, and 0.1 mM MnCl₂). The chemically synthesized CPs from commercial sources were dissolved in dimethylformamide (DMF) prior to adding to the assay solution.
5 DIRECTED EVOLUTION OF MtCM

Parts of this chapter will soon appear in print:

5.1 Introduction

As discussed in section 1.3, directed evolution is a powerful tool for elucidation of factors needed for catalytic perfection. This method is in particular useful when identification of residues contributing to catalysis is not possible by rational approaches due to our limited understanding of protein structure-function relations.

Figure 5.1 Directed evolution of MtCM. (A) Sequence alignment of the wild-type MtCM, the His\textsubscript{6}-tagged construct, and the mutants 3p3 and 3p5 obtained after the first two rounds of evolution\cite{26} (active site residues are emphasized in red and mutations are in blue). The secondary structure elements are indicated by H (for helix) and L (for loop) according to the crystal structure of wild-type MtCM. (B) The structures of malate-bound (yellow, PDB ID: 2VKL) and MtDS-bound activated MtCM (green, PDB ID: 2W1A) are superimposed. The helices of one subunit are labelled H1, H2, and H3, and the regions that changing conformation (loop between H1 and H2 helices, and C-terminus) are indicated by arrows.

MtCM is a mediocre enzyme on its own and needs the interaction partner MtDS to attain the catalytic efficiency typical for natural CMs.\cite{48} In this study, we set out to investigate whether high catalytic activity of MtCM can be achieved by mechanisms other than complex formation with a partner enzyme.

Directed evolution subsequently requires biochemical characterization of many mutants; therefore, an N-terminally His\textsubscript{6}-
tagged version of the wild-type MtCM (H₆-MtCM) had been constructed to facilitate easy purification of evolved variants[26] (Figure 5.1A). This wild-type variant showed a roughly 2-fold reduced catalytic efficiency and could not complement the growth of the CM deficient E. coli strain KA12/pKIMP-UAUC (see section 1.3.1) under selective conditions.[26]

During the first two evolutionary cycles, the flexible loop between helices H1 and H2 as well as the C-terminus were targeted by cassette mutagenesis. Crystal structures of wild-type MtCM revealed that these regions change their conformation upon complex formation with MtDS (Figure 5.1B) and play an important role in positioning active site residues in a catalytically competent conformation.[48] Also, directed evolution showed that C-terminal residues Gly84, Arg85, and Gly86 are crucial for the catalytic machinery, whereas the very last four residues Arg87, Leu88, Gly89, and His90 are important for the activation mechanism by MtDS.[126]

The best variant obtained after the first round of evolution, when the H1-H2 loop was targeted by cassette mutagenesis, was H6-PHS08-3p3 (3p3). It acquired Thr52Pro and Val55Asp exchanges (Figure 5.1) and was 23-fold more active (k_{cat}/K_m = 2.0 \times 10^4 \, M^{-1} \, s^{-1}) than the starting point, the His₆-tagged enzyme (k_{cat}/K_m = 0.09 \times 10^4 \, M^{-1} \, s^{-1}).[26] When the last four residues of the 3p3 variant were targeted during the second round of evolution, the best obtained catalyst pHS10-3p5 (3p5) had Arg87Pro, Leu88Asp, Gly89Ala, and His90Met replacements (Figure 5.1) and could catalyze the reaction four times faster (k_{cat}/K_m = 8.3 \times 10^4 \, M^{-1} \, s^{-1}) than 3p3.[26]

Despite a total of a 92-fold increase in catalytic efficiency, the variant 3p5 was still 3 times less active than the MtDS-bound wild-type MtCM (k_{cat}/K_m = 24 \times 10^4 \, M^{-1} \, s^{-1}). This indicated, that the full catalytic capacity of a MtCM had not been reached yet, leaving room for further improvement.

In this chapter, we perform additional two cycles of evolution to increase the catalytic activity of variant 3p5. However, along with the increasing activity of mutant CMs, suitable dynamic range of KA12/pKIMP-UAUC-based selection system was lost. The most common approaches to increase the stringency of a selection system is regulation of the intracellular concentration of the target enzyme. This
can be accomplished (i) at the transcriptional level by using low copy number plasmids and inducible promoters[209] (ii) translational level by using riboswitches and modified ribosome binding sites (RBS)[210, 211, 212], and (ii) post-translational level by using protein degradation tags[213, 214] or incorporation unnatural amino acids (E.g. DL-para-fluoro-phenylalanine (pFPhe)[126]) that makes them susceptible to proteases. Finally, catalytic efficiency of an enzyme can be directly reduced the by introducing catalytic lesions, which was first implemented by Knowles and coworkers to study TIM by directed evolution.[215] This and several other studies showed, that perturbation-compensation strategy could be successful in obtaining catalytically more active and stable proteins.[216, 217, 218, 219] Therefore, we employed structural destabilization to temporarily bring down the catalytic activity of the parent mutants. This allowed for differentiation of active and highly active destabilized clones during the selection process, which subsequently were re-stabilized.

5.2 Results and Discussion

5.2.1 Directed evolution via inter-subunit destabilization

The rather high catalytic activity of 3p5 granted the bacterial growth of the transformant in KA12/pKIMP-UAUC under the highest selection stringencies available.[26] Therefore, the selection stringency was augmented further using an inter-subunit destabilization strategy, which introduces structural (and indirectly catalytic) lesions at the interface of the two 3p5 subunits.

The side chains of Leu24 and Leu31 point towards the interface of the H1 helices of the two subunits and might play a role in dimer stabilization via leucine-zipper-like interactions (Figure 5.2A). It was assumed that replacing these residues with large, polar, or charged residues would destabilize the dimer lowering the catalytic efficiency of 3p5 and thus raise the bar for successful complementation during in vivo selection.
Chapter 5: Directed evolution of MtCM

5.2.1.1 Generation the inter-subunit destabilized 3p5 variants

Four gene libraries were constructed to randomize Leu24 (library L1), Leu31 (library L2), or both Leu24 and Leu31 simultaneously (library L3) via the degenerate codon NNK (N = A/C/G/T, K = G/T). To prevent incorporation of the stop codon the RNK (R = A/G) codon was used to simultaneously randomize positions Leu24 and Leu31 for the L4 library. The RNK codon also reduced the amino acid variety that could be incorporated at its position and enriched it with charged residues Arg, Lys, Asp, and Glu.

The libraries L1, L2, L3, and L4 were individually transformed into KA12/pKIMP-UAUC to give library sizes of 5.6×10^3, 1.3×10^4, 8.4×10^3, and 4.9×10^3, respectively. The corresponding MtCM genes were under the control of the \( P_{tet} \) promoter, allowing for induction of gene expression with tetracycline (Tet). The transformed cells were grown on weakly selective M9c+F+Tet\(^{500}\) ng/mL agar plates (containing Phe and 500 ng/mL of the inducer Tet) to remove toxic or inactive MtCM variants. A complementation rate of 64 %, 52 %, 32 %, and 13 % was observed for libraries L1, L2, L3, and L4, respectively. This suggested that substitution at either Leu24 or Leu31 (L1 and L2 libraries) did not impair the variant 3p5 as much as substitutions at both positions.

Figure 5.2 Destabilization of variant 3p5. (A) Leu21 (red spheres) and Leu31 (green spheres) are localized at the interface of the H1 helices of the two intertwined subunits (shown in cyan and wheat). (B) Sequencing results (only translated codons from Glu19 to Ala36) of the weak complementers from libraries L1, L2, L3, and L4. Deviations from the parent 3p5 sequence in each variant are highlighted in red.

<table>
<thead>
<tr>
<th>Library</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>3p5</td>
<td>EEIDR LDAEILAL VKRRA</td>
</tr>
<tr>
<td>L1.13</td>
<td>EEIDR EDAEILAL VKRRA</td>
</tr>
<tr>
<td>L1.15</td>
<td>EEIDR DDAEILAL VKRRA</td>
</tr>
<tr>
<td>L2.9</td>
<td>EEIDR LDAEILAP VKRRA</td>
</tr>
<tr>
<td>L3.6</td>
<td>EEIDR WDAEILAT VKRRA</td>
</tr>
<tr>
<td>L3.7</td>
<td>EEIDR RDAEILAR VKRRA</td>
</tr>
<tr>
<td>L3.8</td>
<td>EEIDR QDAEILAG VKRRA</td>
</tr>
<tr>
<td>L4.4</td>
<td>EEIDR GDAEILAR VKRRA</td>
</tr>
<tr>
<td>L4.6</td>
<td>EEIDR TDAEILAT VKRRA</td>
</tr>
<tr>
<td>L4.7</td>
<td>EEIDR TDAEILAS VKRRA</td>
</tr>
</tbody>
</table>
simultaneously (L3 and L4 libraries). Most inactive clones were contained in library L4, which was enriched for charged amino acids. This was expected due to the location of the side chains at positions 24 and 31 buried in a hydrophobic core.

The growing clones were tested on M9c+Tet\(^{80}\) ng/mL, M9c+Tet\(^{40}\) ng/mL, M9c, and M9c+pFPhe\(^{40}\) \(\mu\)M (containing 40\(\mu\)M of the toxic pFPhe) agar plates to find the weakest complementers. The clones that showed no growth on the highest stringency plates (M9c and M9c+pFPhe\(^{40}\) \(\mu\)M) were used for further characterization. Out of twelve tested clones, only four from the L1 and two from the L2 library exhibited no complementation on the highest stringency plates.

Sequence analysis revealed that a substitution by the large charged Asp or Glu at position Leu24 and a helix-disrupting Pro at position Leu31 could sufficiently reduce CM activity of 3p5 \textit{in vivo} (Figure 5.2B). However, when Leu24 and Leu31 were mutagenized simultaneously, destabilization was achieved more readily: six of the twelve tested clones each from libraries L3 and L4 showed no growth on high-stringency plates. Sequencing analysis of the weakest complementers showed that intra-helical Leu24 and Leu31 had been replaced by charged (Glu, Gln, Arg), polar (Thr, Ser), bulky (Trp), or small (Gly) amino acids (Figure 5.2B).

The plasmids of several of these clones were transformed into the KA12/pT7POLTS strain\(^{[126]}\) for protein production. Kinetic and thermal denaturation assays showed that weak \textit{in vivo} complementation correlated well with the loss of thermal stability and catalytic efficiency (Table 5.1). The latter was contributed by decreases in \(k_{cat}\) and increases in \(K_m\) as compared to the parent mutant 3p5. Since residues from both MtCM subunits contribute to the active site\(^{[48]}\) it is likely that impairment of the dimer formation equilibrium or a general misalignment of the active site residues due to the destabilizing mutations cause poorer substrate binding and lower catalytic rate constant.

The substitutions by the small polar residues Ser and Thr had the weakest effect on catalytic activity, leaving unchanged \(k_{cat}\) and only slightly poorer substrate binding (variant L4.7, Table 5.1). However, when the bulky Trp was introduced at position 24, together with the polar Thr at position 31, the dominant effect was reduction in \(k_{cat}\).
whereas substrate binding was hardly affected (variant L3.6). Moreover, substitution Leu24Trp caused a significant decrease in thermal stability ($T_m$ of variant L3-6 was reduced by 39°C). It was also evident, that introduction of Gly at any of the two positions caused an activity loss presumably due to its tendency to destabilize α-helical segment, which is also reflected by a significantly reduced $T_m$ of variant L3-8 (Table 5.1).

### Table 5.1. Effects of destabilizing mutations on the kinetic parameters and melting temperature of destabilized 3p5 variants.

<table>
<thead>
<tr>
<th>His-tagged protein</th>
<th>Destabilizing mutations</th>
<th>$k_{cat}^a$ (s$^{-1}$)</th>
<th>$K_m^a$ (μM)</th>
<th>$k_{cat}/K_m^a \times 10^4$ (M$^{-1}$ s$^{-1}$)</th>
<th>$T_m^b$ (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3p5</td>
<td>-</td>
<td>31 ± 2</td>
<td>370 ± 21</td>
<td>8.3 ± 1.2</td>
<td>84 ± 4</td>
</tr>
<tr>
<td>L4.7</td>
<td>L24T, L31S</td>
<td>28</td>
<td>550</td>
<td>5.0</td>
<td>nd</td>
</tr>
<tr>
<td>L3.6</td>
<td>L24W, L31T</td>
<td>16</td>
<td>380</td>
<td>4.0</td>
<td>49</td>
</tr>
<tr>
<td>L3.8</td>
<td>L24Q, L31G</td>
<td>22</td>
<td>740</td>
<td>3.0</td>
<td>39</td>
</tr>
<tr>
<td>L1.13</td>
<td>L24E</td>
<td>10</td>
<td>880</td>
<td>1.0</td>
<td>nd</td>
</tr>
<tr>
<td>L3.7</td>
<td>L24R, L31R</td>
<td>ND</td>
<td>&gt;1600</td>
<td>0.3</td>
<td>35</td>
</tr>
<tr>
<td>L4.4</td>
<td>L24G, L31R</td>
<td>ND</td>
<td>&gt;1800</td>
<td>0.05</td>
<td>nd</td>
</tr>
</tbody>
</table>

$^a$CM activity was determined in 50 mM potassium phosphate buffer, pH 7.5, containing 0.1 mg/mL BSA. Disappearance of chorismate was monitored at 310 nm at 30°C varying concentrations between 100 and 2000 μM. Kinetic parameters were derived by fitting specific initial velocities to the Michaelis-Menten equation.

$^b$Melting temperature was determined by CD spectroscopy following the procedure describes in section 5.4.9

nd: not determined

ND: not determined, since substrate saturation could not be achieved using up to 2 mM chorismate. Measuring at higher substrate concentrations was not possible due to the high chorismate absorption of the substrate at 310 nm and the relatively high background reaction.

Finally, the most significant effects on catalytic activity were observed when charged residues (Glu or Arg) were introduced at the dimer interface. This substitution had the strongest effect when occurring at position 24. All of the variants found to complement the weakest when Leu24 alone was mutated, contained a charged residue at this position (Figure 5.2B). This could be due to the spatial vicinity of Arg34 and Arg35 in the neighboring subunit, which might be forced to engage in alternative hydrogen bonding due to Glu or Arg at position 24. Such rearrangements could have am impact on dimer formation or
optimal positioning of the active site Arg35 (see Figure 5.9A below). On the other hand, an Arg at position 31 might form hydrogen bonds with Glu27 of the other subunit and thereby alter dimer formation. It is, therefore, not surprising that the strongest reductions in the catalytic activity were observed when Arg was introduced at both L24 and L31 positions simultaneously (variant L3-7, 30-fold lower that 3p5), or in combination with the helix destabilizing Gly (variant L4-4, 180-fold lower) (Table 5.1).

Taken together, the mutations at the intra-helical positions Leu24 and Leu31 successfully diminished the catalytic activity and thermal stability of variant 3p5. This resulted in weaker complementation on high-stringency plates and allowed for further directed evolution experiments. The range of catalytic activities exhibited by the variants L1-13, L3-6, L3-7, L3-8, L4-4, and L4-7, which were subsequently chosen as parent mutants, permitted variation of selection stringencies to obtain optimal selection conditions.

5.2.1.2 Directed evolution of destabilized 3p5 variants

Two mutagenesis strategies – DNA shuffling\textsuperscript{[112]} and epPCR\textsuperscript{[111]} – were employed to mutagenize the genes of the inter-subunit destabilized 3p5 variants. DNA shuffling allowed testing if the previously identified beneficial mutations Gly43Val, Thr52Pro, Val55Asp, Val62Ile, Asp75Tyr, R82Gln, Arg87Pro, Leu88Asp, Gly89Ala, and His90Met\textsuperscript{[26]} can augment catalytic function when introduced in various combinations, whereas epPCR allowed sampling of a new set of mutations. To prevent the reversion of the destabilizing mutations at positions 24 and 31 back to the wild-type Leu, only the downstream region of the gene was subjected to mutagenesis.

The portion of the accordingly mutagenized gene region was cloned into a pKNTTET-0-based acceptor vector\textsuperscript{[26, 126]} already containing the 5’ portion of the gene encoding one of the destabilized variants L1-13, L3-6, L3-7, L3-8, L4-4 or L4-7, controlled by the tandem $P_{tet}/P_{T7}$ promoter. After electroporation of individual libraries into the selection strain KA12/pKIMP-UAUC, twelve libraries of $1\times10^5$ – $8\times10^5$ independent clones were obtained. Complementation rates on selective M9c+F+Tet\textsuperscript{80 ng/mL}, M9c+F+Tet\textsuperscript{40 ng/mL}, M9c+Tet\textsuperscript{80 ng/mL}, M9c+Tet\textsuperscript{40 ng/mL} and M9c agar plates, were < 2 % for all tested libraries.
Chapter 5: Directed evolution of MtCM

Figure 5.3 Frequency of mutations in selected clones from inter-subunit destabilization libraries. Sequencing results of selected clones from epPCR (ep) and DNA shuffling (sh) libraries using destabilized variants L1-13, L3-6, L3-7, L3-8, L4-4 or L4-7 are shown. A total of 26, 21, 24, 12, 9 and 6 clones from epPCR libraries epL1.13, epL4.4, epL4.7, epL3.6, epL3.7, and epL3.8, respectively and 23, 9, 26, 11, 10, and 11 from sh libraries shL1.13, shL4.4, shL4.7, shL3.6, shL3.7, and shL3.8 with correct sequences were analysed (artefacts were omitted). The frequency of appearance of a specific mutation is color coded as purple (<10 %), blue (10-30 %), orange (30-60 %) and red (60-100 %). Only the sequence region C-terminal from Leu31, which has been subjected to mutagenesis, is shown. Mutations that were present in the parent mutant 3p5 are highlighted in red for each library template.
Regulation, effectors, and evolvability of chorismate mutases from *Corynebacteriales*

Figure 5.4 (legend next page)
Figure 5.4 Catalytic efficiency of evolvedMtCM variants from the inter-subunit destabilization libraries. The activity of the destabilized parent variants L1-13, L4-4, L4-7 L3-6, L3-7, or L3-8 is colored in darker red, green, blue, purple, orange and brown, respectively. The evolved variants from each library are presented in the same color as their parent. The boxed bars indicate the activity of the evolved mutants after removal of the destabilizing mutations at positions 24 and 31. The best new variant (re4.7s11) is indicated by asterisk and the activity of the starting variant 3p5 ($k_{cat}/K_m = 8.3 \times 10^3 \text{ M}^{-1} \text{s}^{-1}$) is shown in black. For each variant only new mutations that occur in addition to the ones present in 3p5 are listed. For the individual kinetic parameters, see Tables A.I and A.II in the Appendix.

Sequencing analysis of the selected mutants revealed several regions where most of the new mutations accumulated (Figure 5.3). The most heavily mutagenized was the segment between Glu37 and Ala45 that comprises the C-terminal part of the helix H1. This region is in the immediate vicinity of the active site and adjacent to the active site residue Arg46. It is possible, that mutations in this region may significantly influence the positioning of Arg46, which probably plays a central role in CM catalysis. Among the most frequently found changes was Gly43Val substitution, which might have a helix-stabilizing effect. Other frequently seen replacements were Lys44Glu, Asp72Glu and Gly84Cys.

Importantly, the Thr52Pro and Val55Asp mutations, which were introduced during the first round of directed evolution, were never replaced, indicating their importance in efficient catalysis of autonomous MtCM. On the other hand, several of the C-terminal substitutions, which were introduced during the second round, were affected in individual selected variants (Figure 5.3).

