

Partners in crime

Structure-function studies on chorismate mutases and their associated enzymes



Tamjidmaa Khatanbaatar

Doctoral Thesis

Department of Chemistry

Faculty of Mathematics and Natural Sciences

University of Oslo

2023

© **Tamjidmaa Khatanbaatar, 2023**

*Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 2608*

ISSN 1501-7710

All rights reserved. No part of this publication may be
reproduced or transmitted, in any form or by any means, without permission.

Cover: UiO.

Print production: Graphic Center, University of Oslo.

Acknowledgements

This work was carried out from 2019 Oct to 2023 Jan at the laboratory of Prof. Ute Krenzel, Department of Chemistry of the University of Oslo. This project was funded by a grant from the Swiss National Science Foundation (grant 310030M_182648).

Foremost, I would like to acknowledge my main supervisor, Prof. Ute Krenzel who gave me the opportunity to pursue a doctoral degree in her research group and guided me through the whole project. I am thankful for your share of knowledge and insightful ideas that kept me constantly engaged with my research and made me eager to learn more. I would like to thank my co-supervisors, Dr. Gabriele Cordara for his continuous support and brilliant ideas, which helped me to improve my scientific thinking, and Prof. Michele Cascella for scientific discussions and valuable advice during my research.

Furthermore, I would like to express my gratitude to my co-authors and collaborators. Thank you, Peter Kast group at the ETH, Zurich, including Christian Stocker, Luca Bressan, Kathrin Würth-Roderer, and Peter Kast for your fundamental contribution to this work and all meaningful discussions. Special thanks go to Luca Bressan for helping me with experiments during my visit. I would like to thank our collaborators Dr. Tobias Krojer and Dr. Derek Logan for your collaboration and support in crystallographic fragment screening.

I am grateful that I met and worked with many good people during my time as a PhD student, including members of the Protein Crystallography group: Gabriele Cordara, Hedda Johannesen, Henrik Vinther Sørensen, Dipankar Manna, Giacomo Pesci, Kira Leidl, Clara Helena Klee, Natalia Mojica Cortés, Abeline Tislevoll, Helene Sørsum Kristiansen, Mateu Montserrat Canals, Flore Kersten, Eirik Falao, Mari Stramrud, Ayla Steffensen Coder and students who contributed significantly to the chorismate mutase study: Claire Auvray and Henrik Torjusen. I wish everyone a good health and success in your own meaning. Among them, I would like to thank Clara and Gabriele for being such nice office mates and friends. You guys made the office warm and bright. I thank Natalia for her kindness and all the nice dinners and walks we shared together, and Henrik for his help with SAXS analysis and also with the mastermind. Thanks, Mari, for revising my summary in norwegian.

Furthermore, special honors to my friends: Bilegee, Uyanga, Kherlen, and Jakk for a long-lasting friendship and continuous support. Let's become cool "ajummadeul" together. All my friends in Budapest, thank you for still staying close and welcoming me every time I visit. To Flo and Björn, friends I met during my years in Norway, thank you for the company. I always enjoy spending time together.

I take this opportunity to express my love and thanks to my family, my parents, brother, and sisters. I feel grateful to everyone for giving me a great childhood and loving me so much. Best wishes to my nieces and nephews, who brought great joy to me. (Аав ээж, ах эгч, дүү нартаа үргэлж хайрлаж, тусалж байдагт баярлалаа). My biggest thanks go to my halzan zaluu for encouraging me in everything and making my life more meaningful. You have been a very bumbug hair and will always be, my sunshine.

Tamjidmaa Khatanbaatar

Summary

All metabolic processes require biological catalysts, also called enzymes. Enzymes involved in a specific metabolic pathway are colocalized within the organism, and often associate with each other for more efficient performance. Furthermore, metabolic pathways are often allosterically regulated either through feedback inhibition by the end-products or via enzyme-enzyme interactions. Detailed knowledge of allosteric regulation of metabolic enzymes and their ability to form complexes among each other is crucial for the design of drugs/inhibitors or their exploitation in the synthesis of fine chemicals.

This PhD thesis focuses on chorismate mutase (CM), an enzyme involved in the aromatic amino acids biosynthesis pathway in bacteria, fungi and plants. The pathway is a very promising source of novel drug targets, as it is unique to these organisms, many of them involved in unwanted processes, such as human infection, food spoilage and plant parasitosis. CM is a branch-point enzyme, directing the pathway towards the synthesis of L-Phe or L-Tyr. It catalyzes the conversion of chorismate to prephenate through a pericyclic reaction, one of the few natural enzymes doing so. We tried to lay the foundation for the future discovery of new drug targets by gaining precise understanding on its allosteric regulation and its mode of interaction with associated enzymes.

Nine species of non-pathogenic γ - and β -proteobacteria encode for novel exported (marked with an asterisk, ‘*’, before their name) bifunctional fusion enzymes, carrying a CM domain fused to a cyclohexadienyl dehydratase (CDT) domain in two different topologies: *CMCDT and *CDTCM. CM and CDT catalyze sequential reactions in the terminal pathway leading to the synthesis of L-Phe. Postulating that they could provide a paradigm for communication between *fused* sequential enzymes, we aimed to understand the evolutionary reason for the genetic fusion of CM and CDT. In **Manuscript I** we analyzed periplasmic *CM/CDT fusion enzymes representing both topologies. We probed the presence of a catalytic advantage granted by the fusion of the two domains by using enzyme kinetics. The relative orientation of the CM and CDT active sites and existence of substrate channeling was investigated by X-ray crystallography. Overall, no catalytic benefit or substrate channeling was detected. SAXS data collected on some orthologs and not included in the manuscript suggest that the bifunctional enzyme do not form multimers. Based on the extremely low affinity of the CM domain for its substrate ($K_m < 7 \mu\text{M}$), chorismate, and the presence of genes encoding

transporters and substrate binding proteins in the same operon, we hypothesize that these bifunctional fusion enzymes play a role in signaling, rather than the biosynthesis of aromatic amino acids. We additionally (**Manuscript II**) investigated the quaternary structure and association of exported monomolecular CM and CDT enzymes found in the periplasm of *Pseudomonas aeruginosa* (*PaeCM and *PaeCDT), which have a fused CMPDT cytoplasmic counterpart (PaeCMPDT). *P. aeruginosa* is an opportunistic pathogen targeting immunocompromised patients (e.g. cancer or HIV patients). The structure of *PaeCM and *PaeCDT was studied by X-ray crystallography and small-angle X-ray scattering (SAXS), while their association was probed by size-exclusion chromatography (SEC). The collected data suggest that *PaeCDT and *PaeCM do not form a stable complex; however, we gathered insights into the dynamic behavior of *PaeCM and *PaeCDT, which could be exploited in later studies.

Several CM classes with completely different folds have been identified. Among them AroQ₈-subclass CMs are naturally mediocre enzymes, which need inter-enzyme allosteric activation by forming a complex with DAHP synthase (DS), the first enzyme of the aromatic amino acid biosynthesis pathway. We took as the model system CM from *Mycobacterium tuberculosis* (MtCM), an AroQ₈-subclass CM. As *M. tuberculosis* is the causative agent of the tuberculosis disease, understanding in deep detail the MtCM activation can lead to the design of effective drugs. In **Manuscript III** we aimed to elucidate the structure-functional determinants underlying the allosteric activation of MtCM. We carried out structural studies and enzyme kinetics analysis targeting the wild-type MtCM (MtCM^{WT}), a laboratory-evolved super-active variant (MtCM^V) and two variants boosting the catalytic activity by a factor of six (MtCM^{T52P}) and 12 (MtCM^{V55D}). We investigated the structural importance of mutations introduced by directed evolution by comparing the conformation sampling of three MtCMs: apo wild-type (mediocre), and transition state analog (TSA)-bound wild-type (MtCM^{LC}) and highly active (lab-evolved) using molecular dynamics simulation. Our results revealed that the T52P and V55D substitutions help to rigidify an active-site loop that carries catalytically crucial residues, getting them pre-organized for catalysis. Finally, we aimed to discover potent small-molecule compounds providing leads to inhibitors and allosteric modulators for MtCM by fragment-based screening. The screening yielded a selection of possible ligands, which will be further investigated in future studies.

Sammendrag

Alle metabolske prosesser krever biologiske katalysatorer, også kalt enzymer. Enzymer som er involvert i en spesifikk metabolsk vei, er samlokalisert i organismen og samarbeider ofte med hverandre for å oppnå en mer effektiv ytelse. Videre er metabolske veier ofte allosterisk regulert enten gjennom feedback-inhibering av sluttproduktene eller via enzym-enzym-interaksjoner. Detaljert kunnskap om allosterisk regulering av metabolske enzymer og deres evne til å danne komplekser med hverandre er avgjørende for design av legemidler/inhibitorer eller utnyttelse av dem i syntesen av finkjemikalier.

Denne doktorgradsavhandlingen fokuserer på korismatmutas (CM), et enzym som er involvert i biosyntesen av aromatiske aminosyrer i bakterier, sopp og planter. Veien er en svært lovende kilde til nye legemiddelmål, siden den er unik for disse organismene, og mange av dem er involvert i uønskede prosesser, for eksempel infeksjon hos mennesker, matforringelse og planteparasitose. CM er et forgreningsenzym som styrer veien mot syntesen av L-Phe eller L-Tyr. Det katalyserer omdannelsen av korismat til prefenat gjennom en perisyklisk reaksjon, et av de få naturlige enzymene som gjør dette. Vi forsøkte å legge grunnlaget for fremtidig oppdagelse av nye legemiddelmål ved å få en presis forståelse av den allosteriske reguleringen og interaksjonen med tilknyttede enzymer. Ni arter av ikke-patogene γ - og β -proteobakterier koder for nye eksporterte (merket med en stjerne, '*', foran navnet) bifunksjonelle fusjonsenzymer, som bærer et CM-domene fusjonert til et cykloheksadienyldehydratase (CDT)-domene i to forskjellige topologier: *CMCDT og *CDTCM. CM og CDT katalyserer sekvensielle reaksjoner i terminalveien som fører til syntese av L-Phe. Vi postulerte at de kunne gi et paradigme for kommunikasjon mellom fusjonerte sekvensielle enzymer, og hadde som mål å forstå den evolusjonære årsaken til den genetiske fusjonen av CM og CDT. **I Manuskript I** analyserte vi periplasmatiske *CM/CDT-fusjonsenzymer som representerer begge topologiene. Vi undersøkte tilstedeværelsen av en katalytisk fordel gitt av fusjonen av de to domenene ved å bruke enzymkinetikk. Den relative orienteringen av de CM- og CDT-aktive stedene og eksistensen av substratkanalisering ble undersøkt ved røntgenkrystallografi. Totalt sett ble det ikke påvist noen katalytisk fordel eller substratkanalisering. SAXS-data samlet inn på noen ortologer og ikke inkludert i manuskriptet antyder at det bifunksjonelle enzymet ikke danner multimerer. Basert på den ekstremt lave affiniteten til CM-domenet for dets substrat ($K_m < 7 \mu\text{M}$), chorismat, og tilstedeværelsen av gener som koder for transportører og substratbindende

proteiner i samme operon, antar vi at disse bifunksjonelle fusjonsenzymer spiller en rolle i signalering, snarere enn biosyntesen av aromatiske aminosyrer. I tillegg (**Manuskript II**) undersøkte vi den kvaternære strukturen og assosiasjonen av eksporterte monomolekylære CM- og CDT-enzymmer funnet i periplasma av *Pseudomonas aeruginosa* (*PaeCM og *PaeCDT), som har en fusjonert CMPDT cytoplasmatisk motpart (PaeCMPDT). *P. aeruginosa* er et opportunistisk patogen som angriper pasienter med nedsatt immunforsvar (f.eks. kreft- eller HIV-pasienter). Strukturen til *PaeCM og *PaeCDT ble undersøkt ved hjelp av røntgenkrystallografi og småvinklet røntgenspredning (SAXS), mens assosiasjonen deres ble undersøkt ved hjelp av størrelseseksklusjonskromatografi (SEC). De innsamlede dataene tyder på at *PaeCDT og *PaeCM ikke danner et stabilt kompleks; vi samlet imidlertid innsikt i den dynamiske oppførselen til *PaeCM og *PaeCDT, som kan utnyttes i senere studier. Flere CM-klasser med helt forskjellige folder har blitt identifisert. Blant dem er AroQ δ -underklasse CM-er naturlig middelmådige enzymer, som trenger allosterisk aktivering mellom enzymer ved å danne et kompleks med DAHP-syntase (DS), det første enzymet i biosynteseveien for aromatiske aminosyrer. Vi tok som modellsystem CM fra *Mycobacterium tuberculosis* (MtCM), en AroQ δ -underklasse CM. Etersom *M. tuberculosis* er det forårsakende agens for tuberkulose sykdommen, kan en detaljert forståelse av MtCM-aktiveringen føre til utvikling av effektive legemidler. I **Manuskript III** tok vi sikte på å belyse de strukturfunksjonelle determinantene som ligger til grunn for allosterisk aktivering av MtCM. Vi utførte strukturelle studier og enzymkinetiske analyser rettet mot villtype MtCM (MtCM^{WT}), en laboratorieutviklet superaktiv variant (MtCM^V) og to varianter som øker den katalytiske aktiviteten med en faktor på seks (MtCM^{T52P}) og 12 (MtCM^{V55D}). Vi undersøkte den strukturelle betydningen av mutasjoner introdusert ved rettet evolusjon ved å sammenligne konformasjonsprøvetaking av tre MtCM-er: apo-vildtype (middelmådig) og overgangstilstandsanalog (TSA)-bundet villtype (MtCM^{LC}) og svært aktiv (laboratorieutviklet) ved hjelp av molekylær dynamikk simulering. Resultatene våre avslørte at T52P- og V55D-substitusjonene bidrar til å stivne en aktiv-site-sløyfe som bærer katalytisk viktige rester, og får dem forhåndsorganisert for katalyse. Til slutt tok vi sikte på å oppdage potente småmolekylforbindelser som gir ledetråder til inhibitorer og allosteriske modulatorer for MtCM ved fragmentbasert screening. Screeningen ga et utvalg av mulige ligander, som vil bli undersøkt nærmere i fremtidige studier.

