

Diss. ETH No. 20407

**Structure, Function, and Directed Evolution of Shikimate
Pathway Enzymes from *Mycobacterium tuberculosis***

A dissertation submitted to the
ETH Zürich

For the degree of
Doctor of Sciences

Presented by
Kathrin Roderer

Dipl. Natw. ETH
born May 16, 1983
citizen of Trogen AR

Accepted on the recommendation of
Prof. Peter Kast, examiner
Prof. Ute Krengel, co-examiner
Prof. Donald Hilvert, co-examiner

Zürich, 2012

Summary

Mycobacterium tuberculosis is a deadly organism causing tuberculosis, which is responsible for millions of deaths world wide every year. Its intrinsically high antibiotic resistance, its slow growth, and the fact that the disease is still mainly a 3rd world problem has led to inadequate tackling of the potential of cure so far.

Generally, enzymes of anabolic pathways in pathogens that are absent from their host are considered “safe” drug targets due to reduced cross-reactivity in the host. The shikimate pathway, responsible for production of the aromatic amino acids phenylalanine (Phe), tyrosine (Tyr), and tryptophan (Trp), as well as many aromatic compounds of the secondary metabolism in bacteria, fungi, and plants, is absent in mammals. The enzymatic series leading to these aromatic compounds starts out with the condensation of phosphoenolpyruvate and D-erythrose-4-phosphate to form 3-deoxy-D-*arabino*-heptulosonate-7-phosphate (DAHP) by DAHP synthase (DS). Six consecutive steps convert DAHP to the chorismate molecule, the branch point metabolite of Trp and Phe/Tyr synthesis. Chorismate mutase (CM) catalyzes the first committed step in Tyr and Phe biosynthesis, the pericyclic conversion of chorismate to prephenate. The identification of two new subclasses of CMs in *M. tuberculosis* has led to intense studies on their structure and function. A secreted CM (Rv1885c, *MtCM, AroQ_γ subclass), whose function is so far unknown but was implicated in pathogenesis, does not contribute to the intracellular pool of aromatic amino acids. The housekeeping CM (Rv0948c, MtCM, AroQ_δ subclass) has low enzymatic activity on its own (1% of typical CMs) but can be activated over 100-fold through formation of a non-covalent protein complex with the DS of *M. tuberculosis* (MtDS). The X-ray crystal structure has been determined and showed a hetero-octameric complex with two MtCM dimers associated with an MtDS tetrameric core. Upon complex formation, the CM activity becomes sensitive to inhibition by the end products Phe and Tyr. We could show that the allosteric inhibitors bind to MtDS and thereby lead to 25-fold reduced CM activity of the complex. The effect of Tyr and Phe on the DS activity is only minor, whereas it responds considerably to a mixture of Phe and Trp. The most complete inhibition of MtDS, however, is observed in the presence of all three aromatic

amino acids with half maximal inhibition at a concentration of 17 μM , which is in good agreement with the intracellular concentration of these amino acids of 12 μM (Trp), 18 μM (Phe), and 29 μM (Tyr) in exponentially growing bacteria (*Escherichia coli*). A Tyr binding site could be identified in MtDS from mutational analysis and a concomitant loss of sensitivity to Tyr of the DS activity and the CM activity in the complex. Size-exclusion chromatography experiments with the complex in the presence of chorismate together with Phe and Tyr suggest that the CM activity in the complex is regulated by reversible dissociation of the hetero-octameric complex.

Directed evolution experiments targeting the seven C-terminal residues (positions 84 to 90) of MtCM and selection in a CM-deficient *Escherichia coli* strain in the presence of MtDS yielded a sequence pattern compatible with productive interactions with MtDS. The consensus pattern mirrors a multiple sequence alignment of naturally occurring CMs from the subclass AroQ₈. In particular, an Arg-Gly dyad at positions 85 and 86, which was found to be conserved in AroQ₈ variants, did not show any tolerance to mutation in our selected variants. Furthermore, substitutions for Leu88 and Gly89 strongly favored small and hydrophobic residues. To identify positions that are only involved in the catalytic core of MtCM, the libraries were also subjected to very low stringency selection on minimal medium in the absence of MtDS. We found that residues 84 to 86 are crucial for retention of a functional active site. Our results suggest that the C-terminal MtCM residues Arg87, Leu88, and Gly89 are critically involved in productive complex formation with MtDS.

In another directed evolution project, we explored the difficulty of evolving the naturally inefficient MtCM into a proficient, MtDS-independent enzyme by performing random mutagenesis and selection in a CM-deficient *E. coli* strain. After just two rounds of directed evolution, error-prone PCR of the entire gene yielded a variant with a $k_{\text{cat}}/K_{\text{m}}$ that was 28-fold better than the parent, whereas focused cassette mutagenesis in regions relatively distant from the active site led to a clone with a catalytic efficiency ($k_{\text{cat}}/K_{\text{m}}$) 82-fold better than the parent. The best variant that was discovered has the Thr52Pro and Val55Asp mutations in the loop between helices one and two, and Pro-Asp-Ala-Met in place of the C-terminal Arg-Leu-Gly-His sequence; it is only a factor of 3 less active than the complexed wild-type MtCM.