Several evolved destabilized mutants carrying the most frequent mutations found in the epPCR and DNA shuffling libraries were produced and characterized in vitro. The kinetic analysis revealed either similar or increased catalytic efficiencies ($k_{cat}/K_m$) when compared to the parent proteins (Figure 5.4). This indicated that the newly acquired mutations typically improved the catalytic efficiency of the destabilized variants L1-13, L3-6, L3-7, L3-8, L4-4, and L4-7.

However, the accumulation of mutations that are only beneficial in the context of the destabilized protein is a substantial drawback of the destabilization-compensation strategy. Therefore, the final step of
the 3p5 evolution involved reconstitution of the wild-type Leu residues at the destabilization sites 24 and 31 in the selected evolved variants.

Kinetic measurements of the re-stabilized variants revealed that most of the new mutations actually improved the catalytic efficiency of 3p5 (Figure 5.4). The best variant re4.7s11, which had the additional replacements Val62Ile and Asp72Val, exhibited more than a 3-fold improved catalytic efficiency compared to 3p5 and was chosen as a parent variant for the next round of directed evolution.

### 5.2.2 Directed evolution via terminal truncation

As for the variant 3p5, the selection stringency needed to be increased to evolve re4.7s11. For this, a terminal truncation approach was implemented to restrict its catalytic activity. By targeting the crippling mutations to the termini, the regions N-terminal to the positions 24 and 31 became available for mutagenesis.

Protein termini had often been considered to be non-essential for protein structure and stability since they differ in length and sequence pattern among homologous proteins and are often poorly defined in crystal structures. However, examples where removal of terminal residues caused severe effects including folding and structural perturbations or even turning the protein into a molten globule[220] speak for more complex roles of these parts of a protein.

Several studies have shown, that iterative truncation-optimization-re-elongation cycles can be successfully implemented for directed evolution of thermostable β-lactamases.[221, 222, 223] We hypothesized that MtCM might also be sensitive to C-terminal truncation since C-terminal residues play an important role in its activation by MtDS.[48, 126] Moreover, mutations at the last four positions increased the catalytic efficiency of the 3p3 variant by 4.3-fold (see section 5.1), indicating their importance in catalysis.

However, due to the already very high catalytic activity of variant re4.7s11, truncation at the C-terminus might not be sufficient to cut back catalysis to selectable level. Therefore, deletion at the N-terminus was also considered as means to partially deactivate the re4.7s11 variant. In fact, the observation that exchanging the first four N-terminal residues with the His6-tag reduces the catalytic efficiency of
the wild-type enzyme nearly 2-fold (see below, Table 5.3), suggested that this region might also be sensitive for perturbations.

In addition to the N- and C-terminal truncations, mutagenesis of residues that are close to the active site, as well as the addition of the protein degradation tag RepA[224] were employed to further reduce intracellular CM activity and to facilitate the directed evolution of re4.7s11.

5.2.2.1 Generation of destabilized re4.7s11 variants

The LdNdC library was created to permit simultaneous N- and C-terminal mutagenesis of variant re4.7s11. The N-terminus was randomized in a way that allowed deletion by two residues at a time up to the active site residue Arg18 (Figure 5.5). It was assumed that truncation will introduce higher flexibility in the α-helix and, therefore, preclude a catalytically proficient positioning of Arg18. The three first N-terminal residues Met-Leu-Glu directly following after the His6-tag and the Ser-Ser-Gly linker were not deleted to retain the purification tag and the XhoI site for in-frame cloning.

\[
\begin{align*}
&\text{MHHHHHHSSGMLESQPVPEIDTLREE...} & \text{...LLRLGRGPDPAM} \\
&\text{MHHHHHHSSGMLE--PVPEIDTLREE...} & \text{...LLRLGRGPDA*} \\
&\text{MHHHHHHSSGMLE-----PEIDTLREE...} & \text{...LLRLGRGPD*} \\
&\text{MHHHHHHSSGMLE-------IDTLREE...} & \text{...LLRLGRGP*} \\
&\text{MHHHHHHSSGMLE----------TLREE...} & \text{...LLRLGRG*} \\
&\text{MHHHHHHSSGMLE----------REE...} & \text{...LLRLGR*} \\
\end{align*}
\]

Figure 5.5 Construction of re4.7s11 truncation libraries. The N- and C-termini of His6-tagged MtCM are shown. To destabilize the highly active re4.7s11 variant (i) deletions at the N-terminus were allowed every two residues up to Arg18, (ii) truncations at the C-terminus were allowed every residue up to Arg85 and residues (iii) Leu81 or (iv) Arg85 were randomized. For the LdNdC library (i) and (ii) were combined. For the libraries LdNL81X and LdNR85X, (iii) and (iv), respectively, were combined with the five residue truncation up to Arg 85 at the C-terminus and the N-terminal truncations according to strategy (i).

The C-terminus was randomized allowing truncations of one residue at a time up to Arg85 (Figure 5.5). Together with Gly84, Arg85
had been preserved during previous directed evolution experiments indicating their importance in catalysis.[126] Therefore, truncation of the C-terminus was designed to avoid removal of these residues.

The LdNdC library clones were plated on weakly selective M9c+F+Tet500 ng/mL plates to isolate catalytically impaired but still active clones. The complementation rate was > 80 % suhhesting that most truncations were still compatible with significant CM activity in vivo. 108 colonies that allowed sampling the entire library including 36 truncation combinations from then M9c+F+Tet500 ng/mL plates were tested on selective plates of various stringencies. Less than 20 % of the 108 clones did not (or only weakly) complemented on the highest stringency plates M9c, M9c+pFPhes40 μM and M9c+pFPhes100 μM. Five of the eighteen sequenced weak completerers had extensions at the C-terminus due to unplanned nucleotide deletions and were excluded from further analysis. Out of the remaining thirteen clones only four sequences were unique. They exhibited the maximally allowed truncation of five residues at the C-terminus and various deletions between the His6-tag and Arg18 at the N-terminus.

The clones dNdCs2, dNdCs3 and dNdCs4 that complemented weakly on the highest stringency plates (M9c+pFPhes40 μM and M9c+pFPhes100 μM) had a truncation of five amino acids at the C-terminus and deletions of 0, 2, and 4 amino acids at the N-terminus, respectively. Only clone dNdCs1, which had the maximum deletion at the N-terminus and the maximum truncation at the C-terminus, did not complement on high-stringency plates and seemed to have a sufficiently reduced intracellular CM activity for subsequent evolution experiments.

However, large deletions and truncations also removed sequence regions that could potentially accommodate beneficial mutations during the further directed evolution. Therefore, the LdNR85X and LdNL81X libraries were designed with the goal to identify clones carrying shorter deletions at the N-terminus but still exhibiting sufficient destabilization. To achieve this, the random stepwise truncations at the N-terminus were introduced as described for the LdNdC library (Figure 5.5). At the C-terminus, a fixed truncation up to Arg85 was applied, as it still proved to be compatible with significant in vivo activity of the variants from the LdNdC library. In addition,
positions Arg85 and Leu81 were randomized resulting in libraries LdNR85X and LdNL81X, respectively (Figure 5.5). Arg85 had been shown to be catalytically important\cite{126} whereas Leu81 is a conserved residue that points toward the active site, suggesting its potential contribution to catalysis. We assumed that randomizing Arg85 and Leu81 would disrupt efficient binding of the substrate and together with truncations at the N- and C- termini would reduce the catalytic activity of re4.7s11.

The libraries LdNR85X and LdNL81X in KA12/pKIMP-UAUC were plated on weakly selective M9c+F+Tet\(^{500}\) ng/mL agar plates. The complementation rate of library LdNL81X on these plates was only 50 \%, whereas nearly 100 \% of the clones from library LdNR85X complemented. However, testing 94 slow growers from each library on higher stringency plates (M9c, M9c+pFPhe\(^{40}\) \(\mu\)M, and M9c+pFPhe\(^{100}\) \(\mu\)M) showed that 7 \% and 38 \% of the clones from libraries LdNL81X and LdNR85X, respectively, did not complement anymore under these more challenging conditions.

A low percentage of weak complementers from library LdNL81X would indicate that remaining 93 \% of the selected mutants were still highly active. Out of nine sequenced weak complementers, only five sequences were unique. Mutants without a deletion at the N-terminus together with either a Leu81Thr substitution or the unchanged Leu81 were observed several times. Surprisingly, the maximum deletion at the N-terminus found in selected weak complementers was just two amino acids (Ser8Gln9) combined with the conservative Leu81Ala or Leu81Val substitutions.

A 50 \% complementation frequency of library LdNL81X clones on M9c+F+Tet\(^{500}\) ng/mL plates as well as the fact that most of the so obtained clones could complement at higher stringencies indicated that either fully active or completely inactive clones were present in this library. The pre-selection eliminated the latter clones; therefore, variants with appropriately low \textit{in vivo} activity, useful for further directed evolution, could not be obtained from the LdNL81X library.

In contrast, 38 \% of the preselected clones from library LdNR85X could not complement at higher stringencies, indicating that destabilized but still moderately active clones were more frequent
than in library LdNL81X. Sequencing results of weak complementers from LdNR85X revealed that mutations at Arg85 were sufficient to reduce *in vivo* complementation even when no or only few residues were deleted at the N-terminus. This indicated that randomization of Arg85 in combination with C-terminal truncation of five residues reduces the catalytic efficiency of re4.7.11 more efficiently than truncations alone in the previously described LdNdC library. Interestingly, variant dNR85Xu15 that was identified on the non-selective plates M9c+FY lacked the Arg85 residue making it the shortest characterized MtCM. Even though it could not complement under high-stringency conditions, it complemented well on the weakly selective M9c+F+Tet\(^{500}\) ng/mL plates, making it interesting for further investigation.

### Table 5.2 The effects of truncations on the kinetic parameters and the melting temperature of truncated re4.7s11 variants.

<table>
<thead>
<tr>
<th>His-tagged protein&lt;sup&gt;c&lt;/sup&gt;</th>
<th>N-terminus&lt;sup&gt;d&lt;/sup&gt;</th>
<th>C-terminus&lt;sup&gt;e&lt;/sup&gt;</th>
<th>(k_{\text{cat}}^a) (s(^{-1}))</th>
<th>(K_m^a) (μM)</th>
<th>(k_{\text{cat}}/K_m^a) ×10(^4) (M(^{-1}) s(^{-1}))</th>
<th>(T_m^b) (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>re4.7s11</td>
<td>5MLESPVPEIDTLR&lt;sup&gt;18&lt;/sup&gt;</td>
<td>85RGPDAM&lt;sup&gt;90&lt;/sup&gt;</td>
<td>23</td>
<td>72</td>
<td>32</td>
<td>&gt;90</td>
</tr>
<tr>
<td>dNR85Xu15</td>
<td>5MLE----------R&lt;sup&gt;18&lt;/sup&gt;</td>
<td>85----------90</td>
<td>ND</td>
<td>ND</td>
<td>0.002</td>
<td>77</td>
</tr>
<tr>
<td>dNdCs1</td>
<td>5MLE----------R&lt;sup&gt;18&lt;/sup&gt;</td>
<td>85R----------90</td>
<td>ND</td>
<td>ND</td>
<td>0.01</td>
<td>74</td>
</tr>
<tr>
<td>dNdCs3</td>
<td>5MLE--PVPEIDTLR&lt;sup&gt;18&lt;/sup&gt;</td>
<td>85R----------90</td>
<td>ND</td>
<td>&gt;1300</td>
<td>0.4</td>
<td>nd</td>
</tr>
<tr>
<td>dNdCs4</td>
<td>5MLE--PEIDTLR&lt;sup&gt;18&lt;/sup&gt;</td>
<td>85R----------90</td>
<td>ND</td>
<td>&gt;2300</td>
<td>0.1</td>
<td>86</td>
</tr>
<tr>
<td>dNdCs2</td>
<td>5MLESPVPEIDTLR&lt;sup&gt;18&lt;/sup&gt;</td>
<td>85R----------90</td>
<td>ND</td>
<td>&gt;3700</td>
<td>0.3</td>
<td>nd</td>
</tr>
<tr>
<td>dNR85Xs5</td>
<td>5MLESPVPEIDTLR&lt;sup&gt;18&lt;/sup&gt;</td>
<td>85L----------90</td>
<td>ND</td>
<td>&gt;1000</td>
<td>0.1</td>
<td>&gt;90</td>
</tr>
<tr>
<td>dNR85Xs10</td>
<td>5MLESPVPEIDTLR&lt;sup&gt;18&lt;/sup&gt;</td>
<td>85M----------90</td>
<td>ND</td>
<td>&gt;3600</td>
<td>0.3</td>
<td>&gt;90</td>
</tr>
<tr>
<td>dNR85Xs48</td>
<td>5MLESPVPEIDTLR&lt;sup&gt;18&lt;/sup&gt;</td>
<td>85K----------90</td>
<td>ND</td>
<td>&gt;2900</td>
<td>0.2</td>
<td>nd</td>
</tr>
</tbody>
</table>

<sup>a</sup> See footnote<sup>a</sup> of Table 5.1
<sup>b</sup> See footnote<sup>b</sup> of Table 5.1
<sup>c</sup> All truncated variant are descendants of re4.7s11
<sup>d</sup> The N-terminal region staring from position 5 and ending with position 18 is represented.
<sup>e</sup> The C-terminal region staring from position 85 and ending with position 90 are represented.

ND and nd are described in the footnote of Table 5.1

Several destabilized variants from libraries LdNdC and LdNR85X were produced and characterized to investigate the effects of terminal truncations and substitutions on the catalytic activity of re4.7s11. We
found that the weak *in vivo* complementation of these variants correlated well with a reduced catalytic efficiency *in vitro* (Table 5.2). The extensive deletions at the N-terminus of dNdCs1 and dNR85Xu15 caused an activity loss of three to four orders of magnitude as compared to the parent re4.7s11, exhibiting the largest catalytic lesions (Table 5.2). The catalytic activities of variants that had fewer amino acids deleted at the N-terminus (variants dNdCs2, dNdCs3, and dNdCs4), were reduced by about two orders of magnitude.

Another observation that followed from the kinetic analysis of the truncated variants indicated that mutations at Arg85 did not further reduce the catalytic efficiency. In fact, catalytic efficiencies of dNR85Xs5 (Arg85Leu), dNR85Xs10 (Arg85Met), and dNR85Xs48 (Arg85Lys) were comparable with the one of dNdCs2, which had an intact Arg85 (Table 5.2).

The melting points determined for a few variants indicated that truncation of five residues at the C-terminus together with Arg85 mutations do not reduce the thermal stability of these variants (Table 5.2). However, the variants with N-terminal deletions exhibit a substantial stability loss, reflected by reduction of a $T_m$ value by more than 10 degrees. It can, therefore, be concluded that the C-terminal truncation alone does not affect the stability of re4.7s11.

The reduction in catalytic activity in C-terminally truncated variants was possibly caused by the loss of interactions between the active site residues and the ones at the C-terminus, as seen in the crystal structure of MtDS-bound wild-type MtCM. Moreover, the activity decrease of simultaneously N-terminally truncated mutants is not only due to poorer substrate binding, reflected by the high $K_m$, but also due to overall stability loss as seen by the reduced $T_m$ values of the variants dNsCs1, dNsCs3 and, dNsCs4 (Table 5.2).

5.2.2.2 Directed evolution of destabilized re4.7s11 variants

The variety of catalytic activities and thermal stabilities exhibited by the truncated variants dNR85Xu15, dNdCs1, and dNR85Xs10 allowed testing various selection stringencies. The variants dNR85Xu15 and dNdCs1 exhibited the largest catalytic and stability impairments. Therefore, they were deemed suitable as parent mutants for the highest stringency selection of better variants. Variant
dNR85Xs10, however, retained a high thermal stability despite of a substantial reduction in catalytic efficiency and it offered the intact N-terminus for the potential accumulation of beneficial mutations. Moreover, evolving the less stable dNR85Xu15 and dNdCs1, as well as a highly stable dNR85Xs10 allowed testing the assumption, that higher thermal stability might be advantageous for accumulation of catalytically beneficial mutations that might otherwise be detrimental to stability.\[105\]

The genes for variants dNR85Xu15, dNdCs1 were mutagenized by epPCR and cloned either with or without the His\(_6\)-tag. The tag occurs in the immediate vicinity of the active site in the N-terminally truncated variants and might play a role in the accumulation of mutations that are only beneficial in its context. Therefore, removal of the tag prior to mutagenesis and selection should avoid these artifacts. However, there were essentially no complementers even on the lowest tested stringency plates M9c+F+Tet\(^{80}\)ng/mL, indicating that the removal of the His\(_6\)-tag destabilizes dNR85Xu15 and dNdCs1 variants further making it very difficult for the new mutations to rescue the CM activity. In fact, even in the presence of the His\(_6\)-tag the dNR85Xu15-based library did not yield any complementers. This demonstrated that the catalytic efficiency of this variant was too low for directed evolution. Therefore, only the His\(_6\)-tagged dNdCs1 and dNR85Xs10 were employed as starting variants for further evolution experiments.

Collections of 10\(^5\)-10\(^6\) epPCR mutagenized independent variants for libraries LdNdCs1 (His\(_6\)-tagged dNdCs1 as a parent) and LdNR85Xs10 (His\(_6\)-tagged dNR85Xs10 as a parent), were subjected to selection. On low stringency plates M9c+F+Tet\(^{80}\)ng/mL and M9c+Tet\(^{40}\)ng/mL approximately < 1 % of LdNdCs1 library clones and 5 % of LdNR85Xs10 library clones could complement. On the highest stringency plates M9c and M9c+pFPhe\(^{40}\)\(\mu\)M only < 0.1 % of clones from dNdCs1 and dNR85Xs10 libraries could complement.

The clones from LdNdCs1 and LdNR85Xs10 libraries, which were able to complement the CM deficiency of KA12/ppKIMP-UACU at the highest stringencies, were sequenced (Figure 5.5). As seen before for the inter-subunit destabilized variants (Figure 5.3), most of the mutations in the truncated variants appeared in the region between Glu37 and Ala45. Again, residues Lys40 and Gly43 were often targeted,
even though there were many substitutions that were unique to the truncated variants. As expected, the longer N-terminus of dNR85Xs10 allowed accumulation of new mutations, however, without any clear preference. Surprisingly, a large number of selected mutants from library LdNR85Xs10 had a Leu81Met mutation. As described previously, we mutagenized Leu81 in order to destabilize the truncated re4.7s11 variant (see section 5.2.2.1); however, re-appearance of this mutation in the selected clones from the high-stringency plates might, in fact, indicate that it increases the catalytic efficiency of the truncated re4.7s11.

Figure 5.5 Frequency of mutations in selected clones from truncation libraries. Sequencing results of selected clones from epPCR libraries using truncated variants dCdNs1 (deletion at the N-terminus is indicated by a dashed line) and dNR85Xs10 as well as library Ls4RepA2 (RepA tag is underlined) after epPCR and two cycles of DNA shuffling. A total of 23, 21, and 16 clones from LdCdNs1, LdNR85Xs10, and Ls4RepA2 libraries with correct sequences were analysed, respectively (artefacts were omitted). The frequency of appearance of a specific mutation is color coded as purple (<10 %), blue (10-30 %), orange (30-60 %) and red (60-100 %). Mutations that were present in the parent mutant re4.7s11 are highlighted in red for each library template.
Moreover, the appearance of a large number of artifacts (such as an elongated C-terminus, data not shown) in library LdNR85Xs10 and very few unique sequences found in LdNdCs1, indicated that one round of epPCR might not be enough to accumulate a sufficient number of mutations to improve these variants. Therefore, we resorted to a stepwise evolution protocol providing several rounds of mutagenesis and selection with gradual increase in selection stringency.