Abbreviations	3
List of publications.....	7
1 Introduction.....	9
1.1 Understanding bacterial metabolic pathways.....	9
1.2 Aromatic amino acid biosynthesis pathway	10
1.2.1 Shikimate pathway- common route to the biosynthesis of the aromatic amino acids.....	10
1.2.2 Post-chorismate routes to the biosynthesis of aromatic amino acids	11
1.2.3 Subcellular localization of the aromatic amino acid biosynthesis pathway	13
1.2.4 Regulatory points in the biosynthesis of aromatic amino.....	13
1.2.4.1 DAHP synthase (DS)	14
1.2.4.2 Chorismate mutase (CM)	16
1.2.4.2.1 CM classes and folds.....	16
1.2.4.2.2 CM active site	18
1.2.4.3 Pathogenic organisms depending on the aromatic amino acid biosynthesis pathway..	18
2 Aim of the thesis	23
3 How to block the chain- understanding communication between sequential enzymes in the pathway	25
3.1 Enzyme colocalization	25
3.2 Substrate exchange mechanisms between colocalized enzymes	26
3.3 Model system: bifunctional extracytoplasmic *CDTCM and *CMCDT enzymes.....	29
3.3.1 Manuscript I, Summary	30
3.3.2 Further insights from the crystal and solution structures of the bifunctional CM/CDTs..	35
3.4 Model system: extracytoplasmic aromatic amino acid biosynthesis pathway enzymes from <i>Pseudomonas aeruginosa</i>	38
3.4.1 Manuscript II, Summary.....	39
3.5 Additional insights into CM and CDT domains	44
3.5.1 CDT domains.....	44
3.5.1.1 Hinge-mediated open and closed conformation of the active site of CDT domain.....	44
3.5.1.2 Intrinsically flexible region neighboring the β -strand (hinge region) of the CDT	46
3.5.1.3 Extended C-terminus enabling interactions between the large and small subdomains	47
3.5.2 CM domains	49
3.5.2.1 Binding affinity for chorismate for the CM domains of extracytoplasmic CM/CDT fusion enzymes	49
4 Jamming the engine - understanding fine details of enzyme function (allostery, new inhibitors)	53
4.1 Allostery in enzymes	53
4.2 Model system: CM and DAHP synthase from <i>Mycobacterium tuberculosis</i>	54
4.2.1 CM-DS inter-enzyme allostery.....	54
4.2.2 Allosteric CM-DS feedback regulation by aromatic amino acids.....	55
4.2.3 Manuscript III, Summary	58
4.3 Fragment-screening for MtCM (preliminary results from FragMAX and DSF)	61
5 Conclusions and future prospects	67
6 Materials and Methods.....	71
6.1 Recombinant gene expression and protein purification.....	71

6.1.1	Novel and exported bifunctional *CMCDT and *CDTCMs.....	71
6.1.2	Untagged and His ₆ -tagged *PaeCDT and *PaeCM	71
6.1.3	<i>Mycobacterium tuberculosis</i> CM wild type expression and purification.....	74
6.2	Differential scanning fluorimetry (DSF):	74
6.2.1	Introduction to the method	74
6.2.2	Experimental procedure.....	75
6.2.3	<i>Mycobacterium tuberculosis</i> chorismate mutase.....	75
6.2.4	Bifunctional *CDTCM and *CMCDT	76
6.3	Protein X-ray crystallography	76
6.3.1	Overview of the method	76
6.3.2	Protein crystallization	78
6.3.2.1	Introduction to the method.....	78
6.3.2.2	Crystallization experiments.....	79
6.3.2.3	Ligand-bound complexes crystallization	80
6.3.3	Data collection.....	81
6.3.4	Data processing and the structure determination.....	82
6.4	Fragment-based screening	85
6.4.1	Introduction to the method	85
6.4.2	Fragment-based screening using X-ray crystallography	85
6.4.2.1	Crystallization and soaking.....	86
6.4.2.2	Data collection and processing	89
6.4.3	Orthogonal validation using thermal shift assay	89
6.4.3.1	Experimental work.....	89
6.5	Small-angle X-ray scattering	94
6.5.1	Principles of the method.....	94
6.5.2	Sample preparation	95
6.5.3	Data collection and processing	95
7	References	99
8	Publications	111

Abbreviations

*AfCMCDT	<i>Aequoribacter fuscus</i> periplasmic chorismate mutase-cyclohexadienyl dehydratase fusion enzyme
*DsCDTCM	<i>Duganella sacchari</i> periplasmic cyclohexadienyl dehydratase-chorismate mutase fusion enzyme
*JbCDTCM	<i>Janthinobacterium sp.</i> periplasmic cyclohexadienyl dehydratase-chorismate mutase fusion enzyme
*MpCDTCM	<i>Massila phosphatilytica</i> periplasmic cyclohexadienyl dehydratase-chorismate mutase fusion enzyme
*MtCM	<i>Mycobacterium tuberculosis</i> ‘ secreted ’ chorismate mutase
*PaeCDT	<i>Pseudomonas aeruginosa</i> periplasmic cyclohexadienyl dehydratase
*PaeCM	<i>Pseudomonas aeruginosa</i> periplasmic chorismate mutase
*SbCMCDT	<i>Shewanella baltica</i> periplasmic chorismate mutase-cyclohexadienyl dehydratase fusion enzyme
*ScCMCDT	<i>Steroidobacter cummioxidans</i> periplasmic chorismate mutase-cyclohexadienyl dehydratase fusion enzyme
*TaCMCDT	<i>Thalassomonas actiniarum</i> periplasmic chorismate mutase-cyclohexadienyl dehydratase fusion enzyme
4-HPPY	4-hydroxy-phenylpyruvate
5-EPSP	5-enolpyruvylshikimate-3-phosphate
AA	anthranilate
AABP	amino acid binding protein
AAT	aminotransferases
ADT	arogenate dehydratase
Af	<i>Aequoribacter fuscus</i>
AGN	arogenate
Amp	ampicillin
AncCDT	ancestral solute-binding protein
AS	anthranilate synthase
Bis-Tris	bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
CDH	cyclohexadienyl dehydrogenase
CDT	cyclohexadienyl dehydratase
*CDTCM	Extracytoplasmic (periplasmic, exported), fusion enzyme, cyclohexadienyl dehydratase-chorismate mutase topology
CHA	chorismate
CM	chorismate mutase
CMPDT	cytoplasmic fusion enzyme, chorismate mutase-prephenate dehydratase topology
*CM/CDT	Extracytoplasmic (periplasmic, exported) chorismate mutase-prephenate dehydratase fusion enzyme, any topology
*CMCDT	Extracytoplasmic (periplasmic, exported) chorismate mutase-cyclohexadienyl dehydratase fusion enzyme
CS	chorismate synthase
Da	Dalton
DAHP	3-deoxy-DD-arabino-heptulosonate-7-phosphate
DHD	3-dehydroquinate dehydratase
DHDS	3-dehydroquinate synthase

DHD-SDH	3-dehydroquininate dehydratase-shikimate dehydrogenase
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DS	DAHP-synthase
Ds	<i>Duganella sacchari</i>
DSF	Differential Scanning Fluorimetry
<i>E. coli</i>	<i>Escherichia coli</i>
E4P	erythrose-4-phosphate
EcCM	<i>Escherichia coli</i> chorismate mutase
EPSP	5-enolpyruvylshikimate-3-phosphate
ES	5-enolpyruvylshikimate-3-phosphate synthase
ESRF	European Synchrotron Radiation Facility
ETH Zürich	Die Eidgenössische Technische Hochschule Zürich
FPLC	fast protein liquid chromatography
HIV	human immunodeficiency virus
IPTG	isopropyl β -D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
Jb	<i>Janthinobacterium sp.</i> HH01
MD	molecular dynamics
MES	2-(<i>N</i> -morpholino)ethanesulfonic acid
Mp	<i>Massila phosphatilytica</i>
MPD	2-methyl-2,4-pentadiol
Mt	<i>Mycobacterium tuberculosis</i>
MtCM	<i>Mycobacterium tuberculosis</i> cytoplasmic chorismate mutase
MtCM ^{3p3}	<i>Mycobacterium tuberculosis</i> cytoplasmic chorismate mutase, Thr52Pro-Val55Asp double variant
MtCM ^{T52P}	<i>Mycobacterium tuberculosis</i> cytoplasmic chorismate mutase, Thr52Pro variant
MtCM ^{T52P}	<i>Mycobacterium tuberculosis</i> cytoplasmic chorismate mutase, Val55Asp variant
MtCM ^V	<i>Mycobacterium tuberculosis</i> cytoplasmic chorismate mutase, superactive variant
MtCM ^{WT}	<i>Mycobacterium tuberculosis</i> cytoplasmic chorismate mutase, wild-type
MtDS	<i>Mycobacterium tuberculosis</i> DAHP-synthase
NMA	normal mode analysis
NMR	nuclear magnetic resonance
Pae	<i>Pseudomonas aeruginosa</i>
PaeCMPDT	<i>Pseudomonas aeruginosa</i> cytoplasmic chorismate mutase- prephenate dehydratase fusion enzyme
PanDDA	Pan-Dataset Density Analysis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PDH	prephenate dehydrogenase
PDT	prephenate dehydratase
PEG	polyethylene glycol
PEP	phosphoenolpyruvate
PPA	prephenate
PPY	phenylpyruvate
RNA	ribonucleic acid

r.m.s.d.	Root mean square deviation
rpm	revolutions per minute
SAXS	small-angle X-ray scattering
Sb	<i>Shewanella baltica</i>
SBP	solute-binding protein
Sc	<i>Saccharomyces cerevisiae</i>
ScCM	<i>Saccharomyces cerevisiae</i> chorismate mutase
SDH	shikimate dehydrogenase
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
SK	Shikimate kinase
SPR	surface plasmon resonance
Ta	<i>Thalassomonas actiniarum</i>
Tris	tris(hydroxymethyl)aminomethane
TSA	Transition State Analog

List of publications

This thesis is based on the following publications, which will be referred to in the text by their Roman numerals:

I. Stocker, C., **Khatanbaatar, T.**, Würth-Roderer, K., Cordara, G., Kregel, U., and Kast, P. Novel exported bifunctional fusion enzymes with chorismate mutase and cyclohexadienyl dehydratase activity: shikimate pathway enzymes teamed up in no man's land.

II. **Khatanbaatar, T.**, Bressan, L., Würth-Roderer, K., Kast, P., and Kregel, U. Structural analysis of chorismate mutase and cyclohexadienyl dehydratase from *Pseudomonas aeruginosa*.

III. Thorbjørnsrud, H.V., Bressan, L., **Khatanbaatar, T.**, Carrer, M., Würth-Roderer, K., Cordara, G., Kast, P., Cascella, M., and Kregel, U. (2023). "What drives chorismate mutase to top performance? Insights from a combined in silico and in vitro study." Biochemistry.

1 Introduction

1.1 *Understanding bacterial metabolic pathways*

Despite their essential role in sustaining life, a small percentage of bacteria infect humans, causing major health problems (Wilson *et al.* 2002, Doron and Gorbach 2008). The discovery of antibiotics was one of the greatest medical breakthroughs in last century. Antibiotics are natural products of microorganisms, originally used to fight each other (Hutchings *et al.* 2019). Before the discovery of penicillin, the first microbial antibiotic (1928), German scientist Paul Ehrlich developed in the beginning of the 1900s the first synthetic antibiotic, the arsenic-based *salvarsan* (Kohanski *et al.* 2010, Hutchings *et al.* 2019). Since then, hundreds of natural and synthetic antibiotics have been developed to fight against infectious bacteria. Antibiotics kill bacteria by targeting cellular functions essential for survival such as DNA replication, protein biosynthesis, energy metabolism and cell wall synthesis (Kohanski *et al.* 2010, Walsh and Wencewicz 2020). For example, the first-line antitubercular drugs isoniazid and ethambutol target *Mycobacterium tuberculosis* cell wall synthesis and protein synthesis, respectively (Timmins and Deretic 2006, Zhu *et al.* 2018).

Misuse of antibiotics led to the emergence of drug-resistant strains. Bacteria adapt to resist the action of drugs made to kill them, thereby requiring the discovery of drugs based on novel principles of action (Tenover and McGowan 1996, Michael *et al.* 2014, Darby *et al.* 2022). The first critical step in drug discovery is to identify a molecular target. The target needs to be essential for the survival of the pathogen and non-homologous to proteins within the host organism (Bakheet and Doig 2009, Serral *et al.* 2021, Khan *et al.* 2022). Novel bacterial targets can be identified by experimental and computational methods (Weinstein *et al.* 1997, Sundberg 2000, Zheng *et al.* 2004). A possible experimental approach is to screen a library of small compounds against a potential target using high-throughput screening techniques (Mayr and Bojanic 2009, Hughes *et al.* 2011). Hits obtained by high-throughput screening are further optimized to yield so-called ‘lead’ molecules, a step which is followed by lead optimization, drug candidate selection and finally three stages of clinical trials, taking 10-15 years before approved for marketing (Hughes *et al.* 2011, Mohs and Greig 2017).

Availability of extensive genomic databases for pathogenic bacteria and the full human genome sequence has made computational methods more competitive, as they require less

time and labor compared to experimental methods (Lionta *et al.* 2014, Sliwoski *et al.* 2014). Once a possible drug target has been identified, its structure can be determined and used in structure-based drug discovery (Greer *et al.* 1994, Lionta *et al.* 2014, Batool *et al.* 2019). Structural determinants of potential binding cavities serve as the basis to identify efficient ligands (Greer *et al.* 1994, Batool *et al.* 2019). Detailed structural knowledge of a drug target protein allows to employ computational techniques such as structure-based virtual screening, molecular docking, and molecular dynamics simulations, significantly shortening time and cost of conventional drug design (Meng *et al.* 2011, Lionta *et al.* 2014).