The structurally quite distinct CMs from the new AroQ_γ and AroQ_δ subclasses discovered in mycobacteria may be inhibited by compounds that were not known as CM inhibitors to date. We therefore set out to identify new CM inhibitors from a large compound library in a collaboration with the FMP Berlin. Specific inhibitors could be useful for further characterization of the mycobacterial CMs *in vitro* and for *in vivo* studies in the model organism *Mycobacterium marinum*. To facilitate compound screening in a high-throughput format, a fluorescence-based CM assay was developed that couples the CM reaction to NADH production. Application of this assay to a 32,032 compound library to identify inhibitors for both the secreted and the intracellular CMs from *M. tuberculosis* revealed highly potent compounds with preliminary IC₅₀ values about 3-fold better than the best known CM inhibitor available to date. Further experiments are required to confirm the effect on enzyme function and explore their potential for studies *in vivo*.

Zusammenfassung

Mycobacterium tuberculosis ist ein tödliches Bakterium, welches Tuberkulose (TB) verursacht und jedes Jahr weltweit für Millionen von Todesfällen verantwortlich ist. Seine intrinsisch hohe Antibiotikaresistenz, sein langsames Wachstum und die Tatsache dass TB immer noch vor allem die dritte Welt betrifft, hat dazu geführt, dass das Problem bisher nicht adäquat angegangen worden ist.

Im Allgemeinen werden Enzyme von anabolischen Synthesewegen in Pathogenen, die nicht im Wirt vorkommen, als „sichere“ Ziele für neue Medikamente betrachtet, da meist weniger Nebenwirkungen zu erwarten sind. Der Shikimat-Weg, welcher für die Synthese der aromatischen Aminosäuren Phenylalanin (Phe), Tyrosin (Tyr) und Tryptophan (Trp), sowie von vielen aromatischen Substanzen im sekundären Metabolismus von Bakterien, Pilzen und Pflanzen verantwortlich ist, kommt in Säugern nicht vor. Die Enzymkaskade, die zur Synthese von diesen aromatischen Produkten führt, beginnt mit der Kondensation von D-Erythrose-4-Phosphat und Phosphoenolpyruvat zu 3-Deoxy-D-Arabino-Heptulosonat-7-Phosphat (DAHP), die vom Enzym DAHP Synthase (DS) katalysiert wird. Sechs nachfolgende Schritte konvertieren DAHP zum zentralen Chorismat Molekül, das über zwei Äste zu Trp oder Phe und Tyr umgesetzt wird. Chorismat-Mutase (CM) katalysiert den ersten Schritt in Richtung Tyr und Phe, nämlich die perizyklische Umwandlung von Chorismat zu Prephenat. Die Identifizierung von zwei CMs in *M. tuberculosis*, welche zwei neuen Subklassen dieser Enzyme angehören, hat zu intensiven Studien über die Struktur und die Funktion der zwei Enzyme geführt. Die eine CM (Rv1885c, *MtCM, AroQ_γ Subklasse), deren Funktion für das Bakterium noch ungeklärt ist, aber sehr wahrscheinlich eine Rolle in der Pathogenese übernimmt, trägt nicht zum intrazellulären Pool von aromatischen Aminosäuren bei. Die CM die dafür verantwortlich ist (Rv0948c, MtCM, AroQ_δ Subklasse), ist zwar intrazellulär, zeigt aber nur etwa 1% der Aktivität von typischen CMs. Sie kann aber über die Ausbildung eines nicht-kovalenten Komplexes mit der DAHP Synthase von *M. tuberculosis* (MtDS) über 100-fach aktiviert werden. Die Kristallstruktur dieses Proteinkomplexes ist bestimmt worden und zeigt eine hetero-oktamere Zusammenlagerung

von zwei MtCM-Dimeren auf entgegengesetzten Seiten eines MtDS-Tetramers. Die Komplexbildung hat zur Folge, dass die CM-Aktivität im Komplex über die Endprodukte des Shikimat-Wegs, Phe und Tyr, reguliert werden kann. Wir konnten zeigen, dass diese allosterischen Inhibitoren an MtDS binden und dadurch die CM-Aktivität des Komplexes um einen Faktor 25 verringern. Der Effekt von Phe und Tyr auf die DS-Aktivität ist nur gering, dafür reagiert das Enzym stark auf die Kombination von Phe und Trp. Die vollständigste Inhibition von MtDS kann jedoch bei gleichzeitiger Anwesenheit aller drei aromatischen Aminosäuren beobachtet werden, mit halbmaximaler Inhibition bei einer Konzentration von 17 μM , was ungefähr den intrazellulären Konzentrationen dieser Aminosäuren in exponentiell-wachsenden Bakterien (*Escherichia coli*) von 12 μM (Trp), 18 μM (Phe) und 29 μM (Tyr) entspricht. Auf MtDS konnte eine Tyr-Bindungsstelle durch eine Mutation, und dem einhergehenden Verlust der Sensitivität auf Tyr, sowohl der DS-Aktivität als auch der CM-Aktivität im Komplex, identifiziert werden. Gelfiltrationsexperimente mit dem Komplex in Anwesenheit von Chorismat, Phe und Tyr weisen darauf hin, dass die Regulation auf reversible Dissoziation des hetero-oktameren Komplexes zurück zu führen ist.