The variant dNdCs4 was chosen as a parent, due to its mediocre initial catalytic activity (Table 5.2). It was assumed that this activity would readily be improved during the first cycle of evolution and the resulting variants would be active enough to be tagged with protein degradation tag RepA\textsuperscript{[214]} for the next two cycles of DNA shuffling and selection. Furthermore, dNdCs4 variant had intact Arg85, reducing the risk of mutations that would only be beneficial in the context of substitutions at this position, as well as a minor deletion at the N-terminus, providing a larger portion of the gene for mutagenesis.

Since the His-tagged dNdCs4 variant complemented weakly on high stringencies the untagged version of dNsCs4 was used for the first cycle of evolution to increase the selection stringency. A library of 1.6×10^6 independent clones was plated on M9c+Tet\textsuperscript{80} ng/mL, M9c+Tet\textsuperscript{40} ng/mL, M9c+Tet\textsuperscript{5} ng/mL, M9c, M9c+pFPhe\textsuperscript{40} μM and M9c+pFPhe\textsuperscript{100} μM. While M9c+Tet\textsuperscript{80} ng/mL, and M9c+Tet\textsuperscript{40} ng/mL proved little selective with 20 % and 5 % complementers, respectively, the complementation rate on the other plates was < 0.1 %. Three out of seven sequenced clones, which complemented on M9c plates, did not have any mutations indicating that the initial selection stringency was relatively low even on M9c+pFPhe\textsuperscript{100} μM plates. Therefore, 180 clones, which complemented on M9c plates, were used for DNA shuffling and selection after providing them with a RepA protein degradation tag. A library of 9×10^5 independent RepA-tagged variants was plated on M9c+Tet\textsuperscript{80} ng/mL, M9c+Tet\textsuperscript{40} ng/mL, and M9c agar plates. As for the initial library, the first two selection conditions were only weakly selective, providing a complementation rate of 26 % and 7 %, respectively, whereas only < 0.01 grew on stringent M9c plates. For the last round of dNdCs4 evolution, DNA from 190 clones growing on M9c+Tet\textsuperscript{80} ng/mL and M9c+Tet\textsuperscript{40} ng/mL plates was isolated and shuffled to obtain the LS4RepA2 library. 3×10^7 independent variants of the LS4RepA2 library.
were plated on the highest available stringency plates M9c, M9c+pFPhe\textsuperscript{40} \textmu M, and M9c+pFPhe\textsuperscript{100} \textmu M to give < 0.001 \% complementers.

Sequencing of 29 selected LS4RePA2 clones revealed that > 30 \% lost their RepA tag due to insertions and 14 \% due to deletions. Sixteen clones that could complement on the highest selection stringency plates M9c+pFPhe\textsuperscript{100} \textmu M with the RepA tag retained exhibited several mutations unique to this library (such as Glu13Asp and Leu24Ile) as well as mutations seen before in the region between Glu37 and Ala45 (Figure 5.5).

In general, the directed evolution of truncated re4.7s11 variants dNdCs1, dNR85Xs10, and dNsCs4 yielded mutations that have been seen before for some of the evolved inter-subunit destabilized variants of 3p5. This indicated that acquired mutations are beneficial, regardless of the destabilization type. Moreover, none of the mutations in re4.7s11 (Thr52Pro, Val55Asp, Val62Ile, and Asp72Val) have been replaced, even though they were exposed to mutagenesis.

The proteins encoded by several evolved mutant genes carrying the most frequent mutations from libraries LdNdCs1, LdNR85Xs10 and LS4RePA2 were produced and purified. The kinetic analysis revealed that most mutants, including the ones that had Leu8Met substitution, exhibited increased catalytic efficiencies ($k_{cat}/K_m$) as compared to the corresponding parent proteins (Figure 5.6).

Several evolved variants from each of the libraries were elongated and exhibited higher catalytic efficiency than their truncated versions (Figure 5.6). However, there were no enzyme variants that exhibited largely increased activity as compared to the parent protein re4.7s11. The improved variants were only up to 30 \% better than re4.7s11 and the best among them, termed s10es4.15, contained mutations Val11Leu, Asp15Val, and Lys40Q.

As expected, the one tested variant containing the Leu81Met substitution from library LS4RePA2, exhibited substantially lower catalytic efficiency when the truncations were removed (Figure 5.6). Therefore, all mutants from the LdNR85Xs10 library containing this mutation were also excluded from further analysis. This result demonstrated that epistasis played an important role during this
Regulation, effectors, and evolvability of chorismate mutases from *Corynebacteriales*.

![Figure 5.6](legend next page)
Figure 5.6 Catalytic activity of selected mutants from truncation libraries. The activity of destabilized parent variants dNR85Xs10, dCdNs1 is shown in darker green and red, respectively. The evolved mutants (only the new mutations are shown) from each library are presented in the same colour as their parent mutant. In blue are the descendants from Ls4RepA2 library after removal of the RepA tag. The catalytic activity of the RepA tagged-dCdNs4 was not determined. The best variants s10es4.15 (V11L, D15V, K40Q) and s1es5.1 (A41V) are indicated by asterisk and the activity of the starting variant re4.7s11 is shown in black. For each variant only new mutations that occur in addition to the ones present in re4.7s11 are listed. For the individual kinetic parameters, see Tables A.III and A.IV in the Appendix.

evolution round meaning that the Leu81Met substitution is context-dependent.

5.2.3 Biochemical characterization of the evolved MtCM variants

To follow the effects of mutations gained throughout the evolution of MtCM, we characterized the best variants from each cycle. For that, the variants 3p3 (from the 1st evolution cycle), 3p5 (2nd cycle), re4.7s11 (3rd cycle), as well as s1es5.1 and s10es4.15 (both from the 4th cycle) were chosen.

The steady-state kinetics and melting temperature analysis showed that the gain in activity was achieved alongside increasing thermal stability (Table 5.3). The largest enhancement of catalytic efficiency was attained after the first evolution cycle when mutations Thr52Pro and Val55Asp were introduced into the H1-H2 loop of variant 3p3. The improvement was due to an increase in $k_{\text{cat}}$ and a decrease of $K_m$, indicating simultaneous improvement of catalytic machinery and substrate binding. Testing the effects of the individual mutations Thr52Pro and Val55Asp revealed that each individual exchange improves the catalytic efficiency of MtCM. These two mutations alter the wild-type sequence ($^{52}$TRLV$^{55}$) of MtCM to ($^{52}$PRLD$^{55}$), which resembles the corresponding sequence in the structurally similar EcCM from E. coli ($^{45}$PVRD$^{48}$). EcCM achieves a high catalytic efficiency without an interaction partner; therefore, Pro52/Asp55 in MtCM and Pro45/Asp48 in EcCM can play a similar role in positioning active site residues in a catalytically competent conformation (see section 5.2.4 below).
Table 5.3. Comparison of the catalytic parameters of His$_6$-tagged and untagged evolved MtCM variants

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Protein$^d$</th>
<th>Mutations</th>
<th>$k_{cat}^a$ (s$^{-1}$)</th>
<th>$K_m^a$ (μM)</th>
<th>$k_{cat}/K_m^a \times 10^4$ (M$^{-1}$ s$^{-1}$)</th>
<th>$T_m$ (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>H$_6$-MtCM</td>
<td>-</td>
<td>ND</td>
<td>&gt;1700</td>
<td>0.09 ± 0.004</td>
<td>75 ± 1</td>
</tr>
<tr>
<td>I</td>
<td>H$_6$-3p3</td>
<td>PD</td>
<td>12 ± 0.1</td>
<td>570 ± 16</td>
<td>2.0 ± 0.1</td>
<td>80</td>
</tr>
<tr>
<td>II</td>
<td>H$_6$-3p5</td>
<td>PDAM</td>
<td>31 ± 2</td>
<td>370 ± 21</td>
<td>8.3 ± 1.2</td>
<td>84 ± 4</td>
</tr>
<tr>
<td>III</td>
<td>H$_6$-re4.7s11</td>
<td>PD/PDAM, V62I, D72V</td>
<td>23 ± 3</td>
<td>72 ± 8</td>
<td>32 ± 1</td>
<td>&gt;89 ± 3</td>
</tr>
<tr>
<td>IV</td>
<td>H$_6$-s1es5.1</td>
<td>PD/PDAM, V62I, D72V, A41V</td>
<td>35 ± 2</td>
<td>103 ± 19</td>
<td>35 ± 7</td>
<td>85 ± 0.5</td>
</tr>
<tr>
<td></td>
<td>H$_6$-s10es4.15</td>
<td>PD/PDAM, V62I, D72V, V11L, D15V, K40Q</td>
<td>13 ± 2</td>
<td>30 ± 9</td>
<td>47 ± 8</td>
<td>82 ± 0.7</td>
</tr>
<tr>
<td></td>
<td>N-re4.7s11</td>
<td>PD/PDAM, V62I, D72V</td>
<td>28 ± 2</td>
<td>62 ± 16</td>
<td>47 ± 10</td>
<td>87 ± 0.2</td>
</tr>
<tr>
<td></td>
<td>N-s1es5.1</td>
<td>PD/PDAM, V62I, D72V, A41V</td>
<td>34 ± 4</td>
<td>143 ± 27</td>
<td>25 ± 5</td>
<td>82 ± 1</td>
</tr>
<tr>
<td></td>
<td>N-s10es4.15</td>
<td>PD/PDAM, V62I, D72V, V11L, D15V, K40Q</td>
<td>20 ± 3</td>
<td>42 ± 7</td>
<td>47 ± 6</td>
<td>86 ± 0.2</td>
</tr>
</tbody>
</table>

$^a$See footnote $^a$ of Table 5.1.

$^b$Data from the literature. [48]

$^c$Data from the literature. [26]

$^d$H$_6$ in front of the variant name indicates the presence of an N-terminal His-tag, N indicates the native N-terminus (see alignment in Figure 5.1).

$^e$See footnote $^b$ of Table 5.1.

ND is described in the footnote of Table 5.1.

PD indicates mutations T52P and V55D.

PDAM indicates mutations R87P, L88D, G89A, and H90M.

The second cycle of evolution yielded variant 3p5 with the substitutions Arg87Pro, Leu88Asp, Gly89Ala, and His90Met at the C-terminus. [26] Again, as for the 3p3 variant, the improvement of the overall catalytic efficiency was achieved by reduction of $K_m$ and increase of $k_{cat}$.

Introduction of mutations Val62Ile and Asp72Val into the re4.7s11 variant after the third round of evolution significantly improved the substrate binding by reducing $K_m$ by 4.6-fold, most
probably due to Ile62, which provides additional methylene group and might improve the packing of the active site.

The fourth and last round of evolution resulted in only minor gains. However, the two mutants s1es5.1 and s10es4.15 obtained after this cycle exhibited further improvements in $k_{\text{cat}}$ and $K_m$, respectively. Variant re4.7s11 exhibited a slightly reduced $k_{\text{cat}}$ in exchange of a highly improved $K_m$ compared to its precursor 3p5 and the single conservative exchange Ala41Val in the s1es5.1 variant restored the $k_{\text{cat}}$ to the level of 3p5. In the case of variant s10es4.15, the overall activity has been slightly improved by a 2-fold reduced $K_m$, which was now as low as for the MtDS-activated wild-type enzyme. The $k_{\text{cat}}$ and $K_m$ trade-off exhibited by this variant suggests that the mutations helped to stabilize the enzyme-substrate complex.

As mentioned before, the wild-type MtCM has been equipped with a His$_6$-tag to facilitate quick access to the proteins of the evolutionary intermediates. However, tagging included replacement of three N-terminal residues and reduced the catalytic efficiency of the native protein (Table 5.3). Therefore, to assess whether the newly accumulated mutations are also beneficial in the evolved variants without the His$_6$-tag, the original N-terminus (Figure 5.1A) was reconstructed for the best variants from each evolutionary cycle.

As indicated in Table 5.3, the untagged variants N-re4.7s11 and N-s10es4.15 retained their catalytic proficiency suggesting that mutations found in these variants are beneficial regardless the format of the N-terminus. On the other hand, variant N-s1es5.1 lost 19% of the activity exhibited by its His$_6$-tagged counterpart. This indicates that Ala41Val exchange is only beneficial in the presence of the tag. Interestingly, the thermal stabilities of the untagged mutants N-re4.7s11, N-s10es4.15, and N-s1es5.1 were similar to the ones of their His$_6$-tagged counterparts and also higher than that of the untagged wild-type enzyme.

Finally, we examined whether MtDS was still able to activate the evolved MtCM variants originating from the last two evolutionary cycles. As indicated in Table 5.4, such activation capacity was lost. This result was not surprising since the regions that were shown to interact
with MtDS (the H1-H2 loop and the C-terminus\cite{48}) were heavily mutated.

### Table 5.4 Impact of MtDS on the specific activity of untagged wild-type MtCM and the evolved variants

<table>
<thead>
<tr>
<th></th>
<th>N-MtCM</th>
<th>N-re4.7s11</th>
<th>N-s1es5.1</th>
<th>N-s10es4.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>- MtDS</td>
<td>1.3 ± 0.1\textsuperscript{a}</td>
<td>41.3 ± 3.2</td>
<td>52.8 ± 8.6</td>
<td>26.6 ± 3.8</td>
</tr>
<tr>
<td>+ MtDS</td>
<td>14.3 ± 1.3</td>
<td>43.5 ± 4.5</td>
<td>48.4 ± 6.4</td>
<td>24.9 ± 2.3</td>
</tr>
</tbody>
</table>

\textsuperscript{a}The specific initial velocities of chorismate consumption monitored at 274 nm using 5 nM of CM with or without 50 nM MtDS in BTP+, pH 7.5 buffer at 30°C. Standard deviations were determined from at least three measurements.

#### 5.2.4 Crystal structures of the evolved MtCM variants

MtCM undergoes structural rearrangements upon complex formation with MtDS that are associated with a 100-fold activity increase.\cite{48} To monitor how the newly obtained amino acid exchanges in the evolved variants affected the overall fold, we subjected the best variants to crystallographic studies in collaboration with professor Ute Krengel’s group at the University of Oslo. In this study, we aimed to explore whether the mutations induce similar conformational changes, or whether a different mechanism governs the activation of the evolved MtCM variants.

Initially, the N-terminally His\textsubscript{6}-tagged variants re4.7s11, s1es5.1, and s10es4.15 were subjected to crystallization; however, neither of them yielded crystals (Helen Vikdal Thorbjørnsrud, University of Oslo, personal communication). It is possible that the His\textsubscript{6}-tag interferes with the crystal formation since also previous attempts to crystallize the His\textsubscript{6}-tagged MtCM variants were successful only when the tag was removed.\cite{160}

Attempts to obtain crystals for the untagged variants N-re4.7s11, N-s1es5.1, and N-s10es4.15 yielded well diffracting crystals for apo N-s1es5.1 and apo N-s10es4.15 (not described here; work by Helen Vikdal Thorbjørnsrud) as well as crystals obtained after co-crystallization of N-s10es4.15 with Bartlett’s TSA.\cite{33} Crystallization experiments of N-re4.7s11 resulted in poorly diffracting needle-like crystals suggesting that further optimization of the crystallization
conditions is needed (Helen Vikdal Thorbjørnsrud, personal communication).

In this work, synchrotron diffraction data from crystals of N-s10es4.15 co-crystallized with TSA and N-s1es5.1 without a ligand (work by Helen Vikdal Thorbjørnsrud) have been processed and refined. Unlike the wild-type enzyme (space group $P4_32_12$\textsuperscript{48}) and other evolved MtCM variants (space group $P2_12_12_1$\textsuperscript{160}), N-s10es4.15 and N-s1es5.1 crystallized in the space group $P6_4$ with a single protomer in the asymmetric unit.

The crystal structures of N-s10es4.15 and N-s1es5.1 were solved by molecular replacement to 1.5 Å and 1.7 Å resolution, respectively (Table A.V in the Appendix). Even though the 1.5 Å resolution of the N-s10es4.15 variant co-crystallized with TSA would allow identification of a small molecule, there was no convincing difference density ($F_o - F_c$ maps) for TSA. Instead, the electron density in the active site of variant N-s10es4.15 appeared to be identical to the one in N-s1es5.1, which was crystallized in the absence of a ligand. Therefore, a similar network of water molecules was modeled into the active sites of both crystal structures.

5.2.4.1 Crystal packing and the overall fold of the evolved variants

The structures of the evolved variants N-s10es4.15 and N-s1es5.1 are nearly identical and fold into three helices per subunit like the wild-type MtCM (Figure 5.7A). The N- and C-termini of these models have not been fully modeled due to the lack of convincing electron density indicating that these regions are rather flexible. However, a few resolved residues make it clear that the C-terminus adopts a different conformation in the variants than in wild-type MtCM with or without MtDS. In the MtDS-bound MtCM, the C-terminal residues are docked into a hydrophobic groove of the complex partner, whereas the positioning of the C-terminal residues in the free wild-type MtCM as well as in the evolved variants s10es4.15 and s1es5.1 is influenced by the crystal contacts (Figures 5.7 and 5.8).

Minor differences between the monomers of the evolved variants N-s10es4.15 and N-s1es5.1 can be seen at the C-terminal part of helix H1 and the flexible loop between helices H1 and H2 (Figure 5.7 A). This region, together with the N-terminal part of the helix H2, assumes a
different conformation in the crystal structures of unactivated (apo MtCM, PDB ID: 2QBV, malate-bound MtCM, PDB ID: 2VKL) and MtDS-activated (MtDS-bound MtCM, PDB ID: 2W1A, MtDS-bound MtCM with TSA, PDB ID: 2W19) wild-type enzyme and is associated with the activation mechanism.[48] Remarkably, the conformation of the loop in the activated MtCM is nearly identical to the one observed for the corresponding H1-H2 loop of the structurally related CM from *E. coli* (EcCM). The highly active EcCM shares 39% sequence identity with MtCM and is structurally very similar (Figure 5.7C). Therefore, understanding how the loop conformation affects catalytic activity is essential for understanding how these closely related AroQ enzymes operate.

![Figure 5.7 Overall structure of evolved MtCM variants.](image)

The significant structural changes in the H1-H2 loop upon MtCM-MtDS complex formation are coupled with repositioning of MtCM residues Val55 and Arg46. It is likely that the conformation of this region is responsible for the high enzymatic activity of the activated MtCM and the evolved variants N-s10es4.15 and N-s1es5.1. Indeed, the
H1-H2 loop of the evolved variants adopts a conformation that is similar to the ones observed in EcCM and the activated MtCM (Figure 5.7 C), but is largely different from the one in the inactive MtCM (Figure 5.7 B). However, can the observed conformation of the flexible loop be adopted in solution, too, or is it a consequence of the crystal packing? To answer this questions the forces that keep the loop in the “activated” conformation in the MtDS-bound MtCM and EcCM need to be analyzed.

For instance, in EcCM, the arrangement of the H1-H2 loop does not depend on a partner enzyme or on crystal contacts and relies on the intra-dimer interactions and properties of specific residues. The turn in the loop is induced by Pro45 and is further extended by Arg47, which forms hydrogen bonds with Glu52 in the H2 helix. The N-capping of the H2 helix is achieved by Asp48, which forms a network of polar interactions with Arg11′, a catalytic residue provided by the other subunit.

The structures of the apo and malate-bound unactivated MtCMs exhibit a conformation of the flexible H1-H2 loop, which may be highly influenced by interactions with neighboring crystallographic partners. Due to the lack of specific intra-dimer interactions it is safe to assume that this region is flexible and might not be in a dominant single conformation in solution. Instead, when MtCM binds to MtDS, the loop can be stabilized via non-covalent interactions with the specific regions at the interface with MtDS, like those seen for the C-terminal residues of MtCM. For example, the interaction with the MtDS helix between Phe384 and Leu401 induces the characteristic turn in the loop. It is further stabilized by Arg53 from the H1-H2 loop of MtCM interacting with the backbone carbonyl of His90, which is brought in close proximity as a result of the interaction with MtDS.