1.2 Aromatic amino acid biosynthesis pathway

Aromatic amino acids are essential building blocks for proteins. Humans lack enzymes that produce aromatic amino acids, thus they must consume them through their diet (Matthews 2007, Sprenger 2007, Kapnick and Zhang 2008). In contrast, plants, bacteria, and fungi are capable of *de novo* synthesis of aromatic amino acids (Gilchrist and Kosuge 1980, Sprenger 2007). The synthesis of proteinogenic aromatic amino acids takes place through the aromatic amino acid biosynthesis pathway (Gilchrist and Kosuge 1980, Sprenger 2007). The pathway itself is compounded by several metabolic routes: a common pathway, the shikimate pathway, leading to the synthesis of the branch point compound chorismate, and three terminal pathways, in which chorismate is converted to phenylalanine, tyrosine and tryptophan, respectively (Figure 1) (Herrmann and Weaver 1999, Maeda and Dudareva 2012).

1.2.1 Shikimate pathway- common route to the biosynthesis of the aromatic amino acids

The biosynthesis of the three aromatic amino acids starts with a shared route, the shikimate pathway (Figure 1) (Weaver and Herrmann 1997, Herrmann and Weaver 1999, Maeda and Dudareva 2012). The shikimate pathway utilizes the products of glycolysis and pentose phosphate pathway as substrates (phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P), respectively) and ends with the synthesis of chorismate (CHA) (Herrmann and Weaver 1999, Maeda and Dudareva 2012). The pathway, which starts with the condensation of PEP and E4P to 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) catalyzed by DAHP-synthase (DS), consists of seven enzymatic reactions and six intermediate reaction products (Figure 1). Most shikimate pathway enzymes are monofunctional in bacteria, whereas in plants the enzyme catalyzing the third and fourth steps is bifunctional (3-

dehydroquinase dehydratase-shikimate dehydrogenase, DHD-SDH) (Weaver and Herrmann 1997, Maeda and Dudareva 2012). The intermediates of the shikimate pathway serve as starting points for several secondary aromatic metabolites in both prokaryotes and eukaryotes (Gilchrist and Kosuge 1980, Becerril *et al.* 1989, Maeda and Dudareva 2012, Aversch and Kromer 2018). These include alkaloids, aromatic vitamins (phylloquinone and folic acid), and flavonoids (Gilchrist and Kosuge 1980, Becerril *et al.* 1989, Maeda and Dudareva 2012, Aversch and Kromer 2018).

1.2.2 Post-chorismate routes to the biosynthesis of aromatic amino acids

Chorismate, the final product of the shikimate pathway, is defined as a ‘branch-point metabolite’ (Weaver and Herrmann 1997, Herrmann and Weaver 1999). Three metabolic routes branch off from the shikimate pathway, leading to the biosynthesis of the three proteinogenic amino acids (Pittard and Yang 2008, Maeda and Dudareva 2012) (Figure 1). Anthranilate synthase (AS) catalyzes the conversion of chorismate to anthranilate (AA), which is then converted to tryptophan by four more enzymatic reactions (Figure 1) (Gilchrist and Kosuge 1980, Pittard and Yang 2008). Chorismate mutase (CM) catalyzes the conversion of chorismate to prephenate (PPA), which can be converted to either phenylpyruvate (PPY) or 4-hydroxy-phenylpyruvate (4-HPPY) by prephenate dehydratase (PDT) and prephenate dehydrogenase (PDH), respectively (Pittard and Yang 2008, Maeda and Dudareva 2012) (Figure 1). The transamination of PPY and 4-HPPY by aminotransferases (AAT) leads to phenylalanine and tyrosine, respectively (Pittard and Yang 2008, Maeda and Dudareva 2012) (Figure 1). In plants, the biosynthesis of phenylalanine and tyrosine from chorismate proceeds via arogenate (AGN), which is a transamination product of prephenate (Maeda and Dudareva 2012, Jan *et al.* 2021). Arogenate can be converted to both tyrosine and tryptophan by arogenate/cyclohexadienyl dehydrogenase (CDH) and cyclohexadienyl dehydratase (CDT) (Maeda and Dudareva 2012, Qian *et al.* 2019) (Figure 1).

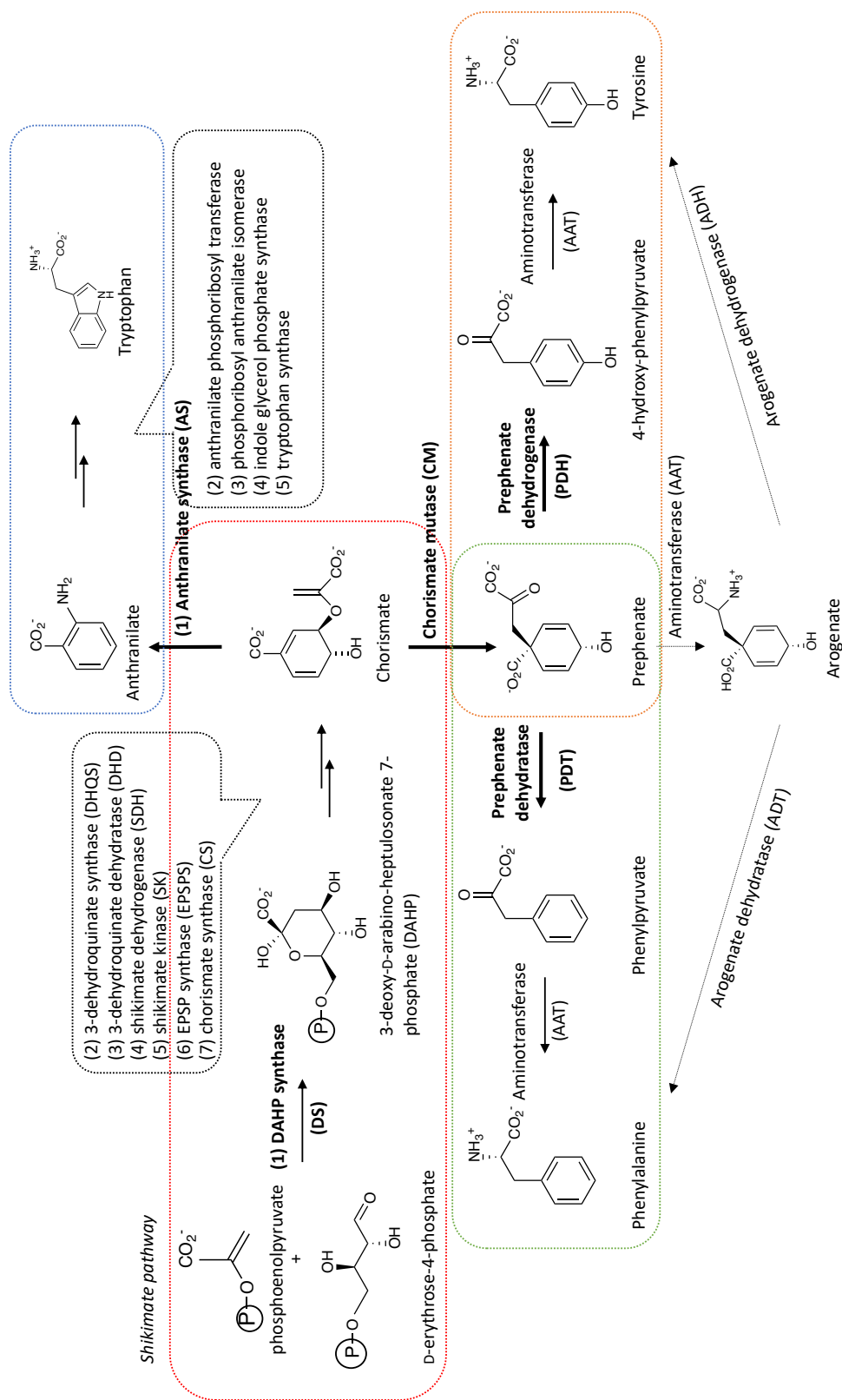


Figure 1. The biosynthesis pathway for aromatic amino acids. The first enzyme 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase catalyzes the aldol-like condensation of PEP and E4P to produce a seven-carbon molecule, DAHP. The subsequent enzymes in the pathway (3-dehydroquininate synthase (DHQS), 3-dehydroquininate dehydratase (DHD), shikimate dehydrogenase (SDH), shikimate kinase (SK), 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), and chorismate synthase (CS)) converts DAHP to chorismate. Chorismate is an end-product of shikimate pathway, and a common precursor for aromatic amino acids (Herrmann and Weaver 1999, Maeda and Dudareva 2012). From chorismate, terminal pathways lead to the biosynthesis of perspective aromatic amino acids. In plant, biosynthesis of the aromatic amino acids proceeds mainly via arogenate (dash arrow) (Weaver and Herrmann 1997, Herrmann and Weaver 1999).

1.2.3 Subcellular localization of the aromatic amino acid biosynthesis pathway

In bacteria, the biosynthetic pathway of aromatic amino acids is localized in the cytoplasm (Tzin *et al.* 2012). However, post-chorismate pathway enzymes equipped with a N-terminal signal sequence characteristic for periplasmic transport have been found in some pathogenic bacteria (Xia *et al.* 1993, Calhoun *et al.* 2001, Sasso *et al.* 2005). For example, *Pseudomonas aeruginosa*, an opportunistic pathogen targeting immunocompromised patients (*e.g.* cancer or HIV patients) (Basseti *et al.* 2018), encodes for periplasmic isoforms of chorismate mutase, cyclohexadienyl dehydratase, and several aminotransferases (Fiske *et al.* 1983, Zhao *et al.* 1992, Tam and Saier 1993, Zhao *et al.* 1993, Calhoun *et al.* 2001). Cyclohexadienyl dehydratase exhibits both prephenate dehydratase (PDT) and arogenate dehydratase activity (ADT), hinting to its possible role in the periplasmic synthesis of phenylalanine (Fiske *et al.* 1983, Zhao *et al.* 1992, Zhao *et al.* 1993, Clifton *et al.* 2018). This hypothesis has been formulated by Jensen, R.A and coworkers, who went as far as postulating the existence of a periplasmic *spillover pathway* for the synthesis of aromatic amino acids (Zhao *et al.* 1992, Zhao *et al.* 1993). *M. tuberculosis*, the causative agent of the disease tuberculosis (Sudre *et al.* 1992), encodes two chorismate mutases, a cytoplasmic (MtCM) and an extracellular isoform (*MtCM) (Sasso *et al.* 2005, Okvist *et al.* 2006, Kim *et al.* 2008). Due to the lack of a periplasmic space in *M. tuberculosis*, *MtCM may be secreted to the extracellular space (henceforth referred to as ‘secreted CM’) (Sasso *et al.* 2005).

Plant enzymes involved in the aromatic amino acid biosynthesis are expressed with N-terminal sequences characteristic for plastid import (Weaver and Herrmann 1997, Herrmann and Weaver 1999). However, some shikimate enzymes (*e.g.* shikimate kinase and 5-EPSP synthase) can be found in plant cytosol and are even functional with their unprocessed signal sequence (Della-Cioppa *et al.* 1986, Schmid *et al.* 1992). In contrast, cytosolic chorismate synthase is not enzymatically active when its signal sequence is attached, thereby cannot produce chorismate in cytosol (Henstrand *et al.* 1995). Hence, the biosynthesis of the aromatic amino acids in plants occur exclusively in plastids (Weaver and Herrmann 1997).

1.2.4 Regulatory points in the biosynthesis of aromatic amino

Metabolic pathways are subjected to tight regulation, and the aromatic amino acids metabolic pathway is no exception (Yates and Pardee 1956) (Weaver and Herrmann 1997). Enzymes involved in a specific metabolic pathway are colocalized within the organism, and often

associate with each other for more efficient performance (Kohnhorst *et al.* 2017). Furthermore, metabolic pathways are often regulated either through feedback inhibition by the end-products or allosteric regulation via enzyme-enzyme interactions (Motlagh *et al.* 2014) (Light and Anderson 2013). Here we focus on chorismate mutase and DAHP synthase, their interplay and how they act as regulatory points for the aromatic amino acids biosynthesis pathway.

1.2.4.1 DAHP synthase (DS)

As the first enzyme in the shikimate pathway, DAHP synthase (DS) is the main gateway to the metabolic route. Both plants and bacteria genomes encode for several DAHP synthase paralog genes, yielding different isozymes (Doy and Brown 1965, Brown and Doy 1966, Bentley and Haslam 1990, Dyer *et al.* 1990, Keith *et al.* 1991). All DAHP synthases contains a conserved Cys-X-X-His motif for binding metal-ions, mainly Mn^{2+} and Co^{2+} , which are required for catalysis (Stephens and Bauerle 1991, Stephens and Bauerle 1992, Herrmann and Weaver 1999). Based on their sequence, DAHP synthases can be classified into two classes: type I, small (39 kDa) and usually found in bacteria, and type II, large (54 kDa) and usually found in plants (Bentley and Haslam 1990, Walker *et al.* 1996) (Figure 2). The two DAHP synthases classes carry a common $(\beta/\alpha)_8$ -barrel fold, despite less than 10% sequence identity (Walker *et al.* 1996, Jensen *et al.* 2002, Webby *et al.* 2005) (Figure 2).

Both type I and type II DAHP synthases have unique amino acid binding elements; however, the binding site position relative to the core barrel structural motif differs between the two classes (Webby *et al.* 2005). A different regulation mechanism underlies the structural difference between bacterial and plant DAHP synthase classes. The bacterial DAHP synthases are subjected to allosteric inhibition by aromatic amino acids, making it an important regulatory point in the pathway (Figure 2) (Pinto *et al.* 1988, Sasso *et al.* 2009, Burschowsky *et al.* 2018). *Escherichia coli* encodes three isozymes of Type I DAHP synthase (genes *aroF*, *aroG* and *aroH*); they are inhibited by tyrosine, and phenylalanine, and tryptophan, respectively – the terminal product of the amino acid biosynthesis pathway (Schoner and Herrmann 1976, McCandliss *et al.* 1978, Ray and Bauerle 1991). The Tyr- and Trp-sensitive DAHP synthases are dimers, while Phe-sensitive isozyme forms a tetramer (Figure 2). The DAHP synthase activity is decreased by tryptophan to 45% of the wild-type enzyme, while Tyr and Phe provided almost complete inhibition (Ray and Bauerle 1991,

Akowski and Bauerle 1997). The residual activity of the aromatic amino acid-inhibited DS is required in bacteria for the biosynthesis of other aromatic compounds (Herrmann and Weaver 1999). *M. tuberculosis* DAHP synthase enzyme, which is co-regulated by all three amino acids, is representative of the Type II class (Munack *et al.* 2016).