Gerichtete Evolutionsexperimente, welche die sieben C-terminalen Aminosäuren auf MtCM (an den Positionen 84 bis 90) untersuchten, und Selektion bei gleichzeitiger Anwesenheit von MtDS in einem CM-defizienten *E. coli*-Stamm beinhalteten, lieferten ein Sequenzmuster, welches mit einer produktiven MtDS-Interaktion kompatibel ist. Das Muster der am häufigsten gefundenen Aminosäuren widerspiegelt einen Sequenzvergleich von natürlich vorkommenden CMs der AroQ₈-Subklasse. Im Besonderen zeigte eine Arg-Gly Dyade an den Positionen 85 und 86 keine Toleranz für Mutationen in unseren selektierten Varianten, und auch der Sequenzvergleich der AroQ₈-Proteine zeigt vollständige Konservierung dieser zwei Positionen. An den Positionen Leu88 und Gly89 waren neben den Wildtyp Aminosäuren, vor allem kleine, hydrophobe Reste erlaubt. Um ausschliessen zu können, dass Reste, welche nur in der katalytischen Maschinerie von MtCM wichtig sind, als Interaktionsdeterminanten definiert werden, wurden die MtCM-Genbibliotheken auch auf wenig selektivem Minimalmedium in Abwesenheit von MtDS getestet. Dabei haben wir gefunden, dass die Aminosäuren an den Positionen 84 bis 86

wichtig sind um eine funktionelle aktive Tasche zu erhalten. Unsere Resultate zeigen, dass vor allem die Aminosäuren Arg87, Leu88 und Gly89 kritisch sind für eine produktive Ausbildung des Komplexes mit MtDS.

In einem weiteren Experiment, welches die gerichtete Evolution beinhaltet, haben wir versucht heraus zu finden, wie schwierig es sein würde, aus der natürlichen, ineffizienten MtCM aus dem Komplex eine hoch aktive CM zu evolvieren. Dazu wurden Runden von Mutagenese und Selektion in einem CM-defizienten *E. coli*-Stamm durchgeführt. Nach nur gerade zwei Evolutionszyklen hat eine Methode, welche zufällige Mutationen über das ganze Gen verteilt einführt, zu einer MtCM-Variante mit 28-fach höherem k_{cat}/K_m -Wert geführt. Eine zweite Strategie, die bestimmte Regionen mittels Kassettenmutagenese untersuchte, hat nach zwei Runden MtCM-Varianten mit 82-fach höherer katalytischer Effizienz (k_{cat}/K_m) geliefert. Die beste Variante trägt zwei Mutationen im Loop zwischen der ersten und der zweiten Helix (Thr52Pro und Val55Asp), sowie Pro-Asp-Ala-Met an Stelle von Arg-Leu-Gly-His ganz am C-Terminus von MtCM. Die Aktivität dieser MtCM-Variante liegt nur einen Faktor drei unter dem komplexierten Wildtyp Enzym.

Die strukturell sehr unterschiedlichen CMs der neuen AroQ $_{\gamma}$ - und AroQ $_{\delta}$ -Subklassen, die in *M. tuberculosis* gefunden worden sind, könnten von Substanzen inhibiert werden, welche wir bis jetzt noch nicht als CM-Inhibitoren kennen. Deshalb haben wir uns in einer Kollaboration mit dem FMP in Berlin in einer grossen Substanzbibliothek auf die Suche nach neuen CM-Inhibitoren gemacht. Spezifische Inhibitoren könnten nützlich sein, um die mykobakteriellen CMs *in vitro* besser charakterisieren zu können, und könnten auch für *in vivo*-Studien am Modellorganismus *Mycobacterium marinum* dienen. Um ein Screening im Hochdurchsatz zu ermöglichen, wurde ein Fluoreszenz-basierter Assay entwickelt, welcher die CM-Reaktion an NADH-Produktion knüpft. Die Anwendung dieses Assays für die Suche nach CM Inhibitoren unter 32'032 Substanzen hat in neuen potenziellen Kandidaten resultiert, welche in ersten Tests bis zu dreifach bessere Affinitäten zu den CMs zeigten als der bisher beste verfügbare Inhibitor. Weitere Experimente sind nötig um den Effekt auf die Enzymfunktion zu verifizieren und um das Potential der Inhibitoren für *in vivo*-Studien zu erforschen.