The H1-H2 loop of the evolved variants N-s10es4.15 and N-s1es5.1 does not rely on the same type of stabilization as in the activated (MtDS-bound) wild-type MtCM. Also, the C-terminus of these variants is affected by the crystal contacts (Figure 5.8), therefore, an electrostatic interaction between His90 and Arg53 as observed in the activated wild-type enzyme is not present. Instead, an EcCM-like intra-dimer stabilization of the loop exists in the evolved variants. The equivalents of EcCM residues Pro45 and Asp48 in s10es4.15 and
s1es5.1 are Pro52 and Asp55. Similarly as in the *E. coli* enzyme, Pro52 induces the turn and Asp55 stabilizes the N-terminus of the H2 helix via polar interactions with the backbone of Ser49 in the helix H2 as well as Arg18’ from the neighboring subunit.

Figure 5.8 Crystal packing. Stereo representation of the unit cell contents for the s10es4.15 crystal in space group P6_4 (the unit cell of s1es5.1 crystals is identical). The unit cell (black box) is shown from the perspective of the γ (120°) angle with the c-axis being vertical. The asymmetric unit contains one subunit of a homodimer. Symmetry-related protomers are shown in black or grey, except for the highlighted blue and orange central protomers. The protomers that are positioned on top of each other crystalize in close contact between the N-terminus of the H1 helix of the protomer below and the active site region of the protomer above. These contacts may influence the conformation of the loop between helices H1 and H2 (brown arrow) and the C-terminus (blue arrow).

However, despite the similarities of possible loop stabilization between the evolved MtCM variants and EcCM, the loop conformation in s10es4.15 and s1es5.1 is also susceptible to crystal contacts. A careful inspection of the crystal packing revealed that the side chains of the loop residues are involved in a distinct network of electrostatic interactions with the residues of the H1 helix in the crystallographically related subunit (Figure 5.8). As a consequence,
Arg53 cannot hydrogen-bond to Glu59, in the same way as Arg47 interacts with Glu52 in EcCM, due to its interaction with Asp27 of the crystallographic neighbor. The different crystal packing providing distinct stabilizing interactions and the lack of stretching induced by the Arg53-Glu59 interaction might explain why the overall fold and the kink in the H1-H2 loop of the evolved variants is less pronounced than in EcCM and in the activated MtCM.

Nevertheless, the general pattern of the loop conformation typical for the activated MtCM can be observed in the crystal structures of the evolved variants (Figure 5.7). Supported by the similar loop conformation in EcCM, this suggests, that it might be a common mechanism for positioning relevant active site residues in a catalytically competent conformation of AroQ enzymes.

5.2.4.2 The active site of the evolved variants

Residues Arg35, Ser39, Arg46, Val55, Arg58, Glu59, and Val62 of one subunit together with Arg18’ of the other subunit bind Bartlett’s TSA in the active site of MtDS-bound MtCM and have been suggested to form the catalytic site (Figure 5.9 A). The side chains of Arg35 and Val62 assume nearly the same conformations in the activated (MtDS-bound MtCM with TSA and without TSA, PDB IDs 2W1A and 2W19, respectively) and non-activated (apo and malate-bound MtCM, PDB IDs 2QBV and 2VKL, respectively) enzyme and do not seem to play a major role in the activity switch. Therefore, it is not surprising that Arg35 and Ile62 in the evolved variants s10es4.15 and s1es5.1 adopt the same conformations as in the wild-type enzyme. However, introduction of the Val62Ile mutation greatly reduced the $K_m$ of variant re4.7s11 (Table 5.3). The Val62Ile exchange occurred several times in independent directed evolution experiments and alone provided a 25-fold improvement in the catalytic efficiency of MtCM.[26] The side chain of Ile62 in s10es4.15 and s1es5.1 does not change its overall position as compared to the Val62 in the wild-type MtCM, but could be associated with a better packing of the active site since it provides an additional methylene group (Figure 5.9).

The active site residues Ser39, Glu59, Val55, Arg46, and Arg18’ undergo significant changes once MtCM binds to MtDS and adopt conformations that are more competent for binding TSA. Therefore, we
expected that a similar pre-organization of these active site residues would occur in the apo structures of s10es4.15 and s1es5.1, explaining their high catalytic activity. However, the positioning of these residues in the evolved variants draws a relatively complex picture (Figure 5.9).

For example, the conformation of Ser39 is identical to the one seen in the unactivated form of MtCM and cannot provide its hydroxyl group for coordination with TSA via a water molecule. Moreover, the side chain of Glu59, which hydrogen bonds to the hydroxyl group of TSA in the activated wild-type enzyme, adopts a very unusual position in the evolved variants. In fact, the orientation of Glu59’s carboxyl group would result in steric clashes with TSA if it would be bound in the same position as in the MtDS-bound MtCM with TSA. Regardless of whether MtCM is in complex with MtDS or not, Glu59 is similarly positioned and forms hydrogen bonds with Arg53 (from the H1-H2 loop) in the MtCM-malate structure, or with Arg85 (from the C-terminus) in the MtCM-MtDS structure. Neither of these interactions is available in the crystal structures of the evolved variants, due to the crystal contacts. Therefore, the side chain of Glu59 turns and interacts with the water molecules in the active site (Figure 5.9 B, C, D, and E).

Despite the apparently catalytically incompetent positioning of Ser39 and Glu59, other active site residues adopt conformations that resemble the activated form of MtCM. For instance, the side chain guanidinium of Arg18’ and the backbone of Asp55 are clearly shifted to adopt positions similar to the ones in the MtDS-bound MtCM (Figure 5.9 B, C, D and E). Also, the side chain of Arg18’ appears to be less flexible than in the apo wild-type MtCM structure, where two conformations are observed, and it forms a hydrogen bond with Asp55 that was introduced during the directed evolution (Figure 5.9B). Therefore, it can be assumed with confidence, that Arg18’ and Asp55 are pre-arranged for the efficient binding of the transition state of the reaction.

Interestingly, residue Arg58 is clearly shifted in the MtDS-bound apo MtCM. This could be associated with the fact that Arg18’ has not been modeled in this structure since the entire N-terminus appears to be disordered (Figure 5.9D). Meanwhile, Arg58 in the evolved variants is pre-organized in the same way as in MtDS-bound MtCM with TSA as well as apo MtCM without the interaction partner, which have Arg18’
modeled (Figure 5.9C and E). This indicates that stable positioning of Arg18' plays a role for catalytically competent conformation of Arg58.

Finally, the most important residue for CM catalysis is Arg46, which interacts with the ether oxygen of TSA in the MtDS-bound MtCM. It was suggested to stabilize the negative charge developing in the transition state (Figure 1.4) and, therefore, is considered to be the essential active site residue.\[^{48}\] This residue adopts slightly different orientations in s10es4.15 and s1es5.1. In the latter variant it appears to be more flexible since two side chain conformations could be modeled into the electron density map (Figure 5.9B). Moreover, the guanidinium group of Arg46 in both s10es4.15 and s1es5.1 points away from the active site and adopts a different orientation than in MtDS-bound MtCM with or without TSA. This can be explained in part by the absence of the ligand in the active site of the evolved variants, which could provide a moiety for interaction. Also, the absence of a defined Arg18' conformation in the MtDS-bound apo MtCM most probably allows Arg46 to adopt conformation similar to the one in the TSA-bound activated enzyme. Since Arg18' seems to be more stably positioned in the evolved variants, it could cause repulsion of the charged guanidinium of Arg46 and thus shift it away from the active site. Nevertheless, despite several differences, the overall position of Arg46 in the evolved variants is more similar to the one observed in the activated than unactivated MtCM (Figure 5.9 C, D, and E).

Stable pre-positioning of Arg46 into a catalytically competent conformation might explain the high enzymatic efficiencies of the evolved s10es4.15 and s1es5.1. In the crystal structures of the unactivated wild-type MtCM, Arg46 occupies the space assumed by the flexible loop between helices H1 and H2 in the activated MtCM. Once the complex with MtDS is formed, the loop descends and occupies the position assumed by Arg46 in the unactivated MtCM structure. Therefore, the positioning of Arg46 can be connected with the conformational changes adopted by the H1-H2 loop upon complex formation (Figure 5.10). In this way, stabilization of Arg46 is reminiscent to the stabilization of its homolog Lys39 in the highly active EcCM. Due to a specific conformation of the corresponding H1-
Regulation, effectors, and evolvability of chorismate mutases from *Corynebacteriales*

**Figure 5.9** (legend next page)
Figure 5.9 Active site of MtCM. (A) Scheme of MtCM’s key active site residues with bound TSA (interactions according to PDB ID: 2W1A). Mutations Val55Asp and Val62Ile found in the evolved variants are in blue (B) Comparison of active site residues in s10es4.15 (magenta) and s1es5.1 (purple). Hydrogen bonds between residues Asp55 and Arg18’ of the other subunit are indicated by the dashed lines. (C) Comparison of active site residues in s10es4.15, s1es5.1, and activated MtDS- and TSA-bound wild-type MtCM (green, PDB ID 2W1A). (D) Comparison of active site residues in s10es4.15, s1es5.1, and activated MtDS-bound apo wild-type MtCM (green, PDB ID 2W19). (E) Comparison of active site residues in s10es4.15, s1es5.1, and apo wild-type MtCM (yellow, PDB ID 2QVB) (E) In (B), (C), (D), and (E) Arg35 is omitted for clarity reasons, since it assumes the same conformation in the activated and unactivated MtCM forms.

H2 loop, residue Lys39 cannot adopt a conformation other than the catalytically competent one without steric clashes with the loop residues. The H1-H2 loop in the evolved variants adopts a conformation that is observed both in EcCM and the activated MtDS-bound MtCM. This conformation reduces the flexibility of Arg46 and locks it in a catalytically competent position. Therefore, we assume that the closed loop conformation, which most probably also occurs in solution, is essential for efficient catalysis of these closely related AroQ enzymes.

Figure 5.10. Position of the H1-H2 flexible loop. The Arg46 (sticks) from s10es4.15 (magenta), s1es5.1 (purple), activated wild-type MtCM (green, PDB ID 2W1A), and non-activated malate-bound wild-type MtCM (yellow, PDB ID 2QVB) is indicated. Hydrogen bonds between Arg46 fromin the activated MtCM (green) and TSA (blue lines) are indicated by dashed lines.
5.2.4.3 Effects of other mutations in the evolved s10es4.15 and s1es5.1 variants

The crystal structures of the evolved N-s10es4.15 and N-s1es5.1 variants reveal the major contribution of the mutations Thr52Pro and Val55Asp in positioning Arg18' and the H1-H2 flexible loop, which in turn directs the essential active site residue Arg46 to its catalytically competent conformation. Also, the Val62Ile mutation might significantly improve the active site packing allowing for better substrate binding (Figure 5.11).

However, the contribution of other mutations is less clear. For example, the crystal packing influences the C-terminus, which holds Arg87Pro, Leu88Asp, Gly89Ala, and His90Met mutations, occluding the potential interactions between these residues and the rest of the protein in the evolved variants.

Figure 5.11 X-ray structures of the evolved MtCM variants. The mutation acquired by (A) s10es4.15 and (B) s1es5.1 are shown in green sticks.
The Asp72Val mutation occurs at the interface of the H3 helices of the two interacting subunits. In wild-type MtCM, Asp72 forms hydrogen bonds with Arg82 of the other subunit and its replacement with Val would disrupt this interaction. However, Val increases the overall hydrophobicity of the H3 helix interface and might improve the interaction between the two monomers.

The two mutations Val11Leu and Asp15Val acquired by the s10es4.15 variant occur at the very end of the N-terminus and could potentially stabilize this otherwise rather flexible region. The Lys40Gln substitution introduces hydrogen bonds with the backbone of Arg85, which in turn has been shown to play a role in the catalytic mechanism\textsuperscript{126} and was used to destabilize the truncated re4.7s11 variants in this work. Therefore, manipulation of Arg85 via Gln40 might play a role in efficient catalysis.

5.3 Conclusions

Diffusion is the only known limitation for the efficiency of enzyme catalysis. However, most metabolic enzymes have evolved to catalyze the reactions three to four orders of magnitude slower.\textsuperscript{98} It is of great interest to find the factors that govern nature’s decisions for making mediocre catalysts – is it the protein’s intrinsically limited potential to evolve or the lack of sufficient selection stringency? In this work, we aimed to answer the question whether MtCM holds the inherent property to achieve higher than MtDS-mediated catalytic activity and whether such an achievement would require a mechanism distinct from the one utilized by nature.

Directed evolution is a powerful method to investigate enzyme catalysis. We applied a CM selection system that couples the enzyme’s activity to bacterial growth under stringent selective conditions. The limitation of the selection system was reached when the target enzyme was active enough to fully complement the metabolic deficiency under the highest possible stringency thus precluding further evolution. In practice, this limitation can be overcome by applying a variety of strategies to increase the stringency. In this study, we successfully implemented structural destabilization, regulatable promoters, a protein degradation tag, and the toxic pFPhe to reduce both catalytic activity (temporarily) and cellular concentration of our target enzyme.
MtCM. This allowed the discovery of MtCM variants that were nearly 2-fold more active than the fully activated wild-type enzyme. We obtained variants that had $k_{\text{cat}}$ and $K_m$ values improved indicating better chemistry at the active site and more efficient substrate interactions. However, most of the evolved destabilized mutants exhibited relatively low activities and could, in principle be evolved further. For example, most of the tested evolved truncated re4.7s11 variants exhibited a $K_m$ higher than 100 μM (Table A. III in the Appendix). Since the $K_m$ values of enzymes tend to be adapted to the normal intracellular concentrations of the substrate, one could assume that further evolution of the destabilized MtCMs might still be possible, thereby improving $k_{\text{cat}}/K_m$.

Finally, we solved the crystal structure of the best-evolved variants and discovered that the structural changes exhibited by these mutants follow the same trend as in the activated wild-type MtCM and its very active counterpart EcCM. It suggests that the high activity of the evolved variants is most probably governed by stable pre-organization of the active site residues into catalytically competent conformations as observed in the MtDS-bound wild-type enzyme and EcCM. However, due to specific crystal contacts, a few active site residues exhibited conformations that would preclude substrate binding in the active site of the evolved N-s10es4.15 and N-s1es5.1. To resolve these issues, crystal structures of the TSA- or product-bound variants are needed.

In conclusion, we show that MtCM has the intrinsic capacity for efficient catalysis; however, it has evolved as a mediocre enzyme to fulfill nature’s need for allosteric regulation at this central location of the shikimate pathway.

5.4 Materials and Methods

5.4.1 Materials and general procedures

DNA manipulations were performed as described in 4.4.1. Chorismate was produced following a published protocol. DL-para-fluoro-phenylalanine (pFPhe) was from Bachem and other chemicals were purchased from Sigma-Aldrich/Fluka.
5.4.2 Strains, plasmids, and general procedures

*E. coli* strains KA12 and KA13 were described in section 4.4.2. The plasmid pKIMP-UAUC [51] was used in selection experiments. It carries the chloramphenicol resistance gene and provides the genes *tyrA* and *pheC* encoding monofunctional forms of PDH from *Erwinia herbicola* and PDT from *Pseudomonas aeruginosa*, respectively. Plasmid pT7POLTS [126] (originally called pAC-Ptet-T7pol-S [225]) was used for protein overproduction in strain KA12. It carries a chloramphenicol resistance gene and provides the genes *tyrA* and *pheC* encoding monofunctional forms of PDH from *Erwinia herbicola* and PDT from *Pseudomonas aeruginosa*, respectively. Plasmid pKTNTET-0 [26] served as the acceptor vector for the CM libraries. It contains the Met5-Ser39 fragment of the *aroQδ* gene N-terminally tagged by the Met-His6 sequence and followed by the SSG linker. Plasmids pKTCMM-H, [48] pHS10-3p5 [26] were described before.

5.4.3 Directed evolution of 3p5 via inter-subunit destabilization

5.4.3.1 Construction of destabilized Leu24 and Leu31 MtCM variants

Libraries L1, L2, L3, and L4 were constructed via overlap extension PCR, using plasmid PHS10-3p5 [26] as the template. The 143 bp 5’ terminal part of the gene was constructed using the forward primer 352-CTLIB-S (5’-CCTTGTTCATATGCACCATCATCATCACCACTCTT) and a reverse primer for each library, which introduced an Ascl restriction site, deleted away the BsaHI site and randomized specific codons. The reverse primers 505-L24-N (5’-AAACCTCGGCAGCCGCGCTTTGACTAACGCGAGGATTTCAGCATC) MNNACCGGTCGATCTCTTCGCGCA, randomized Leu 24 via NNK codon), 506-L31-N (5’-AAACCTCGGCAGCGCGCTTTAACMNNAGCGAGGATTTCAGCATC) CTAGCCCGGTGATCTCTTCGCGCA, randomized Leu 31 via NNK codon), 507-L24-L31_N (5’-AAACCTCGGCAGCCGCGCTTTAACMNNAGCGAGGATTTCAGCATC)
Regulation, effectors, and evolvability of chorismate mutases from *Corynebacteriales*

CMNNACGGTGCATCTCTTCCGCGCA, randomized Leu 24 and L31 via NNK codons), and 508-L24-L31-RNK-N (5’-AAACCTCGGCGCGCCGCTTAACMNYAGCGAGGATTTCAGCATCAGCATC

MNYACGGTGCATCTCTTCCGCGCA, randomized Leu 24 and L31 via RNK codon excluding the stop codon) were used for the gene libraries L1, L2, L3, and L4 respectively.

The 202 bp 3’terminal part of the gene was constructed the same way for all three libraries using forward primer 509-AseI-S (5’-AAGCGGCAGCGCGAGGTTTCAAGCCATCGG, introduced AseI site) and a reverse primer 510-noBsaHI-N (5’-TCACAGCTTTCCGCGCCACTAGTATTACATAGCATCCGGACCACGACCAA GAC).

The 143 bp and 202 bp PCR fragments were assembled using external primers 352-CTLIB-S and 510-noBsaHI-N and the resulting 326 bp fragment was restriction digested with *XhoI* and *SpeI*. The obtained 260 bp fragments from L1, L2, L3, and L4 libraries were ligated with the 2761 bp *XhoI/SpeI* fragments of cut pKTNTET-0 yielding the library plasmid of 3021 bp.

5.4.3.2 Selection of destabilized Leu24 and Leu31 MtCM variants

The transformed electrocompetent KA12/pKIMP-UAUC cells were washed twice in M9 salts (components described in section 4.4.5 and [155]) and plated on non-stringent M9c+F+Tet500ng/mL minimal medium (M9c medium components described in section 4.4.5) with 500 ng/mL tetracycline inducer for gene expression and 20 μg/mL L-phenylalanine. A small fraction of washed cells were plated on non-selective M9c+FY plates (minimal M9c medium additionally containing 20 μg/mL L-tyrosine and 20 μg/mL L-phenylalanine) for library size estimation. Typically 5 clones that grew on these plates were sequenced to determine the quality and mutation rates of the library. The clones that grew on M9c+F+Tet500ng/mL plates were picked, and streaked out on higher stringency plates such as M9c+Tet80ng/mL containing 80 ng/mL tetracycline, M9c without additives and M9c+pFPhe40μM containing 40 μM toxic pFPhe. The clones that did not complement on these conditions were sequenced using 131-TERM (5’-CCCTCAAGACCCGTATTAGA).
5.4.3.3 Acceptor vectors for L1-13, L4-4, L4-7, L3-6, L3-7, and L3-8 libraries

Acceptor vectors for L1.13, L4.4 and L4.7 libraries were constructed by ligating a *NheI/HindIII* restriction digested 1261 bp stuffer fragment from pKTNTET-0 with the correspondingly digested 2873 bp fragment of pKTNTET-L1.13, pKTNTET-4.4 and pKTNTET-4.7, containing the gene of the L1-13, L4-4 or L4-7 variants, respectively, (see section 5.4.3.2). This yielded the 4134 bp acceptor vectors pKTNTET-0-L1.13, pKTNTET-0-4.4 and pKTNTET-0-4.7.