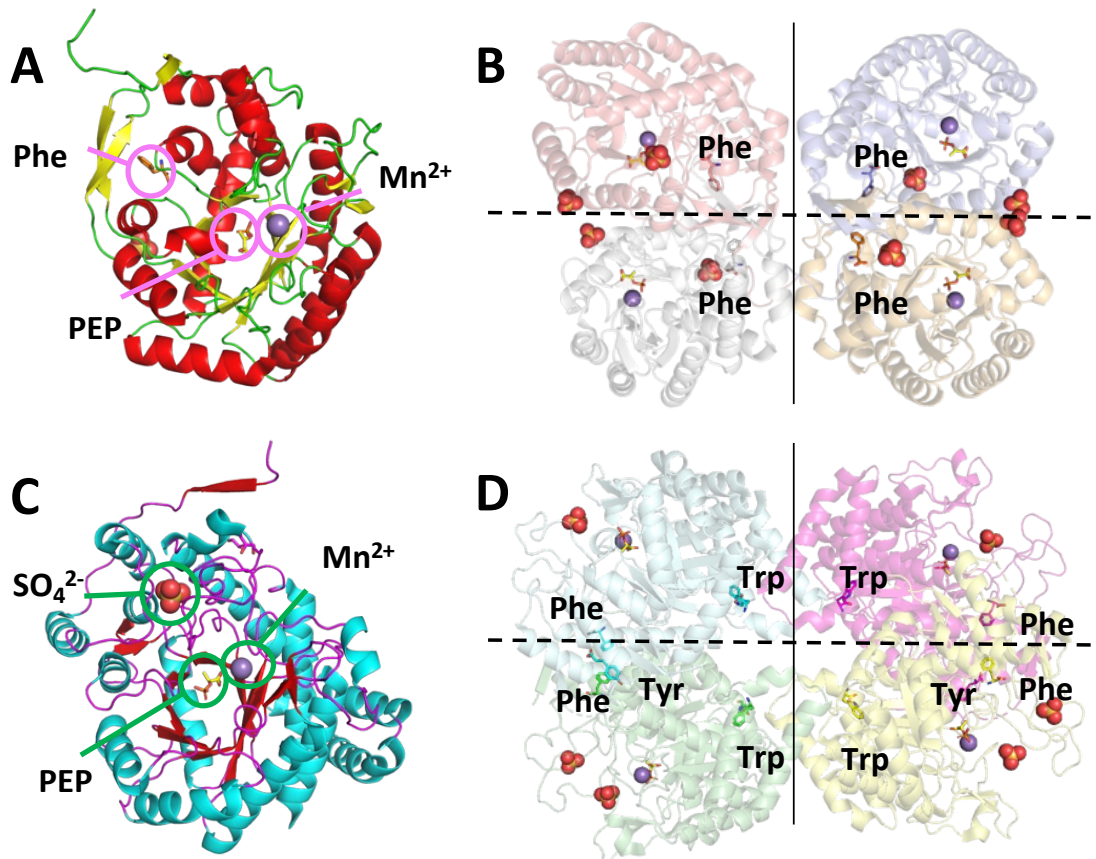


Figure 2. Structural classes of DAHP synthases. A. Secondary structure of Type I DAHP synthase, exemplified by *E. coli* Phe-sensitive DAHP synthase isozyme bound to the substrate (PEP), metal ion (Mn^{2+}) and Phe (PDB ID: 1KFL, (Shumilin *et al.* 2002)). B. Quaternary structure of Phe-sensitive DAHP synthase from *E. coli*. (PDB ID: 1KFL, (Shumilin *et al.* 2002)) C. Secondary structure of Type II DAHP synthase, exemplified by *M. tuberculosis* DAHP synthase complexed with PEP and Mn^{2+} (PDB ID: 3NV8, (Webby *et al.* 2010)). D. Quaternary structure of fully inhibited DAHP synthase from *M. tuberculosis*, (PDB ID: 3NV8, (Webby *et al.* 2010)). The dimerization interface is shown as dash line and tetramerization interface is shown as continuous line. Feed-back regulators are shown as sticks in B and D.

Plants encode two distinct DAHP synthase genes whose expression level varies with environmental stress such as wounding and bacterial infections (Dyer *et al.* 1989, Keith *et al.* 1991). According to the current understanding, the plant enzymes escape feedback inhibition by any amino acids (Suzich *et al.* 1985).

1.2.4.2 Chorismate mutase (CM)

Chorismate mutase is an important regulatory point in the aromatic amino acids biosynthesis pathway, directing it either towards the synthesis of Tyr or the synthesis of Phe (Herrmann and Weaver 1999). It catalyzes the conversion of chorismate to prephenate via a chair-like transition state (Claeyssens *et al.* 2011). As one of few natural enzymes catalyzing a pericyclic reaction, CMs are of great interest to scientists and have long served as a model for mechanistic enzymology studies (DeClue *et al.* 2005, Künzler *et al.* 2005) (Figure 3 A).

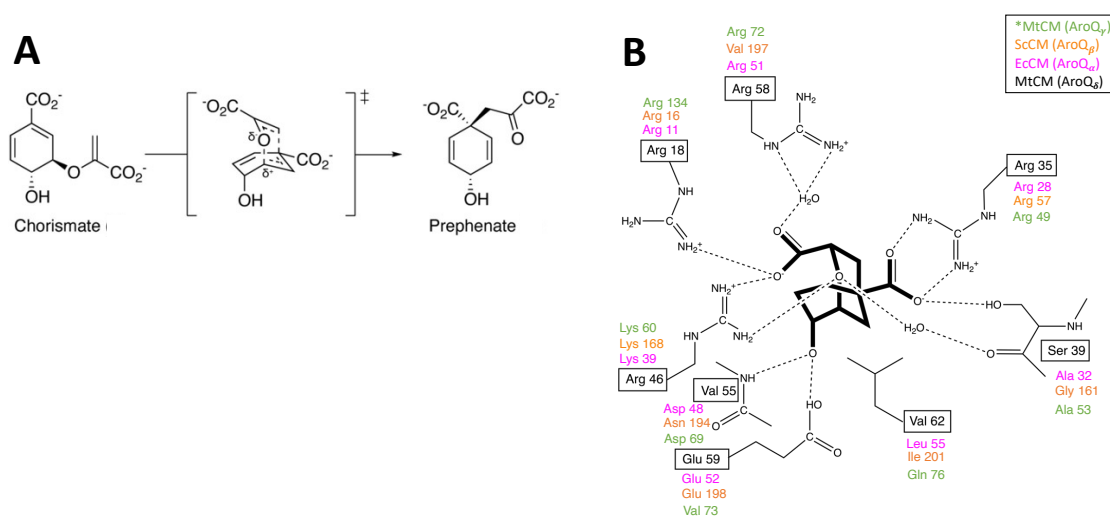


Figure 3. Chorismate mutase catalysis. **A.** Conversion of chorismate to prephenate via a chair-like transition state. **B.** The active site scheme of AroQ δ class CM, exemplified by MtCM (PDB ID: 2VKL, (Sasso *et al.* 2009); black) and the corresponding residues in *MtCM (PDB ID: 2FP2, (Okvist *et al.* 2006); green), EcCM (PDB ID: 1ECM, (Lee *et al.* 1995); magenta) and ScCM (PDB ID: 4CSM, (Sträter *et al.* 1997); orange). Ligand (black stick, bold) is Bartlett's transition state analog (TSA) (Bartlett and Johnson 1985)

1.2.4.2.1 CM classes and folds

Two classes of chorismate mutase with completely different folds have been discovered: AroH, trimeric pseudo- α/β -barrel, and AroQ, dimeric all α -helical (Chook *et al.* 1994, MacBeath *et al.* 1998) (Figure 4). The AroH class is rare in nature and exemplified by CM from *Bacillus subtilis* (Chook *et al.* 1993, Chook *et al.* 1994). The AroQ class is more abundant and divided into four subclasses; AroQ α to AroQ δ (MacBeath *et al.* 1998) (Figure 4). The holotype for the AroQ α subclass is CM from the *E. coli* 'P-protein', a CM-PDT fusion enzyme, which forms an intertwined dimer of three α -helices (Dopheide *et al.* 1972, Gething and Davidson 1976, Lee *et al.* 1995) (Figure 4). The AroQ β -subclass is represented by CM from *Saccharomyces cerevisiae* (ScCM) (Sträter *et al.* 1997) (Figure 4 C). Each ScCM AroQ β -subclass protomer consists of twelve α -helices and has one allosteric binding

site (Sträter *et al.* 1997) (Figure 4 C). The last two subclasses, AroQ_γ and AroQ_δ, can both be found in *M. tuberculosis*. The AroQ_γ subclass of CMs are composed of six α-helices (Figure 4 E) and expressed with a N-terminal signal sequence, which directs the protein out of the cytoplasm, either to the periplasm or extracellular space, and is cleaved upon translocation (Calhoun *et al.* 2001, Sasso *et al.* 2005, Okvist *et al.* 2006).

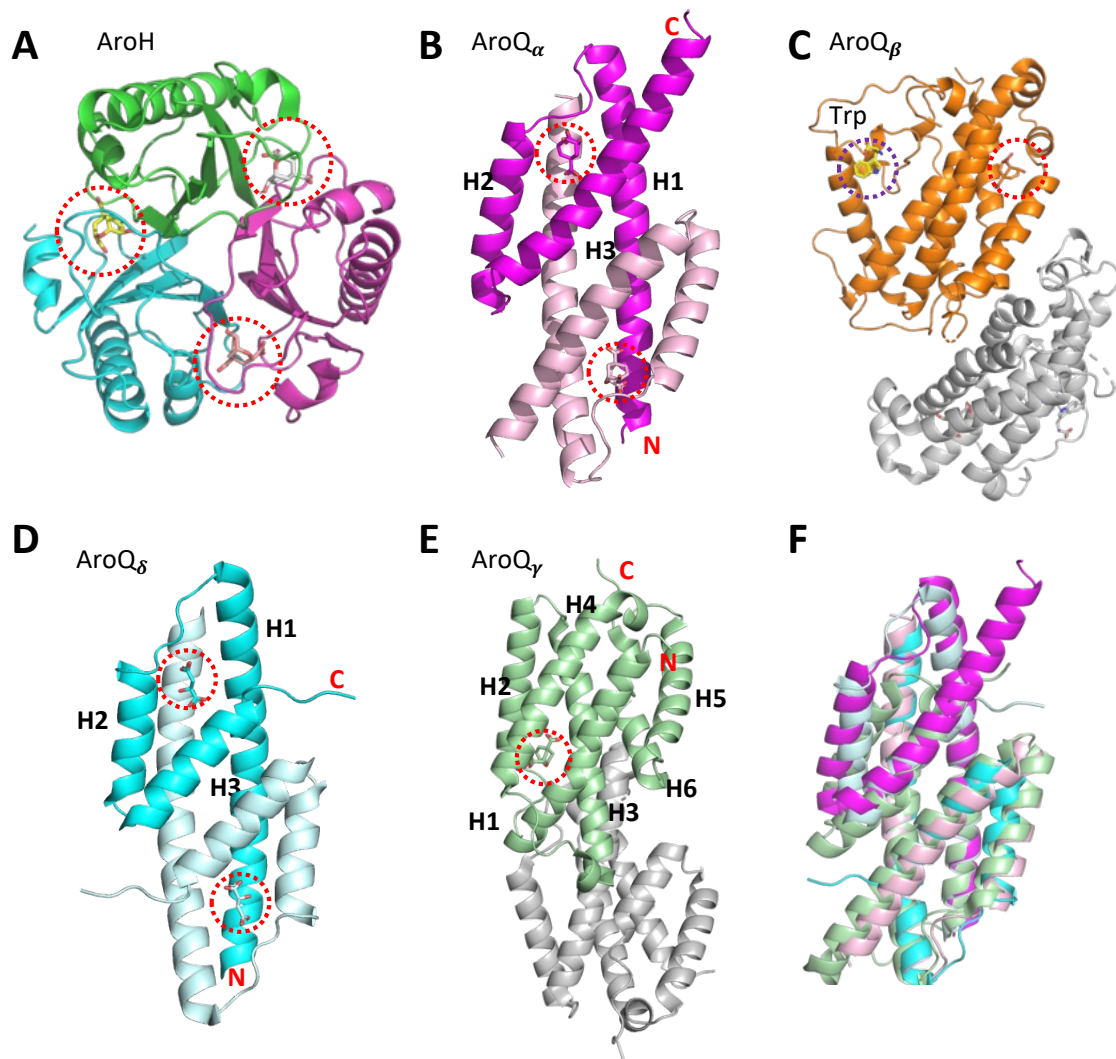


Figure 4. Structural classes of CMs. **A.** Trimeric AroH CM from *B. subtilis*. **B.** Dimeric AroQ_α CM from *E. coli* (PDB ID: 1ECM, (Lee *et al.* 1995)). **C.** Dimeric AroQ_β CM from *Saccharomyces cerevisiae* (PDB ID: 4CSM, (Sträter *et al.* 1997)). Phe is depicted as yellow stick, highlighted by a purple, dash circle. **D.** Dimeric structure of AroQ_δ CM from *M. tuberculosis* (PDB ID: 2VKL, (Sasso *et al.* 2009)). **E.** Dimeric AroQ_γ CM from *M. tuberculosis* (PDB ID: 2FP2, (Okvist *et al.* 2006)) **F.** Structural alignment of AroQ class CMs: EcCM (pink), MtCM (cyan), *MtCM (green). The active sites are highlighted by red, dash circles in A-E.

The AroQ_γ class of CMs are often found in bacteria as well as in fungi and nematodes (Sträter *et al.* 1997, Lambert *et al.* 1999, Bauters *et al.* 2020, Winston *et al.* 2022). Since the

amino acids biosynthesis pathway is localized in cytoplasm, the existence of CMs in periplasmic space or secreted out of the cell raises question about their biological function. Some speculated on their role as virulence factors of the producer organisms, but no conclusive evidence has been collected to corroborate the hypothesis (Xia *et al.* 1993, Calhoun *et al.* 2001, Sasso *et al.* 2005, Assis *et al.* 2021). The AroQ δ subclass is represented by the intracellular CM from *M. tuberculosis* (Sasso *et al.* 2009). AroQ δ CMs resemble the fold of the *E. coli* CM (AroQ α), forming an intertwined dimer of three helices, but they have a shorter sequence (Sasso *et al.* 2009) (Figure 4 D).

1.2.4.2.2 CM active site

The active site of chorismate mutase carries a strong positive charge (Figure 3 B), matching the need to stabilize a negatively charged transition state (Burschowsky *et al.* 2014). The active site of AroQ α -subclass and AroQ δ -subclass CMs is located within a four-helix bundle formed by three helices (H1-H3) from one protomer and H1' from the other protomer (Figure 4 B, D) (Lee *et al.* 1995). This leads to two active sites per dimer, contributed by structural elements from both protomers (Lee *et al.* 1995).

The Aro β and AroQ γ CMs carry one active site per protomer, nested inside a four-helix bundle formed within the protomer itself, without contributions from the second protomer (Sträter *et al.* 1997, Okvist *et al.* 2006, Sasso *et al.* 2009). The catalytic machinery is well-conserved among different class of CMs (Kast *et al.* 1996), accelerating the chorismate to prephenate conversion by 10⁶-fold (Sasso *et al.* 2009, Munack *et al.* 2016, Burschowsky *et al.* 2018). However, the AroQ δ subclass CM shows an extremely low enzymatic activity on its own, requiring allosteric activation by an external partner (Sasso *et al.* 2009, Munack *et al.* 2016, Burschowsky *et al.* 2018). For example, the enzymatic activity of the cytosolic CM from *M. tuberculosis* and *Corynebacterium glutamicum* is boosted upon the non-covalent complex formation with DAHP synthase (Munack *et al.* 2016, Burschowsky *et al.* 2018).

1.2.4.3 Pathogenic organisms depending on the aromatic amino acid biosynthesis pathway

All intermediates and end-products of the aromatic amino acids biosynthesis pathway are precursors for secondary metabolites, many of which are essential for the viability of their host (Maeda and Dudareva 2012, Aversch and Kromer 2018). For example, *p*-aminobenzoic

acid serves as a precursor of folate, which participates in the biosynthesis of nucleotides, and aromatic amino acids are involved in protein synthesis (Averesch and Kromer 2018). Due to its absence in vertebrates, enzymes from the shikimate pathway are valuable targets in the design of bactericides, fungicides, herbicides, and anti-apicomplexan parasitic drug discovery (Gilchrist and Kosuge 1980, Sprenger 2007).

Human apicomplexan-parasitosis

The shikimate pathway is present in the plastid-like organelle of apicomplexan parasites such as *Plasmodium falciparum*, *Toxoplasma gondii* and *Cryptosporidium parvum* (Roberts *et al.* 1998). Apicomplexan parasites are obligate intracellular parasites, which cause a great harm for humans (Morrisette and Sibley 2002). For example, *P. falciparum* causes malaria, a life-threatening disease, while *T. gondii* and *C. parvum* are opportunistic parasites in immunocompromised patients, such as HIV patients and cancer patients (Fichera and Roos 1997). The inhibition of metabolic enzymes in these parasites has shown a promising effect. During chemotherapy, for example, the inhibition of chorismate mutase by 6-fluorishikimate reduced the growth of *P. falciparum* (McConkey 1999).

Plant parasitosis

Nematodes are roundworms that feed on fungi, bacteria, plants, protozoans, and other nematodes (Bird and Bird 2012). The plant-parasitic nematodes manipulate their host for their own benefit by injecting secretions from their esophageal glands (Lambert *et al.* 1999, Jones *et al.* 2003, Bauters *et al.* 2020). Despite the lack of a shikimate pathway, an AroQ γ subclass CM with a signal-sequence is found in secretions of some nematodes, such as cyst nematodes *Globodera pallida* and *Heterodera schachtii*, and root-knot nematode *Meloidogyne javanica*, and is thought to be important in the parasite-host interaction (Lambert *et al.* 1999, Jones *et al.* 2003, Vanholme *et al.* 2009). The CM from parasites can modulate the plant's endogenous chorismate metabolism, thus the level of a variety of compounds important for plant physiology, such as auxin, salicylic acids, and aromatic amino acids (Lambert *et al.* 1999, Jones *et al.* 2003, Vanholme *et al.* 2009).

Mycosis

Fungi cause both human and plant disease. Being eukaryotes, their similarity to the host complicates treatment of fungal infections. *Paracoccidioides brasiliensis* causes a systemic granulomatous disease endemic in Latin America (Bueno *et al.* 2019, Rodrigues-Vendramini *et al.* 2019). The treatment of *paracoccidioidomycosis* using clinical drugs takes several months to a year (Bueno *et al.* 2019, Rodrigues-Vendramini *et al.* 2019). Seixas, F.A and coworkers (2019) targeted the last shikimate pathway enzyme, chorismate synthase (Rodrigues-Vendramini *et al.* 2019). The best-hit identified in their study “CP1” ((1*S*,2*S*,3*aS*,4*S*,9*bR*)-1-chloro-6-nitro-2-(2-nitrophenyl)sulfanyl-2,3,3*a*,4, 5,9*b*-hexahydro-1*H*-cyclopenta[*c*]quinoline-4-carboxylic acid) showed a promising antifungal activity with no cytotoxic effect on human cell lines (Bueno *et al.* 2019, Rodrigues-Vendramini *et al.* 2019). The secreted chorismate mutases from fungi are hypothesized to be a virulence factor and can be a target for antifungal drugs (Djamei *et al.* 2011). For example, the secretome of fungus *Ustilago maydis*, causing maize smut, contains a chorismate mutase ortholog which alters the metabolic status of maize cells (Djamei and Kahmann 2012, Schilling *et al.* 2014).

Bacterial infection / tuberculosis

The emergence of drug-resistant bacterial strains caused a reduction in effectiveness of current antibiotics (Tenover and McGowan 1996, Darby *et al.* 2022). The current treatment of tuberculosis, caused by pathogenic *Mycobacterium tuberculosis*, includes four first-line drugs (isoniazid, rifampicin, ethambutol, and pyrazinamide) which target the host’s cell wall synthesis, DNA synthesis, and fatty acid synthesis (Kapnick and Zhang 2008, Palomino and Martin 2014). However, about half a million people out of 10 million cases of tuberculosis in 2018 developed a drug resistance (Kapnick and Zhang 2008, Palomino and Martin 2014, Tiberi *et al.* 2021). The rise of *multi-*, *extensively-*, *extremely-* and *total-*drug resistant strains requires new antituberculosis drugs based on a novel principle of action. More than 70% drug resistant strains accounts for a multi-drug resistance, which is insensitive to two or more first-

line drugs (Tiberi *et al.* 2021). The total drug-resistant strains cannot be killed by any available drug. In *M. tuberculosis*, all shikimate pathway enzymes have been identified, thus many attempts were made to target this pathway against tuberculosis (Kapnick and Zhang 2008). For example, IMB-T130 ((5-chloro-N-[4-(pyridin-2-yl)-1,3-thiazol-2-yl] thiophene-2-carboxamide)) shows an excellent inhibition on DHQ synthase of *M. tuberculosis* with IC₅₀ value of 0.87 µg/mL (Zhu *et al.* 2018). This compound also exhibits a promising inhibitory activity against drug-resistant tuberculosis (Zhu *et al.* 2018). The shikimate pathway enzymes have also been investigated for vaccine development. The virulence of the auxotrophic mutants of *M. tuberculosis*, deficient in the biosynthesis of tryptophan and proline, was attenuated in immunodeficient mice (Smith *et al.* 2001). The deletion of the EPSP synthase in *Salmonella typhimurium* and in *Listeria monocytis* also showed a significant attenuation in mice (Kapnick and Zhang 2008).

2 Aim of the thesis

The overall aim of this PhD thesis is to investigate the molecular mechanisms regulating selected enzymes from the aromatic amino acid biosynthesis pathway, absent in mammals. Detailed knowledge on structure-function relationship of enzyme activation and regulation will lead the way to design of specific inhibitors or harness their power for the synthesis of fine chemicals.

Our first objective was to study substrate-exchange mechanism within enzymes catalyzing sequential reactions. In this context, we carried out kinetic and structural studies on exported (AroQ γ subclass) chorismate mutase (CM) and cyclohexadienyl dehydratase (CDT) enzymes, which can be found as both monofunctional and bifunctional/fusion forms. The second objective was to elucidate the mechanism underlying allosteric regulation of the enzymatic activity of cytoplasmic AroQ δ -subclass CMs. For this purpose, laboratory-evolved, highly active CMs derived from a naturally sluggish CM from *Mycobacterium tuberculosis* were used as model systems for molecular dynamics study. Also, we tried to discover by fragment-based screening potent small-molecule compounds providing leads to MtCM inhibitors and allosteric regulators that could potentially be developed to drug against tuberculosis.

3 How to block the chain- understanding communication between sequential enzymes in the pathway

3.1 Enzyme colocalization

Enzymes from a metabolic pathway are not homogeneously distributed within the cell. Instead, they are spatially organized to catalyze sequential reaction steps more efficiently (Srivastava and Bernhard 1986, Davis *et al.* 2015). Colocalization can achieve efficient catalysis by optimizing the substrate exchange step or by providing allosteric regulation (Kuzmak *et al.* 2019, Schupfner *et al.* 2020). Enzymes often colocalize as multienzyme complexes, held together by covalent and non-covalent protein-protein interactions or via scaffolding proteins (*e.g.* cellulosomes in plants) (Davis *et al.* 2015). Protein complexes can either be stable or transient. Enzymes forming transient complexes bind each other with weak affinity through small interface areas (Perkins *et al.* 2010, Schilder and Ubbink 2013, Gad and Ayakar 2021). The formation of such complexes is often prompted by external stimuli, such as ligand binding (Perkins *et al.* 2010). Enzymes that form complexes are often co-expressed (Han *et al.* 2004, Adler *et al.* 2009).

Genetic fusion is another strategy to co-localize enzymes. Fused enzymes are single polypeptide chains consisting of two or more domains, each carrying out a different enzymatic reaction and connected to each other by short or long linkers (Brandon 2004, Zhang 2011). Depending on their amino acid composition and length, linkers can either serve as hinge or spacer between domains, modulating their reciprocal orientation (Papaleo *et al.* 2016). Flexible linkers, containing small residues such as glycine and serine, can facilitate spatial proximity of the domains, thus enabling transient domain-domain interaction (Elleuche 2015). Rigid linkers, rich in prolines, ensure structural stiffness and pre-position domains to favor wanted interactions and avoid unwanted ones (Papaleo *et al.* 2016, Aalbers and Fraaije 2019). Besides proximity, linker regions can provide allosteric communication between individual domains. For example, a small-molecule inhibitor binds to the extracellular domain of the fibroblast growth factor receptor (FGF), allosterically inhibiting intracellular signaling (Gokhale and Khosla 2000, Papaleo *et al.* 2016).

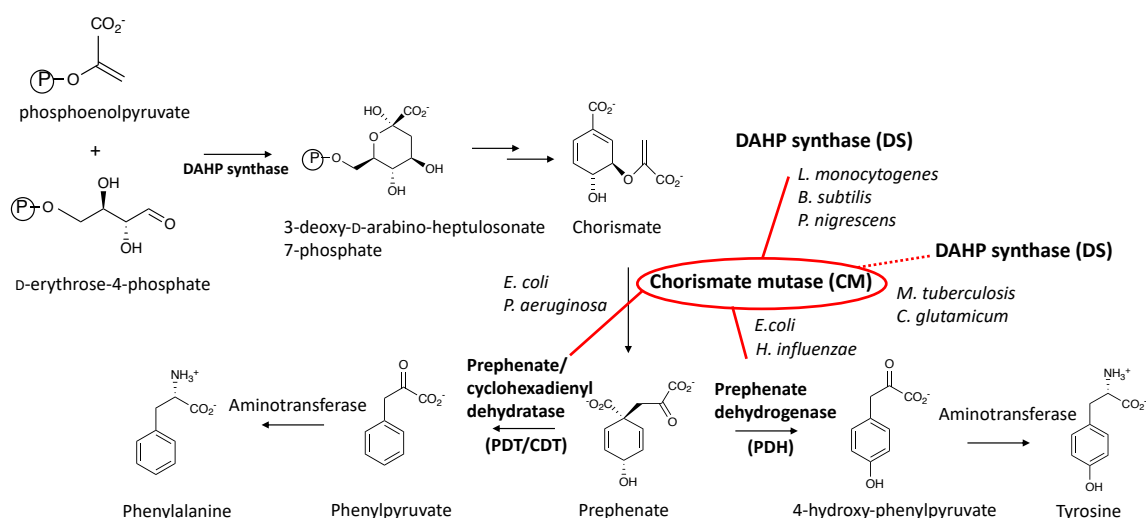


Figure 5. Examples of protein-protein interaction within the biosynthetic pathway for aromatic amino acids. Chorismate mutase (CM) is often found in complex with DAHP synthase (DS) or its subsequent enzymes, prephenate/cyclohexadienyl dehydratase (PDT/CDT) and prephenate dehydrogenase (PDH). For example, non-covalent interaction between CM and DS from *M. tuberculosis* (PDB ID: 2W1A, (Sasso *et al.* 2009)) and *C. glutamicum* (PDB ID: 5HUD (Burschowsky *et al.* 2018)), covalent interactions between CM-DS from *B. subtilis* (Pratap *et al.* 2017) and *Listeria monocytogenes* (PDB ID:3NVT (Light *et al.* 2012)), CMPDH from *E. coli* (Lutke-Eversloh and Stephanopoulos 2005) and *Haemophilus influenzae* (PDB ID: 2PV7, (Chiu *et al.* 2010)) and CMPDT from *E. coli* (Dopheide *et al.* 1972, Gething and Davidson 1976) and *P. aeruginosa* (Fischer *et al.* 1991), *CMCDT from *Aequoribacter fuscus* (**Manuscript I**).