To eliminate the second Ascl site for the L3.6, L3.7 and L3.8 libraries, the new acceptor vectors were constructed by ligating the *Ascl/HindIII* restriction digested 1253 bp stuffer fragment from pKTNTET-0 with the correspondingly digested 2833 bp fragment of pKTNTet-L3.6, pKTNTet-L3.7 or pKTNTet-L3.8 (see section 5.4.3.2), yielding the 4086 bp acceptor vectors pKTNTet-0-L3.6, pKTNTet-0-L3.7 and pKTNTet-0-L3.8.

5.4.3.4 Construction and *in vivo* selection of epL1-13, epL4-4, epL4-7, epL3-6, epL3-7, epL3-8, shL1-13, shL4-4, shL4-7, shL3-6, shL3-7, and shL3-8 libraries

To obtain the epPCR libraries (epL) two rounds of error-prone PCR were performed since the target *aroQ* gene is too short to allow the full mutagenic capacity of the Mutazyme II kit from Stratagene (Agilent Technologies). Initially, the appropriate DNA containing the *aroQ* gene of plasmids pKTNTET-L1.13, pKTNTET-L4.4, pKTNTET-L4.7, pKTNTET-L3.6, pKTNTET-L3.7, and pKTNTET-L3.8 was amplified using primers 372-TetPro (5’-AGCTCTAATGCCTGTTAATCAGT) and 131-TERM (5’-CCCTCAAGACCCGTTTAGA) and 0.5 ng of the template plasmid to yield a 591 bp product. The second round of epPCR for the epL3-6, epL3-7, and epL3-8 libraries was done by using 0.5 ng of the 591 bp epPCR product and the 509-Ascl (5’-AAGCGCGCCGCGGAGGTTTCCAGGCATCGG)/131-TERM primer pair to give the 253 bp product, which was used for cloning after appropriate restriction digestion. For epL1-13, epL4-4 and epL4-7 libraries 0.5 ng of the appropriate 591 bp epPCR product and the 372-TetPro/131-TERM primer pair was used for the second round of epPCR to again generate the a 591 bp product. Therefore, additional non-mutagenic PCR was performed using primers 509-Ascl and 131-TERM to obtain the desired 253 bp product for cloning of epL1-13, epL4-4 and epL4-7 libraries.
DNA shuffling libraries were constructed by shuffling aroQ\textsubscript{6} genes carrying beneficial mutations including the genes from previously constructed plasmids pHS10-3p5 (carrying the aroQ\textsubscript{6} gene coding for variant H\textsubscript{6}-PHS10-3p5, mutations T52P, V55D, R87P, L88D, G89A, H90M), pHS08-3p20 (H\textsubscript{6}-PHS08-3p20, mutation V62I), pHS08-5p12 (H\textsubscript{6}-PHS08-5p12, mutation G43V), pHS10-ANp10 (H\textsubscript{6}-PHS10-ANp10, mutation R82Q) and pHS10-2p14 (H\textsubscript{6}-PHS10-2p14, mutation D75Y).\textsuperscript{[26]} These genes were PCR amplified using primers 131-TERM and 509-AscI to yield a 591 bp product. The PCR products were mixed to have 600 ng of each fragment and treated with 1 μg DNAse until the preferred fragments of approximately 50 bp were generated. The resulting DNA fragments were re-assembled by PCR using primers 131-TERM and 509-AscI to yield a 253 bp product. If the yield of the re-assembly PCR was low, additional PCR amplification with the same primers 131-TERM and 509-AscI was performed.

The 253 bp products of the error-prone PCRs and the assembly PCR of DNA shuffling for L1.13, L4.4 and L4.7 libraries were restriction digested with AcsI and HindIII and the resulting 188 bp fragments were ligated to the correspondingly digested 2833 bp fragment of acceptor vectors pKTNTET-0-L1.13, pKTNTE-0-L4.4 or pKTNTET-0-L4.7, yielding 3021 bp library plasmids. The 253 bp products of error-prone PCR and re-assembly PCR for L3.6, L3.7 and L3.8 libraries were restriction digested with AcsI and SpeI and the resulting 174 bp fragments were ligated to the correspondingly digested 2847 bp fragment of acceptor vectors pKTNTET-0-L3.6, pKTNTET-0-L3.7 and pKTNTET-0-L3.8, yielding the 3021 bp library plasmids.

The library plasmids were transformed into electrocompetent KA12/pKIMP-UAUC cells. The cells were washed twice with 1 × M9 salts and spread on M9c plates of various stringencies. The plates are listed by increasing stringency: M9c+F+Tet\textsuperscript{80ng/mL} containing 80 ng/mL tetracycline and 20 μg/mL L-phenylalanine, M9c+F+Tet\textsuperscript{40ng/mL} containing 40 ng/mL tetracycline and 20 μg/mL L-phenylalanine, M9c+Tet\textsuperscript{80ng/mL} containing 80 ng/mL tetracycline, M9c+Tet\textsuperscript{40ng/mL} containing 40 ng/mL tetracycline, M9c without any additives, M9c+pFPhe\textsuperscript{40 μM} containing 40 μM pFPhe and M9c+pFPhe\textsuperscript{100 μM} containing 100 μM pFPhe. Library size and quality were determined on non-selective M9c +FY plates. Colonies complementing on the highest
selective conditions were picked and sequenced using oligonucleotide 131-TERM (5’-CCCTCAAGACCCGTTTAGA).

5.4.3.5 Removing destabilizing mutations from the inter-subunit destabilized evolved variants

Overlap extension PCR was used to remove L24 and L31 mutations from the evolved aroQδ variants. The PCR I fragment was obtained by using oligonucleotides 372-TetPro (5’-AGCTCTAATGCGCTGTTAATCACT) and 512-x31L-Fw (5’-GACGCCGAAATCCTCGCGTATTAGTCAAGCGACGCGCTGAGG) with using the appropriate pKTNTET-based plasmid as a template. The PCR II fragment was obtained with oligonucleotides 513-x24L-Rev (5’-CGCGAGGATTTCGGCGTCTAGCCGGTCGATCTCTTCGCGC) and 131-TERM (5’-CCCTCAAGACCCGTTTAGA), by using the same pKTNTET-based plasmid as a template. PCR I and PCR II fragments were assembled using oligonucleotides 372-TetPro and 131-TERM. The resulting DNA fragment was restriction digested with XhoI and HindIII to obtain the required 349 bp insert. The insert was ligated with the 2747 bp fragments from the appropriately digested pKTNTET-0.

5.4.4 Directed evolution of re4.7s11 via terminal truncation

5.4.4.1 Construction of truncated re4.7s11 variants

The LdNdC library of truncated re4.7s11 variants was constructed using pKTNTET-re4.7s11 with the His6-tagged version of re4.7s11 as a template and a mixture of several forward and reverse primers. The forward primers 521-MtCMi-N-S8 (5’-GTATGCTCGAGTCCCAACCTGTCCCGAGATCGACGC), 522-MtCMi-N-P10 (5’-GTATGCTCGAGCTGTCCCGAGATCGACGCTGC), 523-MtCMi-N-P12 (5’-GTATGCTCGAGCGCCCGAGATCGACGCTGC), 524-MtCMi-N-I14 (5’-GTATGCTCGAGATCGACGCTGCAGCGAAG), 525-MtCMi-N-T16 (5’-GTATGCTCGAGACGCTGCGAGAGATCGAC), and 526-MtCMi-N-R18 (5’-GTATGCTCGAGGCAGACGATCGACGCTGCAGCGT) introduced a XhoI restriction site via the Leu-Glu sequence right after the start methionine and respective N-terminal truncations up to residue Ser8 (no truncation), Pro10, Pro12, Ile14, Thr16 and Arg18. The reverse primers 527-MtCMi-C-M90 (5’-
CAGCCACTAGTTATTACATAGCATCCGGACCACGACCA), 528-MtCMi-C-A89 (5’-CAGCCACTAGTTATTAAAGCATCCGGACCACGACCAAGC), 529-MtCMi-C-D88 (5’-CAGCCACTAGTTATTAAATCCGGACCACGACCAAGAC), 530-MtCMi-C-P87 (5’-CAGCCACTAGTTATTACGGACAATCCGGACCACGACCAAGAC), 531-MtCMi-C-G86 (5’-CAGCCACTAGTTATTAAACGACCAAGACGACGAAGC), and 532-MtCMi-C-R85 (5’-CAGCCACTAGTTATTAAACGACCAAGACGCAAAAGC) introduced a SpeI restriction site, two stop codons and C-terminal truncations up to Met90, Ala89, Asp88, Pro87, Gly86 and Arg85 respectively.

The PCR products for the libraries LdCdNR85X and LdCdNL81X were constructed using the same template, the mixture of forward primers as described for LdNdC library, and the reverse primers 534-MtCMi-C-R85X (5’-CAGCCACTAGTTATTAMNNACCAAGACGCAAAGCAG) or 535-MtCMi-C-L81X (5’-CAGCCACTAGTTATTAAACGACCAAGACGACGMNNAAAGCAG), which introduced a SpeI restriction site, two stop codons and the randomized R85 or L81, respectively, via NNK codon. The PCR products of different sizes, depending on the truncation extent for libraries LdNdC, LdNR85X and LdNL81X, were digested with XhoI and SpeI and ligated with the 2761 bp fragment of XhoI/ SpeI digested pKTNTET-0.

5.4.4.2 Selection of truncated re4.7s11 variants

The transformed electrocompetent KA12/pKIMP-UAUC cells were plated on non-stringent selective M9c+F+Tet500ng/mL minimal medium as described in section 3.4.3. The library sizes were calculated from growth on non-selective M9c+FY plates. 5 clones that grew on these plates were sequenced to determine the quality and mutation rated of the library.

To further confirm the reduced complementation ability of the truncated variants, the clones that grew on M9c+F+Tet80ng/mL plates were picked, inoculated in 100 μL of M9 salts and drop spotted on higher stringency plates such as M9c+F+Tet80ng/mL and M9c+F+Tet40ng/mL containing 20 μg/mL L-phenylalanine and 80 or 40 ng/mL tetracycline respectively, M9c+Tet80ng/mL and M9c+Tet80ng/mL containing 80 or 40 ng/mL tetracycline respectively, M9c without
additives and M9c+pFPhe_{40} \mu M and M9c+pFPhe_{100} \mu M containing 40 or 100 \mu M toxic pFPhe, respectively. The slowest growing clones on stringent conditions were picked and sequenced with oligonucleotide 131-TERM (5'-CCCTCAAGACCCGTTTAGA).

5.4.4.3 Reengineering the CM selection system for untagged aroQ\delta genes

The acceptor vector for the untagged versions of the aroQ\delta genes was constructed by cloning the PCR fragment generated by using oligonucleotides 131-TERM (5'-CCCTCAAGACCCGTTTAGA) and 533-pKTNTET-dHis (5'-AGATATACATATGCTCGAGTCCCAACCTGTCC), and pKTNTet-0 as the template. The PCR product was restriction digested with NdeI and SpeI and the resultant 1378 bp fragment was ligated with the 2726 bp fragment of accordingly digested pKTNTet-0. The resultant 4104 bp pKTNTET-dHis-0 vector was sequenced using oligonucleotide 60-T7Pro (5'-TAATACGACTCAGTATATAGGG).

5.4.4.4 Reengineering the CM selection system for RepA tagged aroQ\delta genes

The acceptor vector for the Ls4repA library pKTNTET-RepA, which included the RepA degradation tag (MNQSFISDILYADIES), was constructed with pKTNTET-0 as the cloning vector. The first PCR product for overlap extension was generated using primers 131-TERM (5'-CCCTCAAGACCCGTTTAGA) and 538-pKTNTet-Fw-RepA (5'-TTAGCGATATCTGTATGCGGATATTGAATCCCTCGAGTCCCAACCTGT), which contained the XhoI site and excluded the Met codon after RepA sequence while the second PCR product was obtained using primers 372-TetPro (5'-AGCTCTAATGCGCTGTTAATCACT) and 539-pKTNTet-Rev-RepA (5'-ATACAGAATATCGCTAATAAAGCTCGGTTCATATATATATCTCTCCTTC) containing an NdeI site. The resulting 257 bp and 1479 bp PCR products that had 17 bp homologous overlapping sequence were assembled using external 372-TetPro and 131-TERM primers to give a 1719 bp fragment. The 1719 bp product was digested with SpeI and NdeI and the resulting 1423 bp fragment was ligated with the correspondingly digested 2726 bp fragment from pKTNTET-0 to obtain the 4149 bp pKTNTET-RepA vector.
5.4.4.5 Construction and selection of His-tagged and untagged libraries

The variants dNdCs1, dNR85Xu15, and dNR85s10 were mutagenized via two cycles of error-prone PCR. The first cycle was conducted using primers 372-TetPro (5'-AGCTCTAATGCGCTGTAATACACT) and 131-TERM (5'-CCCTCAAGACCCTTGGTTAGA) and 0.5 ng of plasmids pKTNTET-dNdCs1, pKTNTET-dNR85Xu15 or pKTNTET-dNR85s10 carrying the His-tagged versions of the target variants. 0.5 ng of the respective 546 bp, 546 bp and 576 bp PCR products were used for a second cycle using the forward primers 542-Fw-His (5'-ACTCTTCTGGTATGCTCGAG) for the His6-tagged or 543-Fw-dHis (5'-GAGATATACATATGCTCGAG) for the untagged versions, respectively and the reverse primers 547-Rv-dNdCs1 (5'-GCAGCCACTAGTTATTAACG for dNdCs1), 549-Rv-dNR85u15 (5'-GCCACTAGTTATTACTAACC for dNR85u15) and 548-Rv-dNR85s10 (5'-GCAGCCACTAGTTATTACAT, for dNR85s10). The XhoI/SpeI digested 215 bp, 215 bp and 245 bp products were cloned into correspondingly digested pKTNTET-0 or pKTNTET-dHi-0 acceptor vectors. The resultant library plasmids of tagged or untagged mutagenized dNdCs1, dNR85u15 and dNR85s10 were electroporated into KA12/pKIMP-UAUC. To determine the library sizes, quality, and mutation rates, a fraction of cells was plated on non-selective M9c+FY agar plates. The rest of the cells were plated on M9c+F+Tet80ng/mL, M9c+F+Tet40ng/mL, containing 20 μg/mL L-phenylalanine and 80 or 40 ng/mL tetracycline respectively, M9c+Tet80ng/mL, M9c+Tet40ng/mL, M9c+Tet10ng/mL, M9c+Tet5ng/mL, and M9c+Tet2.5ng/mL, containing 80, 40, 10, 5 or 2.5 ng/mL tetracycline, respectively, M9c without additives and M9c+pFPhe40 μM, and M9c+pFPhe100 μM containing 40 or 100 μM toxic pFPhe, respectively. The clones that grew fastest on stringent conditions were picked and sequenced using oligonucleotide 131-TERM.

5.4.4.6 Construction and selection of the RepA tagged library

The dNdCs4 was mutagenized via two cycles of error-prone PCR. First, the gene was amplified using primers 372-TetPro (5'-AGCTCTAATGCGCTGTAATACACT) and 131-TERM (5'-CCCTCAAGACCCTTGGTTAGA) and plasmid pKTNTET-dNdCs4 carrying the His-tagged version of the target variant. The first cycle of epPCR was conducted using 0.5 ng/μl of the amplified gene as a template and
primers 372-TetPro and 131-TERM using the Mutazyme II kit from Stratagene

0.5 ng of the respective 564 bp PCR product was used for the second cycle using the forward primer 542-Fw-His (5’-ACTCTTCTGGTATGCTCGAG) and reverse primer 547-Rv-dNdCs1 (5’-GCAGCCACTAGTTATTAACG) 259 bp fragment. The Xhol/Spel digested 233 bp product was cloned into correspondingly digested pKTNTET-dHis-0 to generate the 2993 bp library plasmid, which was electroporated into KA12/pKIMP-UAU.

To determine the library sizes a fraction of cells was plated on non-selective M9c + FY agar plates. Five clones that grew on these plates were sequenced to determine the quality and mutation rate for the library. The rest of the cells was plated on M9c+F+Tet80ng/mL, M9c+F+Tet40ng/mL, M9c+Tet80ng/mL, M9c+Tet40ng/mL, M9c+Tet10ng/mL, M9c+Tet5ng/mL, M9c, M9c+pFpH40 μM and M9c+pFpH100 μM for selection.

After incubation for 3 days at 30 °C, well-growing colonies were picked on the most stringent plates and were streaked out on M9c agar plates. 192 clones that grew after the streak out on M9c plates were inoculated into 100 μL liquid M9c+FY medium in two 96-well plates and incubated overnight. Eight pools of 24 different clones grown in 96-well plates were collected and the plasmid DNA was isolated. The target gene was amplified from each of the eight pools with oligonucleotides 543-Fw-dHis (5’-GAGATATACATATGCTCGAG) and 547-Rv-dNdCs1 (5’-GCAGCCACTAGTTATTAACG). The PCR products were gel-purified, pooled to have 800 ng/uL of each fragment, and subjected to DNA shuffling as described in section 5.4.3.4. The DNAse digested products were assembled via PCR using oligonucleotides 543-Fw-dHis and 547-Rv-dNdCs1, digested with Xhol and SpeI, and 233 bp fragment was cloned into the accordingly digested pKTNTet-RepA vector (see section 5.4.4.4) to generate 3009 bp library plasmid. After transformation into electrocompetent KA12/pKIMP-UAU cells, the resultant Ls4RepA library plasmid was plated on M9c+F+Tet80ng/mL, M9c+F+Tet40ng/mL, M9c+Tet80ng/mL, M9c+Tet40ng/mL, and M9c plates for selection and on M9c+FY plates for determination of library size, quality and mutation rates.
Two 96-well plates were inoculated with the best complementing clones on highest stringency plates. Again, eight pools of amplified clones were prepared, plasmid DNA was extracted, and the genes were amplified by PCR using oligonucleotides 544-Fw-RepA (5'-GCGGATATTGAATCCCTCGAG) and 547-Rv-dNdcS1. The PCR products were shuffled, assembled and cloned into the pKTNTet-RepA vector as described above for the Ls4RepA library. The resultant Ls4RepA2 library was plated on M9c, M9c+pFPHe40 μM and M9c+pFPHe100 μM agar plates for selection and on M9c + FY agar plates for determination of library size, quality, and mutation rates. 150 clones that complemented on high stringency plates were purified by streaking them out on M9c+pFPHe100 μM agar plates. The best growing clones were sequenced using oligonucleotide 131-TERM.

5.4.4.7 Elongation of evolved truncated variants

The individual genes of evolved, truncated and RepA tagged variants on pKTNTET-based plasmids from libraries LdCdNs1, LdNR85Xs10, and Ls4RepA2 were amplified using an appropriate primer pair (see Table VI in Appendix I). The PCR fragment was digested with SpeI/XhoI and the 260 bp insert was ligated with the 2761 bp fragment of appropriately digested pKTNTET-0 vector.