3.2 Substrate exchange mechanisms between colocalized enzymes

Release/diffusion/uptake is the simplest substrate exchange mechanism between two enzymes catalyzing sequential reactions. However, colocalized enzymes can exchange reaction substrates and products by *substrate channeling*, which is the direct transfer of reaction intermediates between their active sites (Miles *et al.* 1999, Zhang 2011). Understanding the substrate exchange mechanism and blocking the transfer of reaction intermediates among metabolic enzymes catalyzing sequential reactions is an effective strategy to inhibit the growth of pathogenic bacteria.

Substrate channeling can occur through either the presence of a physical tunnel between two active sites, or an ‘electrostatic highway’, or by a proximity exchange (Miles *et al.* 1999, Kuzmak *et al.* 2019, Svedružić *et al.* 2020). The first two mechanisms are called *perfect channeling* (Bauler *et al.* 2010, Zhang 2011). The first example of an intramolecular tunnel was discovered in bacterial tryptophan synthase (Miles *et al.* 1999). The enzyme forms an $\alpha_2\beta_2$ tetramer, where the active sites of each α and β subunits are connected by long hydrophobic tunnel to accommodate indole, the reaction intermediate (Hyde *et al.* 1988)

(Figure 6 A). The presence of an ‘electrostatic highway’ facilitates the transfer (‘crawling’) of charged intermediates along the surface between active sites, as seen in the structure of the dihydrofolate reductase-thymidylate synthase complex (Miles *et al.* 1999, Bauler *et al.* 2010, Zhang 2011) (Figure 6B). The proximity channeling requires structural complementarity between sequential enzymes, which get spatially close and provide a shortened route for reaction intermediates to diffuse between their active sites (Bauler *et al.* 2010, Kuzmak *et al.* 2019). In rat peroxisomal multifunctional enzyme type 1 (rpMFE1), a hinge region near domain A enables close proximity with domain C, bringing the two domains closer and allowing proximity channeling between their active site (Kasaragod *et al.* 2010). Moreover the path between the two active sites is lined by positively charged residues, thus enabling substrate crawling, too (Kasaragod *et al.* 2010) (Figure 6 C).

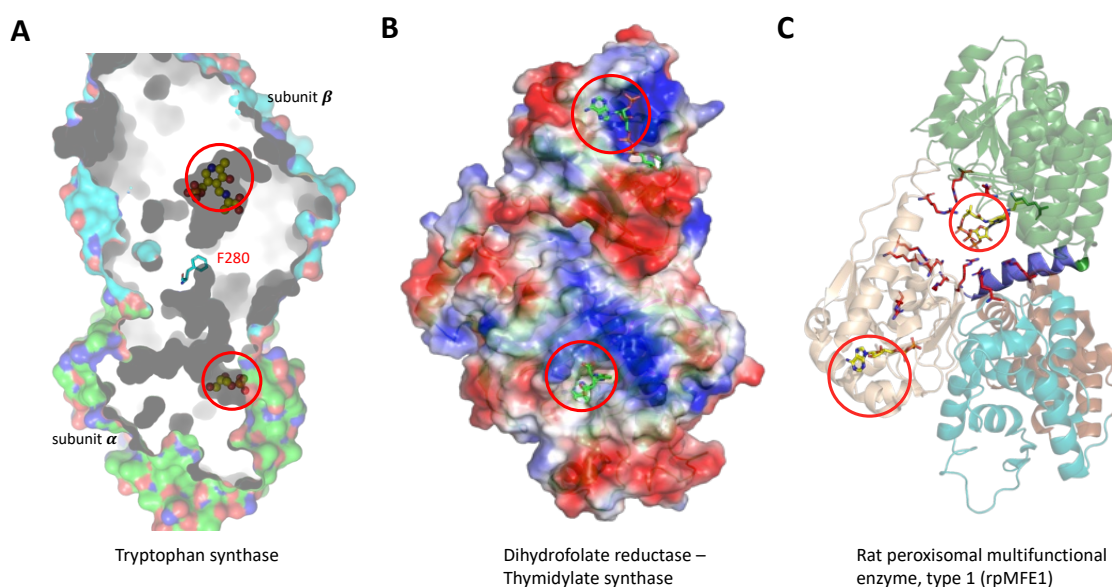


Figure 6. Different substrate-channeling mechanisms. **A.** Physical tunnel between subunit α and β in tryptophan synthase (PDB ID: 2J9X, (Ngo *et al.* 2007)) where Phe280 controls the open and closed conformation of the tunnel. **B.** Electrostatic highway between the active sites of bifunctional dihydrofolate reductase-thymidylate-synthase (PDB ID: 1SEJ, (Svedružić *et al.* 2020)). **C.** Rat peroxisomal multifunctional enzymes, type 1 (rpMFE1). Linker helix (blue) brings domain A (green), domain C (cyan) to proximity (PDB ID: 2X58, (Kasaragod *et al.* 2010)). The active sites involved in direct substrate exchange are highlighted by red circles.

A widely known advantage of substrate-channeling is to accelerate the overall reaction rate by preventing reaction intermediates from diffusing away (Miles *et al.* 1999, Davis *et al.* 2015). However, some studies suggest that this view is true only for reactions for which

diffusion is the rate-limiting step (Zhang *et al.* 2016, Kuzmak *et al.* 2019). In nature, diffusion happens relatively faster than the average turn-over time of typical enzymes (Sweetlove and Fernie 2018). Besides speeding up the reaction rate, substrate-channeling protects intermediates that would otherwise be toxic, unstable, or would diffuse out of the cell (Zhang 2011).

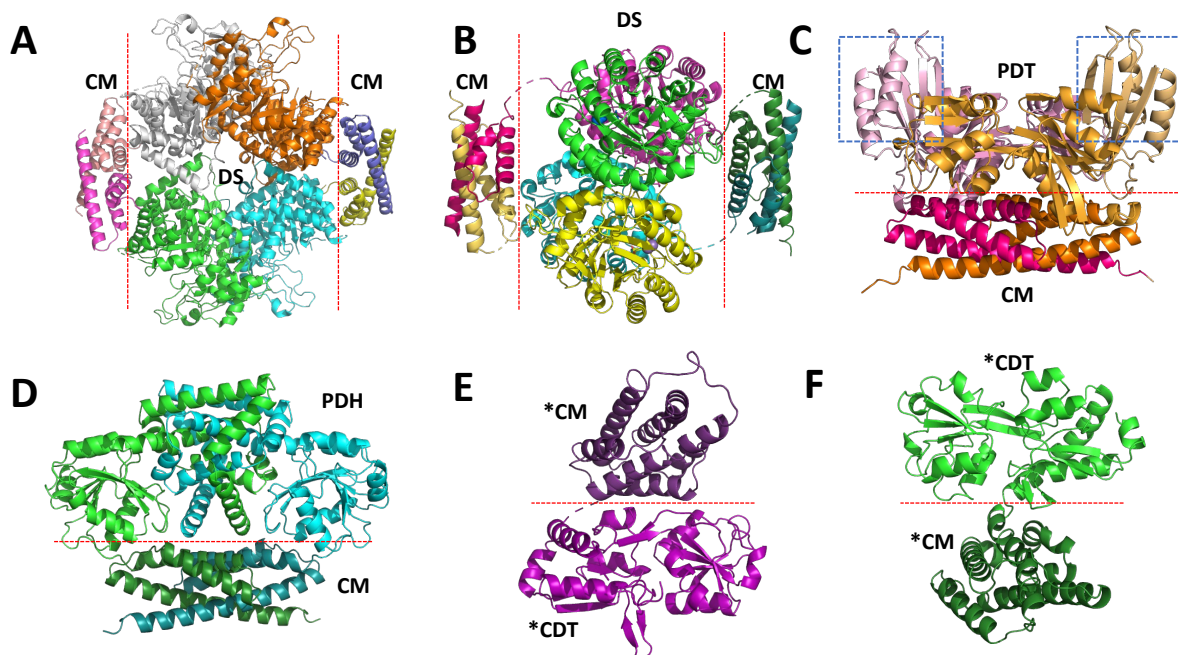


Figure 7. Naturally occurring CM complexes. **A.** Non-covalent hetero-octameric complex of monofunctional CM and DS from *M. tuberculosis* (PDB ID: 2W19, (Sasso *et al.* 2009)) **B.** Non-covalent hetero-octameric complex of bifunctional CM-DS from *Listeria monocytogenes* (PDB ID: 3NVT, (Light *et al.* 2012)). **C.** Biological dimer of the bifunctional CMPDT reconstructed from the AlphaFold2 model of *E. coli* CMPDT (Uniprot: AF-C7EXK8-F1) and crystal structure of *EcCM* (PDB ID: 1ECM, (Lee *et al.* 1995)). The C-terminal regulatory domain is highlighted by dash box in blue. **D.** Bifunctional CM-PDH constructed by alpha-fold structure of *Acidobacteria bacterium* CM-PDH (Uniprot: AF-A0A399YHE9-F1) and crystal structure of *Haemophilus influenzae* bifunctional CM-PDH (PDB ID: 2PV7, (Chiu *et al.* 2010)). **E.** Exported bifunctional *CMCDT from *Aequoribacter fuscus* (**Manuscript I**) **F.** Exported bifunctional *CDTTCM from *Janthinobacterium sp. HH01* (**Manuscript I**). The domain-domain interface is depicted as red, dash line in A-F. Red, dash lines indicate the protein-protein interface.

Another effect of substrate channeling is to direct the metabolic flux towards a specific route. At the branch point of metabolic pathways, enzymes often associate with a specific molecular partner to bypass competing reactions, thus directing the pathway to the production of a certain metabolite. Several examples of that can be found within the shikimate pathway. *E. coli* encodes two cytoplasmic CMs fused to PDH and PDT (Figure 7 C, D): the first steers the pathway to the biosynthesis of L-Tyr and the latter to the biosynthesis of L-Phe (Dopheide *et al.* 1972, Gething and Davidson 1976, Lutke-Eversloh and Stephanopoulos 2005) The PDT

domain in bifunctional *E. coli* CM-PDT has a C-terminal regulatory domain for L-Phe binding, which inhibits both PDT and CM by 80% and 60%, respectively (Gething and Davidson 1976). This is also an example of how colocalization can provide allosteric regulation as described later in details in Chapter 4.

3.3 Model system: bifunctional extracytoplasmic *CDTCM and *CMCDT enzymes

The group of Prof. Peter Kast at the EHT, Zurich, has discovered several exported bifunctional bacterial CMs, either N-terminally or C-terminally fused to CDT (**Manuscript I**). These bifunctional enzymes are extremely rare in nature, found in only nine non-pathogenic bacteria species so far. The *CMCDTs have been found exclusively in γ -proteobacteria and the domain-swapped *CDTCMs in β -proteobacteria (**Manuscript I**). The CM domains of the bifunctional enzymes belong to the AroQ γ CM subclass and share a sequence identity of 15-20 % with the secreted *MtCM, the holotype for the AroQ γ subclass (**Manuscript I**). The linker sequence of *CDTCM is highly conserved throughout orthologs, and it contains a Trp-Leu-Xaa3-Xaa4-Xaa5-Trp (WLXXXW) motif (Figure 8) (**Manuscript I**). Except for *AfCMCDT, all *CMCDTs have a very short linker region consisting of two amino acids, where the second one is Tyr (**Manuscript I**). The *AfCMCDT contains a five-residue-long PNLDP linker (**Manuscript I**).



Figure 8. Linker sequence in the bifunctional enzymes. **A.** *CDTCMs, carrying a well conserved WLXXXW sequence. **B.** *CMCDTs. *AfCMCDT has a longer sequence (PNLDP), while other orthologs have only two residue - long linker, where the second residue is Tyr.

Exported CMs are often correlated to the pathogenicity of their producer organisms (Calhoun *et al.* 2001, Okvist *et al.* 2006, Assis *et al.* 2021). The existence of these bifunctional CMs in non-pathogenic bacteria raises the question of their biological role. In addition to that, the evolutionary reason for having to form a fusion product with CDT, rather than a complex, is

worth further investigation. As stated before, protein-protein interaction through a covalent linker may affect the catalytic activity by allosteric regulation or substrate channeling. Therefore, understanding the interdomain communication would be helpful in defining a way to disturb any beneficial relationship between the two and, as a consequence, inhibit the growth of their host.

In **Manuscript I** we investigated the interplay between the CM and CDT domains in exported bifunctional *CDTCM and *CMCDT enzymes from non-pathogenic bacteria, by kinetics analysis and by determining their crystal and solution structures. Their biological role was inferred by genomic neighborhood analysis.