5.4.5 Production and purification of His6-tagged proteins

N-terminally His6-tagged, pKINT-based MtCM variants were produced in KA12/pAC-Ptet-T7pol-S cells. The cells were grown in 500 mL LB medium containing 150 μg/mL Na-ampicillin and 30 μg/mL chloramphenicol at 30°C. The gene expression was induced with 2 μg/mL tetracycline when an OD600 of 0.3–0.5 was reached. The crude lysate was obtained as described before.[208] (but without the RNaseA and DNase I treatment), and provided with 10 mM imidazole. The sample was loaded onto an equilibrated Ni-NTA column with (His-Select Nickel Affinity Gel (Sigma)). The MtCM variant was eluted with 250 mM imidazole in 50 mM sodium phosphate, pH 8 buffer, containing 0.3 M NaCl, and dialyzed against 20 mM potassium phosphate, pH 7.5. Protein concentration was determined by a calibrated Bradford assay[48] using BSA as a standard. Proteins were assessed by SDS-PAGE using the PhastSystem (20 % homogeneous gels, GE Healthcare) and
the molecular mass was determined by ESI MS at Mass Spectrometry Service at the Laboratory of Organic Chemistry, ETH Zurich.

5.4.6 Construction of the evolved variants containing the wild-type N-terminus

To construct evolved variants with the wild-type (untagged) N-terminus, the required gene fragment of the wild-type MtCM gene of plasmid pKTCMM-H[48] was replaced by the appropriate fragment of the evolved variant. The pKTNTET-based plasmids with the genes of the evolved variants re4.7s11, s1es5.1 and s10es4.15 were restriction digested with XhoI and SpeI to obtain the required 260 bp fragments. These fragments were cloned into the accordingly digested plasmid pKTCMM-H[48].

5.4.7 Production and purification of untagged proteins

The untagged versions of re4.7s11 (predicted pl 6.11), s1es5.1 (predicted pl 6.11), and s10es4.15 (predicted pl 6.09) on a p-KTCMM-H plasmid were produced in strain KA13. The cells were grown in LB medium containing 150 μg/mL Na-ampicillin. At an OD$_{600}$ of 0.3-0.6 the gene expression was initiated by 0.5mM IPTG and continued over night at 37 ºC. The cells were supplied with lysozyme and incubated for 1 hour on ice prior to disruption by sonication. The crude cell lysate was supplemented with 65-70 % (w/v) ammonium sulfate and stirred for 1.5 -2 hours at 4 ºC. The pellet collected by centrifugation was dissolved in 20 mM piperazine, pH 9.0 buffer and dialyzed against the same buffer overnight at 4 ºC. Dialyzed samples were applied onto a pre-equilibrated MonoQ column in buffer A (20 mM piperazine, pH 9.0). The gradient from 0 % to 30 % of buffer B (20 mM piperazine, 1M NaCl pH 9) was applied over 50 -100 mL at a flow rate of 2 mL/min.

The fractions from the MonoQ column were collected, concentrated and dialyzed against 20 mM MES, pH 5.5 buffer. After dialysis, the sample was applied onto a pre-equilibrated MonoS column in buffer A (20 mM MES, pH5.5). The loaded column was first washed with 95 % buffer A, 5 % buffer B (20 mM MES, pH 5.5, 1M NaCl). Then protein were eluted with gradient from 0 % to 40 % of buffer B (20 mM MES, pH 5.5 containing 1M NaCl) over 50 -100 mL, at a flow rate of 2 mL/min.
Finally, the fractions from the MonoS column were pooled, concentrated and applied on a Superdex 75 column. Proteins were eluted in 20 mM BTP, pH 7.5, containing 150 mM NaCl. Roughly 1 mg of pure MtCM per litre of culture was obtained following this procedure. Protein integrity was analyzed by SDS-PAGE and ESI MS at Mass Spectrometry Service at the Laboratory of Organic Chemistry, ETH Zurich.

5.4.8 Enzymatic assays

Steady-state kinetics for AroQ8 variants (with or without the His6-tag and in the absence of MtDS) was performed at 30 °C in 50 mM potassium phosphate, pH 7.5, by varying chorismate concentrations between 100 and ~2000 μM. The initial rates were acquired by fitting a linear function to the absorption decrease curves that were obtained by measurements of chorismate disappearance at 310 nm (ε\textsubscript{310} = 370 M\textsuperscript{-1} cm\textsuperscript{-1}). The data was fitted to the Michaelis-Menten equation using KaleidaGraph (Synergy Software). The kinetic characterization of the AroQ8 variants in the presence of MtDS was performed as described in section 2.4.3.

5.4.9 Circular dichroism spectroscopy

Thermal denaturation curves were determined using circular dichroism measurements at 222 nm (band width was 1 nm) using an Aviv 202 CD spectrometer (Aviv Biomedical). Approximately 4 μM protein in 20 mM potassium phosphate, pH 7.5 was used in quartz cuvettes with a 0.2 cm path length.

After initial equilibration at 10 or 20 °C for 5 min, the sample was heated in 0.2 °C steps with 0.4 min equilibration time and 3 s signal averaging time per point. The melting temperature was determined between 10 or 20 °C and to 95 °C.

5.4.10 Crystallographic data processing and structure refinement

The crystallization and data collection was performed by Helen Vikdal Thorbjørnsrud (University of Oslo) at European Synchrotron Radiation Facility (Grenoble, France). The provided data sets were processed using XDS\textsuperscript{[226]} Phaser MR (from the CCP4 suite\textsuperscript{[227]}) was employed for molecular replacement using the wild-type MtCM
structure (PDB ID: 2W19) as a model. The structures were determined by rounds of rigid body refinement in Refmac5\textsuperscript{[228,229]} (from the CCP4 suite\textsuperscript{[227]}). Model building was undertaken in Coot\textsuperscript{[230]} and the structures were further refined with Refmac5.
6 CONCLUSIONS AND OUTLOOK

This thesis presents new details about the function and regulation of AroQδ CMs that are found in Corynebacteriales\textsuperscript{[28]} featuring several human pathogens and commercially important organisms.

CMs catalyze the first committed step in the biosynthesis of Phe and Tyr and are subjects to tight feedback regulation. Interestingly, AroQδ CMs do not have effector binding sites and feature a unique inter-enzyme allosteric regulation that requires complex formation with the first shikimate pathway enzyme DAHP synthase.

Regulation of MtCM from \textit{M. tuberculosis} relies on its ability to switch between the unactivated ($k_{\text{cat}}/K_m = 1.75 \times 10^3$ M\textsuperscript{-1} s\textsuperscript{-1}) and activated ($k_{\text{cat}}/K_m = 2.4 \times 10^5$ M\textsuperscript{-1} s\textsuperscript{-1}) forms governed by the transient interaction with MtDS\textsuperscript{[48]}. Phe, Tyr, and Trp bind to dedicated sites on MtDS\textsuperscript{[29, 28]} and affect the equilibrium of MtCM-MtDS complex formation.

An analogous protein complex exists in the amino acid producer \textit{C. glutamicum}. In this work we showed a 180-fold activation of CgCM by CgDS and similar regulatory effects of the aromatic amino acids as for the MtCM-MtDS complex.

In Chapter 2, we revealed that Tyr and Phe are allosteric inhibitors, Trp can act as an activator and the presence of all three aromatic amino acids severely inhibits CM activity of both protein
complexes. Using the MtCM-MtDS complex as a model, we show that inhibition is achieved via facilitated dissociation of the complex.

Interestingly, Trp activation is only seen under conditions where protein concentrations are below the $K_{d,\text{app}}$ of the CM-DS complex, or the DS tetramer. Therefore, keeping in mind the relatively high concentrations of these proteins in the cells (3 μM DS, and 300 nM CM, see section 2.3), we hypothesize that due to different $K_{d,\text{app}}$ values (see below) of the CgCM-CgDS and the MtCM-MtDS complexes, Trp activation might be only utilized in *C. glutamicum* but not in *M. tuberculosis*.

The fact that Trp binds at the interface of DS dimers (Chapter 2, Figure 2.3) and activation is only observed at low protein concentration, suggests that Trp induces DS tetramer formation, which creates a platform for CM binding, resulting in CM activity increase. This assumption requires more experimental evidence, which could be provided by a SEC experiment. The DS association induced by Trp can be tested in a similar manner as for the ligand-induced dissociation of the MtCM-MtDS complex, described in Chapter 2. It is important to keep in mind that protein concentrations below the $K_{d,\text{app}}$ are required to observe this phenomenon. The $K_{d,\text{app}}$ of the CgCM-CgDS complex is $> 4$ μM (this work) whereas the $K_{d,\text{app}}$ of MtCM-MtDS is 140 nM; therefore, the protein complex from *C. glutamicum* should serve as a better-suited model due to the concentration requirements for a SEC experiment.

Finally, CgCM exhibits a high $K_m$ of $> 2$ mM, even when activated by CgDS. This is surprising, since metabolic enzymes typically evolve to maintain $K_m$ values in the range of intracellular substrate concentrations,[99] which for chorismate is 40 - 70 μM.[147, 148] However, it might be that CgCM was not fully activated under our assay conditions and requires the presence of Trp for full activity. If our hypothesis that Trp stabilizes the DS tetramer is correct, determination of the kinetic parameters of the CgCM-CgDS complex in the presence of Trp should reveal the maximum substrate affinity of CgCM.

Another goal of this thesis was to identify potential inhibitors of the CMs from *M. tuberculosis*. We used the MtCM-MtDS complex[48] and
Regulation, effectors, and evolvability of chorismate mutases from *Corynebacteriales*

*MtCM (secreted)*\(^{[36]}\) as targets for a small molecule screen using the compound library provided by the FMP Berlin.\(^{[161]}\) At this point, we were not able to confirm any molecules that would act on target CMs, even though many compounds showed promising inhibition profiles when tested in the HTS format using the coupled assay (see Chapter 3).

Judging from the chemical structures of the selected compounds, none of them seemed likely to enter the small and charged active sites of CMs. It is possible, that the compound library used in this experiment was not suitable for these enzymes. Therefore, it is advisable to test different compound libraries that are enriched in small and charged molecules like Bartlett’s TSA,\(^{[33]}\) which remains the best known general CM inhibitor so far.

Moreover, none of the potential small molecule inhibitors found for the MtCM-MtDS assembly seemed to affect complex formation. This is not surprising, since small molecules do not contain enough functional groups for tight interaction with the large and mostly featureless surfaces of PPIs. Therefore, in Chapter 4, a novel CM-deficient *E. coli*-based\(^{[51]}\) selection system was described that would allow identification of head-to-tail cyclized\(^{[175]}\) peptide modulators of CMs. The MtCM-MtDS complex serves as a good model for development of such a system, since it can report direct inhibition of MtCM and disruption of the MtCM-MtDS complex, both resulting in a CM activity change.

We envisioned that our *in vivo* selection system could be used to identify activators as well as inhibitors of CMs, depending on the applied selection strategy. For example, growing the CP producing cells at a high stringency regime could be utilized to identify activators. On the other hand, inhibitors can be isolated by application of penicillin enrichment \(^{[184]}\), which kills the growing cells producing active CMs. However, the preliminary tests performed using both selection strategies did not result in the identification of promising CP modulators so far.

Nevertheless, judging from the identified artifacts, we assume that the proposed selection strategies could in principle work as soon as the experimental issues are resolved. Most importantly, the intracellular CM activity must be reduced to improve the selection for
CP activators and inhibitors. Our selection experiment for activators resulted in identification of CPs that did not have any effect when tested \textit{in vitro}. This suggests that the selection stringency could have been too weak, not suppressing the growth of clones producing inactive CPs. In the case of the penicillin selection, the enrichment for β-lactamase producing artifacts could indicate that the activity of MtCM, even when the complex with MtDS is disrupted, is high enough to allow cell growth finally resulting in penicillin-induced lysis. In both cases, the reduction of basal intracellular CM activity is necessary to increase the stringency, which could be achieved by exchanging the promoter system, using a low copy number plasmid or even constructing a strain that carries the target genes in its chromosomal DNA.

Moreover, it is expected that genuine hits from libraries containing \textit{a priori} high fraction of inhibitors or activators would quickly dominate the cell cultures and thereby reduce the probability of identifying artifacts. Therefore, testing different libraries encoding larger peptides or allowing incorporation of unnatural amino acids\cite{173} could also be considered.

In conclusion, even though the selection system for peptidic modulators of the MtCM-MtDS complex requires further development, we are confident that the presented system holds a great promise in being generally applicable for many metabolic enzymes.

In Chapter 5 we investigated the mechanism by which MtCM achieves high catalytic efficiency. It has been hypothesized, that activation of MtCM occurs due to subtle conformational changes in the flexible H1-H2 loop and at the C-terminus induced by the interaction partner MtDS.\cite{48} These structural modifications would rearrange the active site residues of the wild-type enzyme, and presumably lower the substrate binding energy. In this thesis we employed directed evolution to test whether certain mutations could ‘lock’ MtCM in the activated state without the help of the interaction partner. By this, we aimed to reveal mechanism of the activity switch and the potential of MtCM to achieve high autonomous activity.

The main challenge in evolving highly active CM catalysts is adjusting the selection window that allows distinguishing between
active and very active enzymes. In our case, we employed the destabilization-compensation strategy to first reduce the catalytic activity of the parent mutant and then evolved the impaired variants. The impact of the newly accumulated mutations was then tested after the destabilizing features of the evolved intermediates were removed. We observed that most of the mutations acquired in the destabilized proteins were also beneficial when the crippling modifications were removed. In this way, we were able to obtain MtCM variant s10es4.15 ($k_\text{cat}/K_m = 4.7 \times 10^5$ M$^{-1}$ s$^{-1}$), which was nearly two times more active, that MtDS-activated wild-type enzyme, and variant s1es5.1 ($k_\text{cat}/K_m = 2.5 \times 10^5$ M$^{-1}$ s$^{-1}$), which catalyzed the reaction at the same efficiency. Even though, the activity of these mutants is still 2-3 orders of magnitude lower than the maximum possible diffusion-limited catalysis ($k_\text{cat}/K_m$ of $10^{8-10^9}$ M$^{-1}$ s$^{-1}$) they have reached the same activity level as the best natural CMs ($k_\text{cat}/K_m$ of $10^5-10^6$ M$^{-1}$ s$^{-1}$).[139, 36]

In fact, the s10es4.15 and s1es5.1 variants are the most active CMs known of the AroQ$_\delta$ subclass and they provide insight in how these enzymes achieve their high catalytic efficiency despite lacking several crucial active site residues, conserved in CMs from other AroQ subclasses.

The crystal structures of the evolved s10es4.15 and s1es5.1 variants revealed that introduction of mutations Thr52Pro and Val55Asp induced a similar conformation in the H1-H2 flexible loop as observed in the MtDS-activated wild-type MtCM, and also in the highly active autonomous EcCM (see Chapter 5, figure 5.7). This loop conformation stabilizes the active site residue Arg46 (Lys39 in EcCM), which interacts with the ether oxygen of Bartlett’s TSA in the ligand-bound wild-type MtCM,[48] in a more catalytically competent conformation. Therefore, we hypothesize that in addition to dimer stabilization, a better positioning of residue Arg46 is a key feature for the efficient catalysis by s10es4.15 and s1es5.1.

However, the active site residue Glu59 in the evolved variants was found to adopt conformation that would preclude binding of TSA in the same manner as observed in the crystal structure of the TSA-bound wild-type enzyme.[48] However, the peculiar conformation of this residue could be the consequence of crystal contacts observed in the space group P6$_4$. Therefore, a new screen of crystallization conditions
needs to be found in order to identify the crystals of TSA-bound s10es4.15 and s1es5.1. The crystal structures of ligand-bound MtCM variants are desired to elucidate the precise positioning of the active site residues in these highly active AroQ<sub>s</sub> enzymes.
7 APPENDICES
### Table A. I. Kinetic parameters of evolved inter-subunit destabilized MtCM variants

<table>
<thead>
<tr>
<th>6His Protein</th>
<th>Destabilizing mutations</th>
<th>Parental mutations</th>
<th>New mutations</th>
<th>$k_{cat}^a$ (s$^{-1}$)</th>
<th>$K_m^a$ (μM)</th>
<th>$k_{cat}/K_m^a \times 10^4$ (M$^{-1}$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1.13</td>
<td>L24E</td>
<td>PD/PDM</td>
<td>-</td>
<td>10</td>
<td>880</td>
<td>1.2</td>
</tr>
<tr>
<td>ep1.13s1</td>
<td>L24E</td>
<td>PD/PDM</td>
<td>G43V</td>
<td>3.6</td>
<td>42</td>
<td>8.4</td>
</tr>
<tr>
<td>ep1.13s4</td>
<td>L24E</td>
<td>PD/PDM</td>
<td>K44E</td>
<td>2.5</td>
<td>24</td>
<td>10</td>
</tr>
<tr>
<td>sh1.13s2</td>
<td>L24E</td>
<td>PD/PDM</td>
<td>V38L</td>
<td>ND</td>
<td>&gt;1500</td>
<td>1.5</td>
</tr>
<tr>
<td>sh1.13s13</td>
<td>L24E</td>
<td>PD/PDM</td>
<td>G43S</td>
<td>ND</td>
<td>&gt;3000</td>
<td>0.4</td>
</tr>
<tr>
<td>1-3</td>
<td>L24E</td>
<td>PD/PDM</td>
<td>A45V, D72Y, G89D</td>
<td>27</td>
<td>500</td>
<td>5.4</td>
</tr>
<tr>
<td>1-14</td>
<td>L24E</td>
<td>PD/PDM</td>
<td>E37K, G84C</td>
<td>17</td>
<td>170</td>
<td>10</td>
</tr>
<tr>
<td>2-2</td>
<td>L24E</td>
<td>PD/PDM</td>
<td>V38L</td>
<td>ND</td>
<td>&gt;1600</td>
<td>2.1</td>
</tr>
<tr>
<td>2-20</td>
<td>L24E</td>
<td>PD/PDM</td>
<td>G43A</td>
<td>ND</td>
<td>&gt;1300</td>
<td>1.7</td>
</tr>
<tr>
<td>L4.4</td>
<td>L24G, L31R</td>
<td>PD/PDM</td>
<td>-</td>
<td>ND</td>
<td>&gt;1900</td>
<td>0.05</td>
</tr>
<tr>
<td>ep4.4s1</td>
<td>L24G, L31R</td>
<td>PD/PDM</td>
<td>G84C</td>
<td>22</td>
<td>670</td>
<td>3.3</td>
</tr>
<tr>
<td>ep4.4s8</td>
<td>L24G, L31R</td>
<td>PD/PDM</td>
<td>D75E</td>
<td>ND</td>
<td>&gt;1400</td>
<td>3.4</td>
</tr>
<tr>
<td>3.4-s1</td>
<td>L24G, L31R</td>
<td>PD/PDM</td>
<td>Y66F, G84C</td>
<td>24</td>
<td>942</td>
<td>2.5</td>
</tr>
<tr>
<td>3-8</td>
<td>L24G, L31R</td>
<td>PD/PDM</td>
<td>D72E, G84C</td>
<td>21</td>
<td>160</td>
<td>1.3</td>
</tr>
<tr>
<td>4-s2</td>
<td>L24G, L31R</td>
<td>PD/PDM</td>
<td>G84C, G89D</td>
<td>35</td>
<td>660</td>
<td>5.3</td>
</tr>
<tr>
<td>4-e3</td>
<td>L24G, L31R</td>
<td>PD/PDM</td>
<td>G43V, G89D</td>
<td>4.2</td>
<td>124</td>
<td>3.4</td>
</tr>
<tr>
<td>L4.7</td>
<td>L24T, L31S</td>
<td>PD/PDM</td>
<td>-</td>
<td>28</td>
<td>550</td>
<td>5.0</td>
</tr>
<tr>
<td>ep4.7s11</td>
<td>L24T, L31S</td>
<td>PD/PDM</td>
<td>V62I, D72V</td>
<td>47</td>
<td>180</td>
<td>27</td>
</tr>
<tr>
<td>ep4.7s13</td>
<td>L24T, L31S</td>
<td>PD/PDM</td>
<td>D72E, G84C</td>
<td>22</td>
<td>248</td>
<td>8.9</td>
</tr>
<tr>
<td>sh4.7s2</td>
<td>L24T, L31S</td>
<td>PD/PDM</td>
<td>K40Q</td>
<td>ND</td>
<td>&gt;2300</td>
<td>1.5</td>
</tr>
<tr>
<td>sh4.7s3</td>
<td>L24T, L31S</td>
<td>PD/PDM</td>
<td>V62I</td>
<td>18</td>
<td>209</td>
<td>8.8</td>
</tr>
<tr>
<td>5-8</td>
<td>L24T, L31S</td>
<td>PD/PDM</td>
<td>K44E, G84C</td>
<td>16</td>
<td>390</td>
<td>4.1</td>
</tr>
<tr>
<td>5-15</td>
<td>L24T, L31S</td>
<td>PD/PDM</td>
<td>K44M, D72E</td>
<td>ND</td>
<td>&gt;1600</td>
<td>2.3</td>
</tr>
<tr>
<td>L3.6</td>
<td>L24W, L31T</td>
<td>PD/PDM</td>
<td>-</td>
<td>21</td>
<td>430</td>
<td>4.8</td>
</tr>
<tr>
<td>sh3.6s22</td>
<td>L24W, L31T</td>
<td>PD/PDM</td>
<td>K44E, D72E, G84C</td>
<td>8.5</td>
<td>190</td>
<td>4.5</td>
</tr>
<tr>
<td>ep3.6s23</td>
<td>L24W, L31T</td>
<td>PD/PDM</td>
<td>G43V, S67N, I78N</td>
<td>2.5</td>
<td>93</td>
<td>2.7</td>
</tr>
<tr>
<td>sh3.6s25</td>
<td>L24W, L31T</td>
<td>PD/PDM</td>
<td>V62I, G84C, G89D</td>
<td>2.5</td>
<td>57</td>
<td>4.4</td>
</tr>
<tr>
<td>ep3.6s25</td>
<td>L24W, L31T</td>
<td>PD/PDM</td>
<td>G43V, H90T</td>
<td>1.2</td>
<td>30</td>
<td>4.1</td>
</tr>
<tr>
<td>ep3.6s26</td>
<td>L24W, L31T</td>
<td>PD/PDM</td>
<td>K44Q, V62I</td>
<td>3.5</td>
<td>120</td>
<td>2.9</td>
</tr>
<tr>
<td>ep3.6s27</td>
<td>L24W, L31T</td>
<td>PD/PDM</td>
<td>S57T, G84C</td>
<td>2.3</td>
<td>380</td>
<td>0.6</td>
</tr>
<tr>
<td>L3.7</td>
<td>L24R, L31R</td>
<td>PD/PDM</td>
<td>-</td>
<td>ND</td>
<td>&gt;1700</td>
<td>0.2</td>
</tr>
<tr>
<td>ep3.7s6</td>
<td>L24R, L31R</td>
<td>PD/PDA</td>
<td>V62I, G84C, H901</td>
<td>0.002</td>
<td>140</td>
<td>0.001</td>
</tr>
<tr>
<td>sh3.7s11</td>
<td>L24R, L31R</td>
<td>PD/PDA</td>
<td>A45V, D72E, G84C</td>
<td>13</td>
<td>220</td>
<td>6.1</td>
</tr>
<tr>
<td>L3.8</td>
<td>L24Q, L31G</td>
<td>PD/PDM</td>
<td>-</td>
<td>ND</td>
<td>&gt;1300</td>
<td>1.1</td>
</tr>
<tr>
<td>sh3.8s6</td>
<td>L24Q, L31G</td>
<td>PD/PDM</td>
<td>A45V, D72E, G84C</td>
<td>1.1</td>
<td>54</td>
<td>2.1</td>
</tr>
<tr>
<td>sh3.8s8</td>
<td>L24Q, L31G</td>
<td>PD/PDM</td>
<td>G43V, G89D</td>
<td>0.9</td>
<td>40</td>
<td>2.3</td>
</tr>
<tr>
<td>sh3.8s11</td>
<td>L24Q, L31G</td>
<td>PD/PDM</td>
<td>A45V, D72Y, G84S</td>
<td>2.6</td>
<td>59</td>
<td>4.4</td>
</tr>
<tr>
<td>sh3.8s12</td>
<td>L24Q, L31G</td>
<td>PD/PDM</td>
<td>K44E, D72Y, G89D</td>
<td>5.1</td>
<td>270</td>
<td>1.9</td>
</tr>
</tbody>
</table>

Footnotes next page
Footnotes to the Table A.I.

\(^a\)CM activity was determined in 50 mM potassium phosphate buffer, pH 7.5, containing 0.1 mg/mL BSA. Disappearance of chorismate was monitored at 310 nm at 30°C varying concentrations between 100 and 2000 μM. Kinetic parameters were derived by fitting initial velocities to the Michaelis-Menten equation. ND: not determined.

ND: not determined, since substrate saturation could not be achieved using up to 2 mM chorismate. Measuring at higher substrate concentrations was not possible due to the high chorismate absorption of the substrate at 310 nm and the relatively high background reaction.

PD indicates mutations T52P and V55D.

PDAM indicates mutations R87P, L88D, G89A, H90M.

Table A. II. Kinetic parameters of evolved inter-subunit destabilized MtCM variants after removal of destabilizing mutations

<table>
<thead>
<tr>
<th>6His Protein</th>
<th>Destabilizing mutations</th>
<th>Parental mutations</th>
<th>New mutations</th>
<th>(k_{\text{cat}}^a) (s(^{-1}))</th>
<th>(K_m^a) (μM)</th>
<th>(k_{\text{cat}}/K_m^a \times 10^4) (M(^{-1}) s(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>re1.13s1</td>
<td>-</td>
<td>PD/PDAM</td>
<td>G43V</td>
<td>6.8</td>
<td>48</td>
<td>14</td>
</tr>
<tr>
<td>re1-3</td>
<td>-</td>
<td>PD/PDM</td>
<td>A45V, D72Y, G89D</td>
<td>29</td>
<td>180</td>
<td>16</td>
</tr>
<tr>
<td>re1-14</td>
<td>-</td>
<td>PD/PDAM</td>
<td>E37K, G84C</td>
<td>25</td>
<td>250</td>
<td>9.8</td>
</tr>
<tr>
<td>re4.4s1</td>
<td>-</td>
<td>PD/PDAM</td>
<td>G84C</td>
<td>23</td>
<td>470</td>
<td>4.9</td>
</tr>
<tr>
<td>re4.4s8</td>
<td>-</td>
<td>PD/PDAM</td>
<td>D75E</td>
<td>31</td>
<td>460</td>
<td>6.8</td>
</tr>
<tr>
<td>re3.4-s1</td>
<td>-</td>
<td>PD/PDAM</td>
<td>Y66F, G84C</td>
<td>15</td>
<td>340</td>
<td>4.4</td>
</tr>
<tr>
<td>re4-s2</td>
<td>-</td>
<td>PD/PDAM</td>
<td>G84C, G89D</td>
<td>11</td>
<td>140</td>
<td>8.1</td>
</tr>
<tr>
<td>re4-e3</td>
<td>-</td>
<td>PD/PDAM</td>
<td>G43V, G89D</td>
<td>6.9</td>
<td>230</td>
<td>3.0</td>
</tr>
<tr>
<td>re4.7s11</td>
<td>-</td>
<td>PD/PDAM</td>
<td>V62I, D72V</td>
<td>23</td>
<td>72</td>
<td>32</td>
</tr>
<tr>
<td>re4.7s13</td>
<td>-</td>
<td>PD/PDAM</td>
<td>D72E, G84C</td>
<td>20</td>
<td>280</td>
<td>6.9</td>
</tr>
<tr>
<td>re4.7s3</td>
<td>-</td>
<td>PD/PDAM</td>
<td>V62I</td>
<td>65</td>
<td>440</td>
<td>15</td>
</tr>
<tr>
<td>re3.6s22</td>
<td>-</td>
<td>PD/PDAM</td>
<td>K44E, D72E, G84C</td>
<td>14</td>
<td>160</td>
<td>8.8</td>
</tr>
<tr>
<td>re3.6s23</td>
<td>-</td>
<td>PD/PDAM</td>
<td>G43V, S67N, I78N</td>
<td>7.6</td>
<td>52</td>
<td>15</td>
</tr>
<tr>
<td>re3.6s25</td>
<td>-</td>
<td>PD/PDAM</td>
<td>V62I, G84C, G89D</td>
<td>3.4</td>
<td>37</td>
<td>9.2</td>
</tr>
<tr>
<td>re3.6s25</td>
<td>-</td>
<td>PD/PDAM</td>
<td>G43V, H90T</td>
<td>5.6</td>
<td>24</td>
<td>23</td>
</tr>
<tr>
<td>re3.6s26</td>
<td>-</td>
<td>PD/PDAM</td>
<td>K44Q, V62I</td>
<td>21</td>
<td>120</td>
<td>17</td>
</tr>
<tr>
<td>re3.6s27</td>
<td>-</td>
<td>PD/PDAM</td>
<td>S57T, G84C</td>
<td>9.7</td>
<td>130</td>
<td>7.6</td>
</tr>
<tr>
<td>re3.7s6</td>
<td>-</td>
<td>PD/PDAM</td>
<td>V62I, G84C, H90I</td>
<td>5.7</td>
<td>46</td>
<td>12</td>
</tr>
<tr>
<td>re3.7s11</td>
<td>-</td>
<td>PD/PDAM</td>
<td>A45V, D72E, G84C</td>
<td>6.5</td>
<td>70</td>
<td>9.3</td>
</tr>
<tr>
<td>re3.8s6</td>
<td>-</td>
<td>PD/PDAM</td>
<td>A45V, D72E, G84C</td>
<td>4.3</td>
<td>100</td>
<td>4.3</td>
</tr>
<tr>
<td>re3.8s8</td>
<td>-</td>
<td>PD/PDAM</td>
<td>G43V, G89D</td>
<td>3.4</td>
<td>34</td>
<td>10</td>
</tr>
<tr>
<td>re3.8s11</td>
<td>-</td>
<td>PD/PDAM</td>
<td>A45V, D72Y, G84S</td>
<td>13</td>
<td>91</td>
<td>14</td>
</tr>
<tr>
<td>re3.8s12</td>
<td>-</td>
<td>PD/PDAM</td>
<td>K44E, D72Y, G89D</td>
<td>21</td>
<td>140</td>
<td>15</td>
</tr>
</tbody>
</table>

\(^a\)See footnote \(^a\) of Table A. I.

PD and PDAM are explained in the footnote of Table A.I.
<table>
<thead>
<tr>
<th>6His Protein</th>
<th>Truncation</th>
<th>Parental mutations</th>
<th>New mutations</th>
<th>$k_{\text{cat}}$\textsuperscript{a} (s\textsuperscript{-1})</th>
<th>$K_{m}$\textsuperscript{a} (µM)</th>
<th>$k_{\text{cat}}/K_{m}$\textsuperscript{a} $\times 10^4$ (M\textsuperscript{-1} s\textsuperscript{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>dNR85Xs10</td>
<td>NΔ0, CΔ5</td>
<td>PD/V62I, D72V</td>
<td>ND</td>
<td>&gt;3600</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>s10es5.12</td>
<td>NΔ0, CΔ5</td>
<td>PD/V62I, D72V</td>
<td>A36T</td>
<td>ND</td>
<td>&gt;1500</td>
<td>0.2</td>
</tr>
<tr>
<td>s10es6.12</td>
<td>NΔ0, CΔ5</td>
<td>PD/V62I, D72V</td>
<td>K33R</td>
<td>ND</td>
<td>&gt;1600</td>
<td>0.4</td>
</tr>
<tr>
<td>s10es6.15</td>
<td>NΔ0, CΔ5</td>
<td>PD/V62I, D72V</td>
<td>A30V</td>
<td>ND</td>
<td>&gt;3500</td>
<td>0.3</td>
</tr>
<tr>
<td>1p11</td>
<td>NΔ0, CΔ5</td>
<td>PD/V62I, D72V</td>
<td>K40M, L81M</td>
<td>4.3</td>
<td>420</td>
<td>1.0</td>
</tr>
<tr>
<td>2p13</td>
<td>NΔ0, CΔ5</td>
<td>PD/V62I, D72V</td>
<td>P12S, E37A, L81M</td>
<td>12</td>
<td>230</td>
<td>5.4</td>
</tr>
<tr>
<td>3p15</td>
<td>NΔ0, CΔ5</td>
<td>PD/V62I, D72V</td>
<td>E37K, L81M</td>
<td>3.7</td>
<td>180</td>
<td>2.0</td>
</tr>
<tr>
<td>5p12</td>
<td>NΔ0, CΔ5</td>
<td>PD/V62I, D72V</td>
<td>K40N, I78N</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>6p11</td>
<td>NΔ0, CΔ5</td>
<td>PD/V62I, D72V</td>
<td>K40R, A41S, L81M</td>
<td>4</td>
<td>230</td>
<td>1.7</td>
</tr>
<tr>
<td>dNdCs1</td>
<td>NΔ10, CΔ5</td>
<td>PD/V62I, D72V</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>0.01</td>
</tr>
<tr>
<td>s1es4</td>
<td>NΔ10, CΔ5</td>
<td>PD/V62I, D72V</td>
<td>G43V</td>
<td>ND</td>
<td>ND</td>
<td>0.05</td>
</tr>
<tr>
<td>s1es5.1</td>
<td>NΔ10, CΔ5</td>
<td>PD/V62I, D72V</td>
<td>A41V</td>
<td>ND</td>
<td>ND</td>
<td>0.02</td>
</tr>
<tr>
<td>s10es6.17</td>
<td>NΔ10, CΔ5</td>
<td>PD/V62I, D72V</td>
<td>K40E</td>
<td>ND</td>
<td>&gt;1500</td>
<td>0.01</td>
</tr>
</tbody>
</table>

\textsuperscript{a}See footnote \textsuperscript{a} of Table A. I.
ND and PD are explained in the footnote of Table A.I.
NΔx indicates the number of amino acids (x) deleted at the N-terminus
CΔx indicates the number of amino acids (x) truncated at the C-terminus
Regulation, effectors, and evolvability of chorismate mutases from *Corynebacteriales*

Table A. IV. Kinetic parameters of evolved re-elongated MtCM variants

<table>
<thead>
<tr>
<th>6His Protein</th>
<th>Parental mutations</th>
<th>New mutations</th>
<th>$k_{\text{cat}}$ a (s$^{-1}$)</th>
<th>$K_{\text{m}}$ a (μM)</th>
<th>$k_{\text{cat}}/K_{\text{m}}$ a ×10$^4$ (M$^{-1}$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>re4.7s11</td>
<td>PD/V62I, D72V/PDAM</td>
<td></td>
<td>23</td>
<td>72</td>
<td>32</td>
</tr>
<tr>
<td>s10es2.12</td>
<td>PD/V62I, D72V/PDAM</td>
<td>R34G, A45V</td>
<td>21</td>
<td>62</td>
<td>33</td>
</tr>
<tr>
<td>s10es2.14</td>
<td>PD/V62I, D72V/PDAM</td>
<td>E37K</td>
<td>35</td>
<td>140</td>
<td>23</td>
</tr>
<tr>
<td>s10es3.13</td>
<td>PD/V62I, D72V/PDAM</td>
<td>A41V</td>
<td>6</td>
<td>82</td>
<td>7.3</td>
</tr>
<tr>
<td>s10es4</td>
<td>PD/V62I, D72V/PDAM</td>
<td>L17M, V38A</td>
<td>27</td>
<td>92</td>
<td>29</td>
</tr>
<tr>
<td>s10es4.15</td>
<td>PD/V62I, D72V/PDAM</td>
<td>V11L, D15V, K40Q</td>
<td>13</td>
<td>30</td>
<td>47</td>
</tr>
<tr>
<td>s10es5.12</td>
<td>PD/V62I, D72V/PDAM</td>
<td>A36T</td>
<td>27</td>
<td>100</td>
<td>26</td>
</tr>
<tr>
<td>s10es5.18</td>
<td>PD/V62I, D72V/PDAM</td>
<td>R23W</td>
<td>20</td>
<td>95</td>
<td>21</td>
</tr>
<tr>
<td>s10es6.12</td>
<td>PD/V62I, D72V/PDAM</td>
<td>K33R</td>
<td>52</td>
<td>200</td>
<td>25</td>
</tr>
<tr>
<td>s10es6.14</td>
<td>PD/V62I, D72V/PDAM</td>
<td>K33R</td>
<td>31</td>
<td>110</td>
<td>28</td>
</tr>
<tr>
<td>s10es6.15</td>
<td>PD/V62I, D72V/PDAM</td>
<td>A30V</td>
<td>31</td>
<td>89</td>
<td>38</td>
</tr>
<tr>
<td>s10es6.17</td>
<td>PD/V62I, D72V/PDAM</td>
<td>K40E</td>
<td>40</td>
<td>230</td>
<td>17</td>
</tr>
<tr>
<td>s1es2</td>
<td>PD/V62I, D72V/PDAM</td>
<td>Y66F</td>
<td>43</td>
<td>210</td>
<td>20</td>
</tr>
<tr>
<td>s1es2.10</td>
<td>PD/V62I, D72V/PDAM</td>
<td>I21L, G43S</td>
<td>11</td>
<td>67</td>
<td>16</td>
</tr>
<tr>
<td>s1es2.3</td>
<td>PD/V62I, D72V/PDAM</td>
<td>L24I, K40N</td>
<td>33</td>
<td>140</td>
<td>23</td>
</tr>
<tr>
<td>s1es3</td>
<td>PD/V62I, D72V/PDAM</td>
<td>L29I, K40M</td>
<td>29</td>
<td>160</td>
<td>18</td>
</tr>
<tr>
<td>s1es3.1</td>
<td>PD/V62I, D72V/PDAM</td>
<td>I78V</td>
<td>25</td>
<td>120</td>
<td>20</td>
</tr>
<tr>
<td>s1es3.2</td>
<td>PD/V62I, D72V/PDAM</td>
<td>K40M</td>
<td>32</td>
<td>85</td>
<td>38</td>
</tr>
<tr>
<td>s1es3.4</td>
<td>PD/V62I, D72V/PDAM</td>
<td>I14V, K40E</td>
<td>32</td>
<td>220</td>
<td>15</td>
</tr>
<tr>
<td>s1es3.7</td>
<td>PD/V62I, D72V/PDAM</td>
<td>L54F</td>
<td>23</td>
<td>210</td>
<td>11</td>
</tr>
<tr>
<td>s1es4</td>
<td>PD/V62I, D72V/PDAM</td>
<td>G43V</td>
<td>2.6</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>s1es5</td>
<td>PD/V62I, D72V/PDAM</td>
<td>E37D</td>
<td>27</td>
<td>78</td>
<td>34</td>
</tr>
</tbody>
</table>

Continued on the next page.
### Table A. IV (continued)