3.3.1 *Manuscript I, Summary*

Structural evidence for tunneling (or lack thereof)

To assess the relative arrangement of the two domains in *CDTCM and *CMCDT bifunctional enzymes and the possible occurrence of substrate channeling, we solved the X-ray crystallographic structure of selected representatives from both classes (*JbCDTCM, from *Janthinobacterium sp.* HH01, *DsCDTCM, from *Duganella sacchari* and *AfCMCDT, from *Aequoribacter fuscus*). The CM domains adopt the typical AroQ₇-subclass fold seen in the prototypical *MtCM, having six α -helices (Figure 4 E). The CDT domains exhibit a ‘periplasmic solute-binding protein-like’ fold, highly similar to that of the periplasmic *Pseudomonas aeruginosa* CDT (*PaeCDT, r.m.s.d. values of 0.6-1.0 Å, Figure 9 A-C)

The active sites of the two domains are located 35-40 Å apart from each other, but neither physical tunnel nor electrostatic highway could be observed between the two active sites (Figure 8 A, B). *JbCDTCM and *AfCMCDT crystallized as monomers in the asymmetric unit, and no quaternary structure was observed in the crystal lattice. The asymmetric unit of *DsCDTCM crystals contains four chains: two of them are arranged as a head-to-tail dimer, and thus have the potential to form intermolecular channeling (Figure 9 F). However, the active sites opening of the two domains point to opposite directions, making substrate channeling by proximity exchange unlikely (Figure 9 F). Overall, structural analysis of the fusion enzymes did not support the hypothesis of substrate channeling.

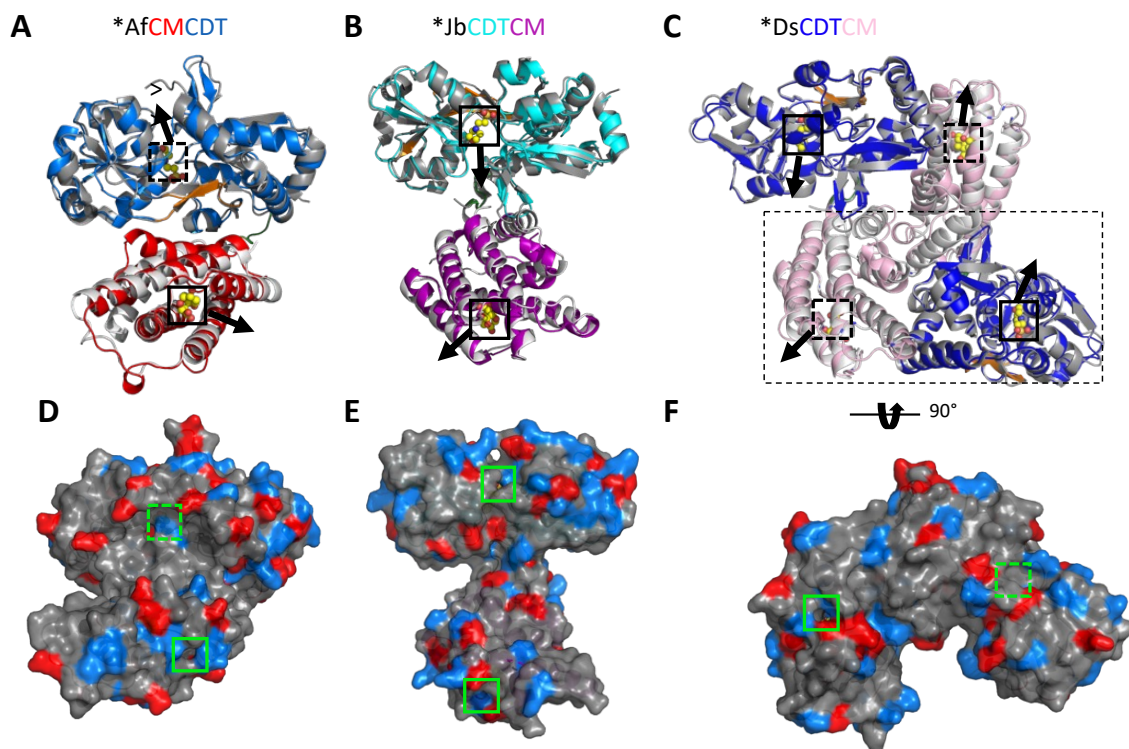


Figure 9. The crystal structures of the bifunctional enzymes A-C. Cartoon presentations of *AfCMCDT (A), *JbCDTCM (B), *DsCDTCM (C). The three structures are superposed matching their CM domain on *MtCM bound to TSA, (PDB ID: 2FP2, Chain B, (Okvist *et al.* 2006); white grey) and *PaeCDT (PDB ID: 5HPQ, (Clifton *et al.* 2018); dark grey). Hinges between CDT subdomains are colored orange, and the expected exit trajectories from the active sites are indicated with black arrows. **D-F.** Surface presentation by charged residues. *AfCMCDT (D), *JbCDTCM (E), *DsCDTCM (F). Positively charged residues are colored in blue; negatively charged residues are colored in red. Note that the domains in the dashed rectangular box are the CM and CDT domains from the two different chains shown in (C), but rotated by 90° in panel F. The active sites are indicated with small squares; those facing backwards drawn with dashed lines. The CM active sites are illustrated with the *endo*-oxabicyclic transition state analog (TSA) bound in the *MtCM structure (PDB ID: 2FP2, Chain B, ((Okvist *et al.* 2006))), and the CDT active sites with a 2-(*N*-morpholino)ethanesulfonic acid (MES) molecule that was co-crystallized with *JbCDTCM.

Table 1. overview on fusion enzyme studies in manuscript 1					
CDTCM bifunctional enzymes produced, tested for activity, and crystallized					
Organisms	CDTCM	CDTCM	CDTCM	Split domains	
				CDT	CM
<i>Duganella sacchari</i>	● ● ●				
<i>Janthinobacterium sp. HH01</i>	● ● ●	● ●	● ●	(Not soluble)	● ●
<i>Massila phosphatilytica</i>	● ●				
CMCDT bifunctional enzymes produced, tested for activity, and crystallized					
Organisms	CMCDT	CMCDT	CMCDT	Split domains	
				CM	CDT
<i>Aequoribacter fuscus</i>	● ● ●	● ●	● ●	● (Low yield)	● ●
<i>Steroidobacter cummioxidans</i>	● ●				
<i>Thalassomonas actiniarum</i>	(Not soluble)				
<i>Thalassomonas viridans</i>					
<i>Shewanella baltica</i>	● ●				
<i>Shewanella psychrophila</i>					

- expressed
- activity tested
- structure solved

Uncertain catalytic benefits of domain fusion

The possible catalytic benefit of domain fusion on the catalytic activity of the CM and CDT domains was investigated by kinetic analysis. To draw more general conclusions on exported CDT/CM fusion enzymes, other orthologs was included (*MpCDTCM, from *Massila phosphatilytica*, *ScCMCDT, from *Steroidobacter cummioxidans*, *SbCMCDT, from *Shewanella baltica* and *TaCMCDT, from *Thalassomonas actiniarum*; see Table 1).

Three sets of kinetic assays were performed to evaluate the effect of fusion on the catalysis of bifunctional enzymes. First, the kinetic parameters (K_m and k_{cat}) of all wild-type fusion enzymes were investigated. In this experiment, kinetic parameters of individual catalytic reactions (CM or CDT) and coupled CM+CDT reaction were determined. The coupled CDT+CM experiment measures two consecutive conversion of chorismate to phenylpyruvate via prephenate. If no substrate-channel is present, a lower catalytic rate step would be a rate-limiting step for overall conversion. Second, split domains (CDT and CM) of *JbCDTCM and *AfCMCDT were produced, and their kinetic parameters were determined independently from the conjugate enzyme. If different from the full-length enzyme, it would hint to an effect of fusion on catalytic efficiency. Third, active site knockout monofunctional variants (*~~CDTCM~~ and *~~CDTCM~~) of *JbCDTCM and *AfCMCDT were produced, and kinetic parameters of the active domain were determined. By comparing the results of the split

domain kinetics to those of the domain knock-outs, it would allow to evaluate the presence of an allosteric effect between the two domains.

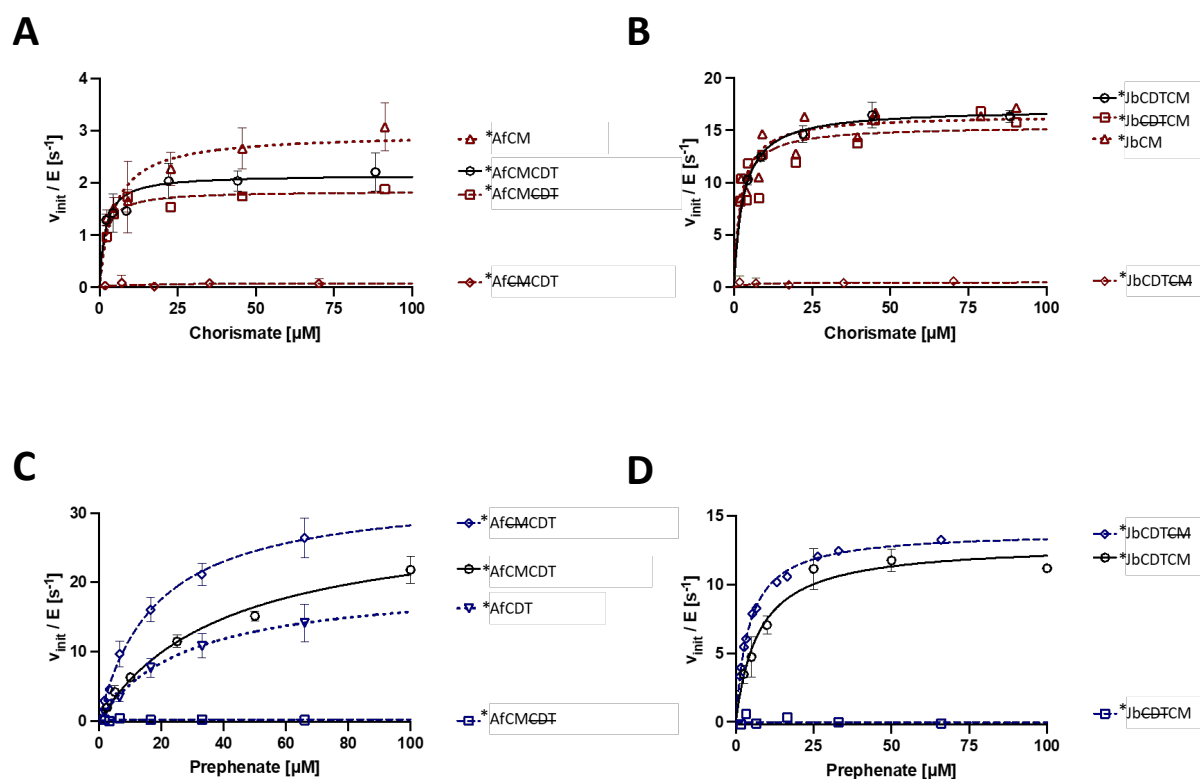


Figure 10. Michaelis-Menten plots of active site KO and single split domain variants of *AfCMCDT and *JbCDTCM. The CM activities (A) and (B), as well as the CDT activity (C) and (D) for the variants *AfCMCDT and *JbCDTCM, respectively, are displayed. Two independently prepared biological replicates were measured of each variant and the curve of the mean calculated. Standard deviation bars are given for identical substrate concentrations in the measurement of the two replicates. Figure adapted from **Manuscript I**, Supplementary information, prepared by Christian Stocker.

Overall, no catalytic benefit of fusion was detected, excluding potential substrate-channeling or interdomain allosteric enhancement of the catalytic activity. Interestingly, CM domains of the bifunctional enzymes exhibit an extremely low $K_m < 7 \mu\text{M}$ compared to its monofunctional ortholog *MtCM ($180 \mu\text{M}$), thus are able to bind very low concentration of chorismate. The low values of K_m of the CM domain of fusion enzymes were accompanied by lower k_{cat} values, resulting into an order of magnitude higher catalytic efficiency (k_{cat}/K_m) of the bifunctional CMs, $1\text{-}4 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ for the *CMCDTs and $5\text{-}8 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ for the *CDTCMs, than *MtCM ($2.7 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$).

The fusion CM/CDT enzymes are probably involved in signaling, not biosynthesis

No feedback response was detected in kinetic assays by adding end-products to the reaction environment (Phe and Tyr; see Figure 11). This suggests that the biosynthesis of aromatic amino acids might not be the main biological role of the *CMCDT and *CDTCM fusion enzymes. Furthermore, their existence in non-pathogenic bacteria weakens the hypothesis of their potential role in pathogenicity, observed for exported monofunctional chorismate mutases (Calhoun *et al.* 2001, Okvist *et al.* 2006).

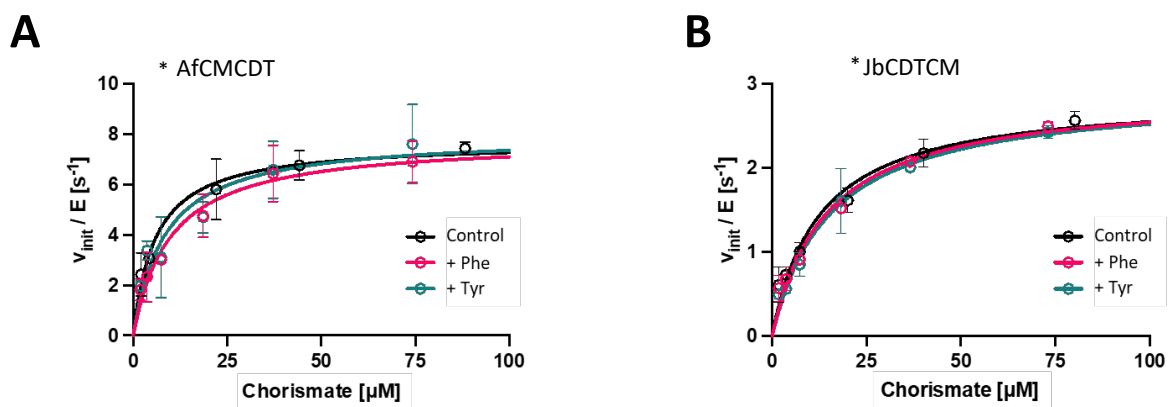


Figure 11. Testing for feedback regulation of CM or CDT activity by Phe or Tyr. Michaelis-Menten plots of coupled CM+CDT kinetic assays with (A) *AfCMCDT and (B) *JbCDTCM in the presence 200 μM Phe or Tyr (80,000-fold molar excess over the enzyme concentration), in comparison to a control in the absence of Phe or Tyr (wild-type). The bars depict standard deviations of two independent biological replicates. Figure adapted from **Manuscript I**, Supplementary information, prepared by Christian Stocker.