<table>
<thead>
<tr>
<th>6His Protein</th>
<th>Parental mutations</th>
<th>New mutations</th>
<th>$k_{\text{cat}}^a$ (s$^{-1}$)</th>
<th>$K_m^a$ (μM)</th>
<th>$k_{\text{cat}}/K_m^a \times 10^4$ (M$^{-1}$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>s1es5.1</td>
<td>PD/V62I, D72V/PDAM</td>
<td>A41V</td>
<td>35</td>
<td>103</td>
<td>35</td>
</tr>
<tr>
<td>s1es6.1</td>
<td>PD/V62I, D72V/PDAM</td>
<td>A41S</td>
<td>28</td>
<td>77</td>
<td>36</td>
</tr>
<tr>
<td>s4-repA2-es10</td>
<td>PD/V62I, D72V/PDAM</td>
<td>L24I, A36T, G84C</td>
<td>11</td>
<td>120</td>
<td>9.3</td>
</tr>
<tr>
<td>s4-repA2-es110</td>
<td>PD/V62I, D72V/PDAM</td>
<td>L24I, R34Q, K40E, L81M, G84S</td>
<td>2.6</td>
<td>300</td>
<td>0.9</td>
</tr>
<tr>
<td>s4-repA2-es113</td>
<td>PD/V62I, D72V/PDAM</td>
<td>L24I, G84C</td>
<td>8</td>
<td>33</td>
<td>24</td>
</tr>
<tr>
<td>s4-repA2-es116</td>
<td>PD/V62I, D72V/PDAM</td>
<td>E13D, L24I, K44E, G84S</td>
<td>18</td>
<td>150</td>
<td>12</td>
</tr>
<tr>
<td>s4-repA2-es16</td>
<td>PD/V62I, D72V/PDAM</td>
<td>L24I, E27G, A45V, G84C</td>
<td>8</td>
<td>48</td>
<td>17</td>
</tr>
<tr>
<td>s4-repA2-es22</td>
<td>PD/V62I, D72V/PDAM</td>
<td>L24I, E27G, A45V, G84C</td>
<td>2.6</td>
<td>300</td>
<td>0.9</td>
</tr>
<tr>
<td>s4-repA2-es3</td>
<td>PD/V62I, D72V/PDAM</td>
<td>E19K, L24I</td>
<td>10</td>
<td>67</td>
<td>15</td>
</tr>
<tr>
<td>s4-repA2-es30</td>
<td>PD/V62I, D72V/PDAM</td>
<td>E19K, L24I</td>
<td>17</td>
<td>74</td>
<td>23</td>
</tr>
<tr>
<td>s4-repA2-es31</td>
<td>PD/V62I, D72V/PDAM</td>
<td>E13D, L83I</td>
<td>45</td>
<td>250</td>
<td>18</td>
</tr>
<tr>
<td>s4-repA2-es33</td>
<td>PD/V62I, D72V/PDAM</td>
<td>L24I, A45V</td>
<td>23</td>
<td>150</td>
<td>15</td>
</tr>
<tr>
<td>s4-repA2-es4</td>
<td>PD/V62I, D72V/PDAM</td>
<td>L24I, E37D</td>
<td>25</td>
<td>150</td>
<td>17</td>
</tr>
</tbody>
</table>

$^a$See footnote $^a$ of Table A. I.
PD and PDAM are explained in the footnote of Table A.I.
Regulation, effectors, and evolvability of chorismate mutases from *Corynebacteriales*

Table A. V. Current status of data collection and refinement statistics for the evolved variants

<table>
<thead>
<tr>
<th></th>
<th>s10es4.15</th>
<th>s1es5.1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection and processing</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.95375</td>
<td>0.95372</td>
</tr>
<tr>
<td>Space group</td>
<td>P 64</td>
<td>P 64</td>
</tr>
<tr>
<td>Unit cell parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a, b, c$ (Å)</td>
<td>54.6, 54.6, 63.2</td>
<td>54.4, 54.4, 63.6</td>
</tr>
<tr>
<td>$\alpha, \beta, \gamma$ (°)</td>
<td>90, 90, 120</td>
<td>90, 90, 120</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>37.86 - 1.49 (1.51-1.49)</td>
<td>47.15 - 1.69 (1.72-1.69)</td>
</tr>
<tr>
<td>$R_{\text{merge}}$</td>
<td>0.05 (3.62)</td>
<td>0.05 (3.81)</td>
</tr>
<tr>
<td>$CC_{1/2}$</td>
<td>99.9 (33.2)</td>
<td>99.9 (98.5)</td>
</tr>
<tr>
<td>$I/\sigma$ (I)</td>
<td>15.0 (0.3)</td>
<td>20.4 (0.9)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>95.5 (82.9)</td>
<td>98.6 (93.7)</td>
</tr>
<tr>
<td>Multiplicity/redundancy</td>
<td>6.1 (1.9)</td>
<td>6.7 (2.8)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>16797 (2358)</td>
<td>12122 (1947)</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>47.27 – 1.49</td>
<td>47.15 – 1.69</td>
</tr>
<tr>
<td>$R_{\text{work}}/R_{\text{free}}$ (%)</td>
<td>18.4/20.5</td>
<td>16.8/21.0</td>
</tr>
<tr>
<td>No. of atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>629</td>
<td>622</td>
</tr>
<tr>
<td>Water</td>
<td>84</td>
<td>84</td>
</tr>
<tr>
<td>B-factors (Å²)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>49.0</td>
<td>47.6</td>
</tr>
<tr>
<td>Water</td>
<td>63.1</td>
<td>63.2</td>
</tr>
<tr>
<td>r.m.s.d. from ideal geometry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>2.05</td>
<td>1.84</td>
</tr>
<tr>
<td>Ramachandran (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favored</td>
<td>100</td>
<td>98.5</td>
</tr>
<tr>
<td>Allowed</td>
<td>0</td>
<td>1.5</td>
</tr>
<tr>
<td>Disallowed</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PDB ID</td>
<td>Not submitted</td>
<td>Not submitted</td>
</tr>
</tbody>
</table>

Values in parentheses refer to the highest-resolution shell.

\[ R_{\text{merge}} = \frac{\sum \sum |I_{hi} - \langle I_h \rangle|}{\sum \sum \langle I_h \rangle} \] where $\langle I_h \rangle$ is the average intensity over symmetry-related measurements.

\[ R = \frac{\sum |F_o - |F_c||}{\sum |F_o|} \] where $F_o$ and $F_c$ are the observed and calculated structure factors, respectively. $R_{\text{free}}$ is $R$ calculated for 5% randomly selected reflections, which were omitted from the refinement.
### Table A. VI Primers pairs for elongating individual truncated variant genes

<table>
<thead>
<tr>
<th>Variant name</th>
<th>Forward primer a</th>
<th>Reverse primer a</th>
</tr>
</thead>
<tbody>
<tr>
<td>s1-es4</td>
<td>561-dNdCs1-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s1-es2.10</td>
<td>561-dNdCs1-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s1-es6.1a</td>
<td>561-dNdCs1-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s1-es5.1</td>
<td>561-dNdCs1-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s1-es3.2</td>
<td>561-dNdCs1-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s1-es3</td>
<td>561-dNdCs1-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s1-es2.3</td>
<td>561-dNdCs1-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s1-es3.4</td>
<td>561-dNdCs1-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s1-es5</td>
<td>561-dNdCs1-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s1-es3.1</td>
<td>561-dNdCs1-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s1-es3.7</td>
<td>561-dNdCs1-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s1-es2</td>
<td>561-dNdCs1-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s10-es 4</td>
<td>60-T7Pro</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s10-es 2.12</td>
<td>60-T7Pro</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s10-es 2.14</td>
<td>60-T7Pro</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s10-es 5.18</td>
<td>60-T7Pro</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s10-es 4.15</td>
<td>60-T7Pro</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s10-es 3.12</td>
<td>60-T7Pro</td>
<td>562-dNdCs4-Rv</td>
</tr>
<tr>
<td>s10-es 3.13</td>
<td>60-T7Pro</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s10-es 5.12</td>
<td>60-T7Pro</td>
<td>563-dNdCs4-Rv</td>
</tr>
<tr>
<td>s10-es 6.12</td>
<td>60-T7Pro</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s10-es 6.13</td>
<td>60-T7Pro</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s10-es 6.14</td>
<td>60-T7Pro</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s10-es 6.15</td>
<td>60-T7Pro</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s10-es 6.17</td>
<td>60-T7Pro</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s4-repA2-s 3</td>
<td>554-dNdCs4-Fw</td>
<td>557-dNdCs4-Rv</td>
</tr>
<tr>
<td>s4-repA2-s 4</td>
<td>554-dNdCs4-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s4-repA2-s 10</td>
<td>554-dNdCs4-Fw</td>
<td>557-dNdCs4-Rv</td>
</tr>
<tr>
<td>s4-repA2-s 16</td>
<td>554-dNdCs4-Fw</td>
<td>557-dNdCs4-Rv</td>
</tr>
<tr>
<td>s4-repA2-s 22</td>
<td>555-dNdCs4-Fw</td>
<td>557-dNdCs4-Rv</td>
</tr>
<tr>
<td>s4-repA2-s 24</td>
<td>554-dNdCs4-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s4-repA2-s 30</td>
<td>554-dNdCs4-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s4-repA2-s 31</td>
<td>555-dNdCs4-Fw</td>
<td>559-dNdCs4-Rv</td>
</tr>
<tr>
<td>s4-repA2-s 33</td>
<td>554-dNdCs4-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s4-repA2-s 60</td>
<td>555-dNdCs4-Fw</td>
<td>557-dNdCs4-Rv</td>
</tr>
<tr>
<td>s4-repA2-s 62</td>
<td>555-dNdCs4-Fw</td>
<td>558-dNdCs4-Rv</td>
</tr>
<tr>
<td>s4-repA2-s 101</td>
<td>554-dNdCs4-Fw</td>
<td>556-dNdCs4-Rv</td>
</tr>
<tr>
<td>s4-repA2-s 110</td>
<td>554-dNdCs4-Fw</td>
<td>560-dNdCs4-Rv</td>
</tr>
<tr>
<td>s4-repA2-s 113</td>
<td>554-dNdCs4-Fw</td>
<td>557-dNdCs4-Rv</td>
</tr>
<tr>
<td>s4-repA2-s 116</td>
<td>555-dNdCs4-Fw</td>
<td>558-dNdCs4-Rv</td>
</tr>
</tbody>
</table>

a Sequences are presented in Table A.VII.
Table A. VII. Primer sequences used for elongating individual truncated variant genes

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>554-dNdCs4-Fw</td>
<td>5'-TATGCTCGAGTCCCAACCTGTCCCCGAGATCGACACGCTGC</td>
</tr>
<tr>
<td>555-dNdCs4-Fw</td>
<td>5'-TATGCTCGAGTCCCAACCTGTCCCCGATATCGACACGCTG</td>
</tr>
<tr>
<td>556-dNdCs4-Rv</td>
<td>5'-CACTAGTTATTACATAGCATTCCGACCACGACCAA</td>
</tr>
<tr>
<td></td>
<td>GACGCAAAAGCA</td>
</tr>
<tr>
<td>557-dNdCs4-Rv</td>
<td>5'-CACTAGTTATTACATAGCATCCGACCACGACCAA</td>
</tr>
<tr>
<td></td>
<td>AGACGCAAAAGCA</td>
</tr>
<tr>
<td>558-dNdCs4-Rv</td>
<td>5'-CACTAGTTATTACATAGCATCCGACCACGACTAAGA</td>
</tr>
<tr>
<td></td>
<td>CGCAAAAGCA</td>
</tr>
<tr>
<td>559-dNdCs4-Rv</td>
<td>5'-CACTAGTTATTACATAGCATCCGACCACGACCAAATA</td>
</tr>
<tr>
<td></td>
<td>CGCAAAAGCA</td>
</tr>
<tr>
<td>560-dNdCs4-Rv</td>
<td>5'-CACTAGTTATTACATAGCATCCGACCACGACTAAG</td>
</tr>
<tr>
<td></td>
<td>ACGCATAAGCA</td>
</tr>
<tr>
<td>561-dNdCs1-Fw</td>
<td>5'-TATGCTCGAGTCCCAACCTGTCCCCGAGATCGACACG</td>
</tr>
<tr>
<td></td>
<td>CTGCCGAAGAGATCGACCGG</td>
</tr>
<tr>
<td>562-dNR85s10-Rv</td>
<td>5'-CACTAGTTATTACATAGCATCCGACCACGATCAAGACG</td>
</tr>
<tr>
<td></td>
<td>CAAAGCA</td>
</tr>
<tr>
<td>563-dNR85s10-Rv</td>
<td>5'-CACTAGTTATTACATAGCATCCGACCACGACCAAGACG</td>
</tr>
<tr>
<td></td>
<td>CAAAGCAGGTGG</td>
</tr>
<tr>
<td>60-T7Pro</td>
<td>5'-TAATACGACTCATATAGGG</td>
</tr>
</tbody>
</table>
**LIST OF ABBREVIATIONS**

Å  Ångstrom (10⁻¹⁰ meters)
Amp  ampicillin (value in superscript indicates the concentration of the sodium salt in μg/mL)
BCG  Bacille Calmette-Guérin
(2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)-1,3-propanediol)
bp  base pairs
BSA  bovine serum albumin
BsCM  monofunctional chorismate mutase from *Bacillus subtilis*
BTP  1,3-bis[tris(hydroxymethyl)methylamino]propane
CD  circular dichroism
CM  chorismate mutase
CP  cyclic peptide
CgCM  chorismate mutase from *C. glutamicum*
CgDS  DAHP synthase from *C. glutamicum*
Da  dalton
DAHP  3-deoxy-D-arabino-heptulosonate-7-phosphate
DMF  N,N-dimethylformamide
DMSO  dimethyl sulfoxide
DNA  deoxyribonucleic acid
DS  DAHP synthase
E4P  D-erythro-4-phosphate
EcCM  CM domain of the bifunctional CM-prephenate dehydratase from *E. coli*
EDTA  ethylenediaminetetraacetic acid
epPCR  error-prone PCR
ESI  electrospray ionization
ESRF  European Synchrotron Radiation Facility
EtBr  EtBr
FACS  fluorescence-activated cell sorting
FMP Berlin  Leibniz Institute for Molecular Pharmacology
FPLC  fast protein liquid chromatography
FY  phenylalanine and tyrosine
GC  gas chromatography
h  hours
H1'  helix 1 of the second protomer
HIV  human immunodeficiency virus
HPLC  high pressure liquid chromatography
HTS  high-throughput screening
I_C  carboxyl terminal intein fragment
IC₅₀  half maximal inhibitory concentration of a compound
I_N  amino terminal intein fragment
in vitro  in the test tube (strictly cell-free)
in vivo  within living cells; also applied to bacteria
IPTG  isopropyl β-D-1-thiogalactopyranoside
IVC  in vitro compartmentalization
Kan kanamycin (value in superscript indicates the concentration of the sulfate salt in µg/mL)

\( k_{\text{cat}} \) turnover number

\( k_{\text{cat}}/K_m \) catalytic efficiency of the enzyme

\( K_m \) Michaelis constant

\( K_d (K_{d,\text{app}}) \) dissociation constant (apparent dissociation constant)

LB Luria-Bertani broth

LC-MS high pressure liquid chromatography couplet to mass spectrometry

LMW LMW low molecular weight marker

M molar concentration

MDR-TB multidrug-resistant tuberculosis

MjCM chorismate mutase from *Methanococcus jannaschii*

MS mass spectrometry

MtCM *Mycobacterium tuberculosis* chorismate mutase (Rv0948c)

*MtCM* secreted CM from *Mycobacterium tuberculosis*

MtDS *Mycobacterium tuberculosis* DAHP synthase (Rv2178c)

NAD\(^+\)/NADH nicotinamide adenine dinucleotide

NMR nuclear magnetic resonance spectroscopy

OD\(_{600}\) optical density at 600 nm

ori origin of replication

PAGE polyacrylamide gel electrophoresis

PCR polymerase chain reaction

PDB Protein Data Bank

PDH prephenate dehydrogenase

PDT prephenate dehydratase

PEP phosphoenolpyruvate

pFPhe DL-para-fluoro-phenylalanine

PPI protein-protein interaction

\( P_{\text{sal}} \) salicylate promoter

PsF Pseudostellarin F

\( P_{\text{prpB}} \) priopionate promoter

\( P_{\text{T7}} \) T7 promoter

\( P_{\text{tet}} \) tetracycline promoter

RBS ribosome binding site

RepA amino-acid protein fragment (MNQSFISDILYADIES) of the replication protein RepA

RNA RNA ribonucleic acid

rpm resolutions per minute

Sal salicylate (value in superscript indicates the concentration in mM)

SAXS small angle X-ray scattering

ScCM CM from *Saccharomyces cerevisiae*

SDS sodium dodecyl sulfate

SICLOPPS split-intein circular ligation of peptides and proteins system

SsrA amino-acid protein sequence (AANDENYALAA) that directs modified proteins to a protein degradation system

TB tuberculosis

TCA trichloroacetic acid

TCEP tris(2-carboxyethyl)phosphine hydrochloride

TDR-TB totally drug-resistant tuberculosis

Tet tetracycline (value in superscript indicates the concentration in ng/mL)
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIM</td>
<td>triosephosphate isomerase</td>
</tr>
<tr>
<td>$T_m$</td>
<td>melting temperature</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)-aminomethane</td>
</tr>
<tr>
<td>TSA</td>
<td>Bartlett’s <em>endo</em>-oxa-bicyclic transition state analog[^33]</td>
</tr>
<tr>
<td>UV</td>
<td>ultra violet</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>XDR-TB</td>
<td>extensively drug-resistant tuberculosis</td>
</tr>
</tbody>
</table>

[^33]: The standard one and three letter codes were used for the 20 proteinogenic amino acids.
ACKNOWLEDGEMENTS

I would like to thank my PhD supervisor Prof. Peter Kast for the opportunity to work on the research projects presented in this thesis and for constant support with experimental issues. I also thank Prof. Stefanie D. Krämer for her interest in my research and for agreeing to co-examine this thesis.

I thank Prof. Donald Hilvert for inviting me for a job interview and the opportunity to work in a close relation to his outstanding research group.

I would like to acknowledge Dr. Jens Peter von Kries, Dr. Martin Neuenschwander, Dr. Edgar Specker, and Carola Seyfahrt from FMP Screening Unit in Berlin for the opportunity to perform the high throughput screen. My special gratitude belongs to Dr. Martin Neuenschwander for the data evaluation and support during the experiment. Also, I would also like to thank Dr. Steffi Munack from Prof. Ute Krengel’s group, for a collaborative work with me at the FMP Berlin.

I am grateful to Prof. Ute Krengel who gave me the opportunity to visit her laboratory at the University of Oslo where I learned about X-ray diffraction data processing and structure refinement. I am thankful to Helen Vikdal Thorbjørnsrud who crystallized and collected the diffraction data of my favorite proteins and measured CgCM-CgDS kinetics with me.

A special thank you belongs to Dr. Marianne Gamper for proofreading parts of this thesis as well as Robert Fahrig and Dr. Felix Kurth for assembling the Zusammenfassung.

I thank my talented and hard-working semester students Elias A. Halabi, Florian Frei, Reto Erni, and Marina Rabinovich for their dedication. I hope I was able to provide you with interesting projects and skills that would help in your future careers.

Finally, my gratitude goes to teaching assistants and students of the one-month laboratory block course “Biological Chemistry B: New Enzymes from Directed Evolution Experiments” in 2012 and 2013, who contributed to Chapter 5.
8 REFERENCES