Genomic neighborhood analysis was carried out to shed light on the physiological role of exported bifunctional enzymes, as functionally related genes often are organized in operons and/or gene clusters (Ralston 2008). The analysis shows that bifunctional enzyme genes were found flanked by genes encoding solute-binding and transport proteins having the same orientation. It might not be a coincidence that monofunctional exported CDTs, too, evolved from ancestral solute-binding protein (Clifton *et al.* 2018). Interestingly, a gene encoding AI-2E family protein is located directly upstream of the *DsCDTCM gene with the same orientation. The AI-2E family is a transporter protein family for AutoInducer-2 (AI-2), which is a signaling molecule used in quorum-sensing (Taga and Xavier 2011, Laganenka *et al.* 2016). Taking into account the extremely low K_m values of CM domains of the fusion enzymes, which allows binding of very low concentrations of its substrate, and genomic neighborhood analysis, we hypothesize that the exported bifunctional CM/CDTs might be involved in quorum sensing and chemotaxis.

3.3.2 Further insights from the crystal and solution structures of the bifunctional CM/CDTs

Thermal stability of *JbCDTCM, *DsCDTCM, *MpCDTCM *AfCMCDT, and *ScCMCDT, were investigated using DSF. These enzymes confirmed to be stable, well-folded proteins, with melting temperatures between 40-46°C (Table2).

Table 2. Thermal stability of of CM/CDT bifunctional enzymes

Sample	T _m (°C)
*JbCDTCM	45.0±0.20
*DsCDTCM	39.8±0.20
*MpCDTCM	45.8±0.78
*AfCMCDT	42.9±0.07
*ScCMCDT	46.4±0.20

Solution structures of *ScCMCDT and *JbCDTCM were studied using small-angle X-ray scattering. No atomistic structural data is available on *ScCMCDT; *AfCMCDT, the only ortholog with a known structure, is an ill suited homology model, as the *CMCDT linker region of the two enzymes significantly differ in length and composition. The amino acid composition of the linker is a key feature in determining the relative orientation of the two domains (Aalbers and Fraaije 2019). A *ScCMCDT model was predicted by AlphaFold2-based ColabFold online tool (Mirdita *et al.* 2022), Indeed, the *ScCMCDT model shows a different relative domain arrangement compared to *AfCMCDT (Figure 12 C).

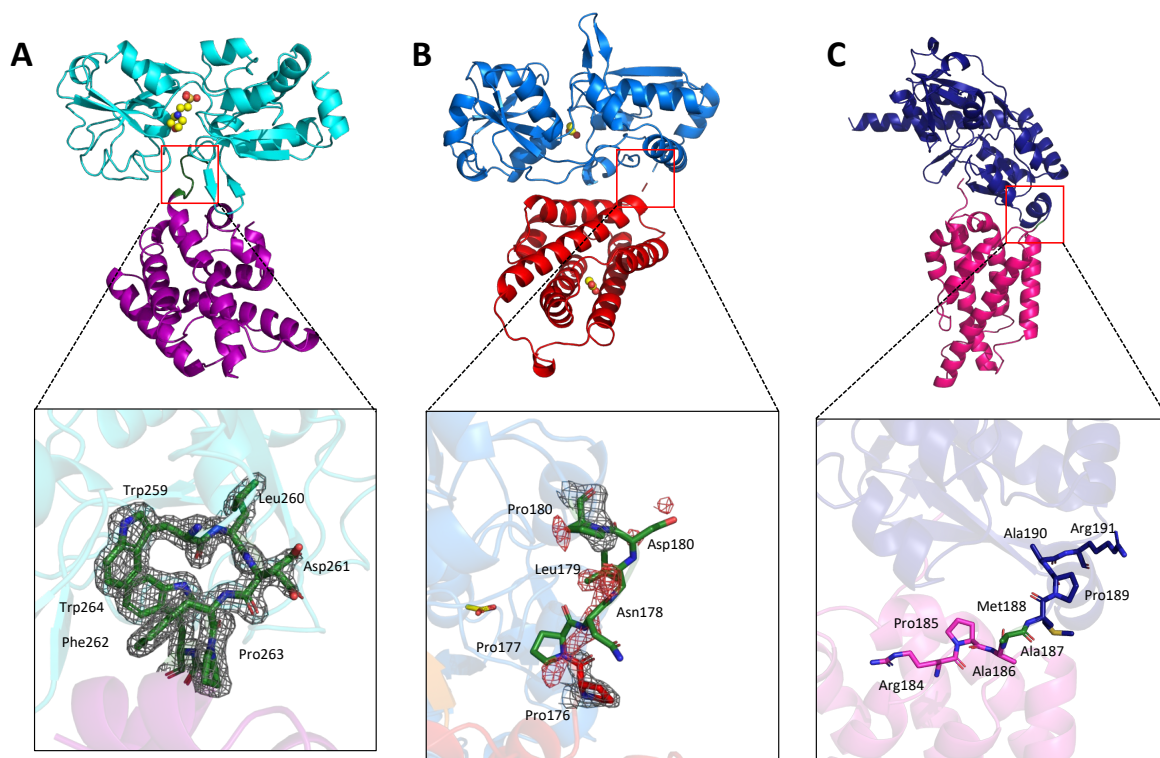


Figure 12. Domain-domain arrangement and linker region in the bifunctional enzymes. **A.** *JbCDTCM crystal structure. The linker region is six-residue and folds into a helical conformation. **B.** *AfCMCDT crystal structure. The linker region is five-residue long, and it showed patchy electron density, suggesting conformational flexibility. **C.** *ScCMCDT, model predicted by AlphaFold2 (Mirdita *et al.* 2022). The *ScCMCDT linker consists of a single residue, Ala287. Linker regions in three bifunctional enzymes are shown in green sticks.

A good fit between the experimental and theoretical scattering curves (χ^2 of 1.0), calculated using *Pepsi-SAXS* from the *JbCDTCM structure and the *ScCMCDT *in silico* model (Grudin *et al.* 2017), suggests that the two proteins are monomeric in solution (Figure 13 A, D). Protein folding was assessed by normalized Kratky plot and the $P(r)$ function. SAXS data suggest that *JbCDTCM is a well-folded, globular protein (Figure 13 B). The $P(r)$ function of the *JbCDTCM shows a main peak flanked by an overlapping shoulder peak, which is the signature for a modular protein with short linker (Receveur-Br  chot and Durand 2012) (Figure 13 C). In contrast, SAXS data suggests that *ScCMCDT adopts an elongated conformation, having the maximum value at $qR_g > \sqrt{3}$ in normalized Kratky plot and extended tail in $P(r)$ function (Figure 13 E, F). This is in good agreement with the *ScCMCDT AlphaFold2 model, showing the two domains arranged as a prolate ensemble (Figure 12 C).

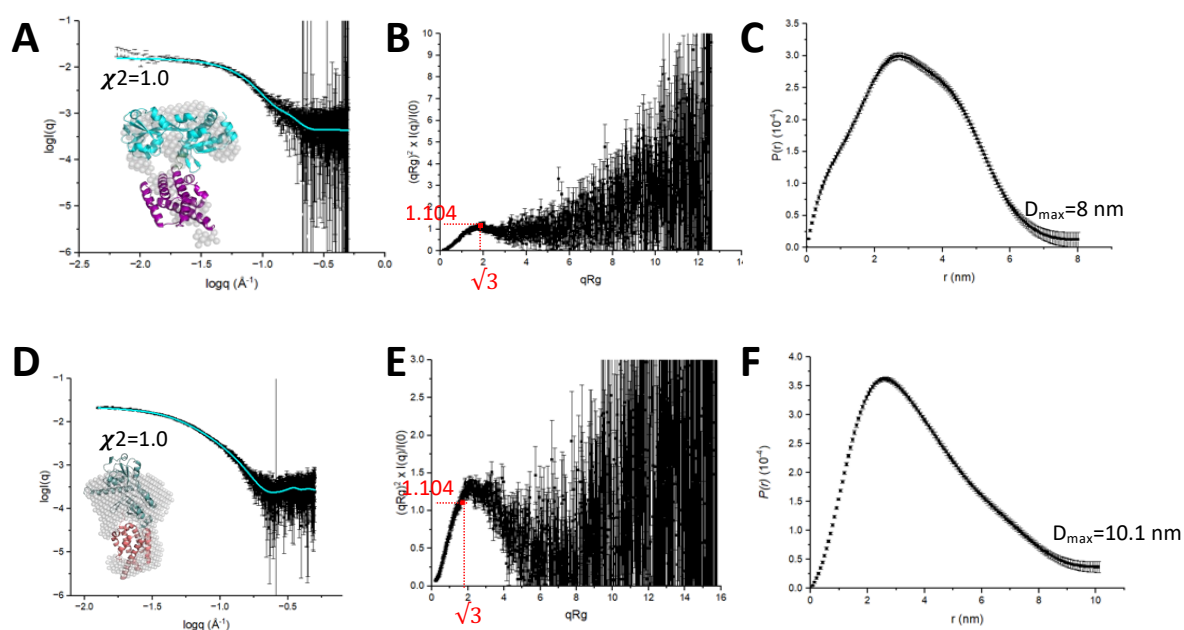


Figure 13. SAXS data for the *JbCDTCM and *ScCMCDT. **A.** Fit between experimental scattering curve (black) and theoretical (cyan) scattering curve of the *JbCDTCM calculated using *Pepsi-SAXS* from its crystal structure, with χ^2 value of 1.0. **B.** Normalized Kratky plot for *JbCDTCM, having maximum value of 1.104 for $qR_g = \sqrt{3}$. **C.** $P(r)$ function for *JbCDTCM, revealing a shoulder peak. **D.** Fit between experimental scattering curve (black) and theoretical (cyan) scattering curve of the *ScCMCDT calculated using *Pepsi-SAXS* from its AlphaFold2 model with χ^2 value of 1.0. **E.** Normalized Kratky plot for *ScCMCDT, having a peak maximum for $qR_g > \sqrt{3}$. **F.** $P(r)$ function for *ScCMCDT, showing an extended tail.

In addition to the crystal structures of the bifunctional enzymes included in **Manuscript I** (MES-bound *JbCDTCM, *DsCDTCM, acetate-bound *AfCMCDT), two crystal structures of the *JbCDTCM, in complex with either PPY alone and in complex with both PPY and TSA, and the TSA-bound *AfCMCDT were solved and refined to 1.8 Å, 1.9 Å, and 2.4 Å, respectively (Table 3 and Table 15 in Methods) (Figure 14). No quaternary structures were observed in the crystal lattice of either JbCDTCM structures, supporting the hypothesis that the fusion enzymes are monomeric.

Table 3 – Structure of CM/CDT bifunctional enzymes

Enzyme	Structure (see Methods)	CM active site	CDT active site	Resolution (Å)
*JbCDTCM	JbCDTCM_MES (Manuscript I)	-	MES	1.7
	JbCDTCM_PPY	-	PPY	1.8
	JbCDTCM_PPY_TSA	TSA	PPY	1.9
*DsCDTCM	DsCDTCM (Manuscript I)	-	-	2.5
*AfCMCDT	AfCMTD_ACT (Manuscript I)	acetate	acetate	1.6
	AfCMTD_TSA	TSA	-	2.4
*ScCMCDT	Alphafold2 model	-	-	-

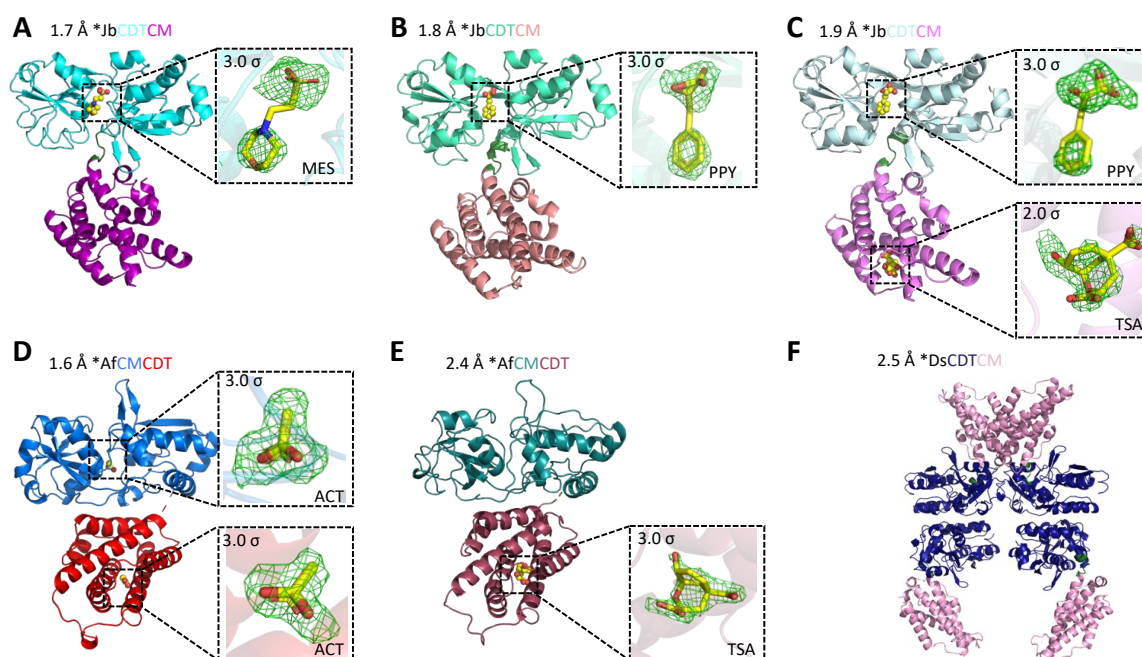


Figure 14. Asymmetric unit content of crystals of the bifunctional enzymes in complex with various ligands. A. MES-bound *JbCDTCM, solved to 1.8 Å. B. PPY-bound *JbCDTCM, solved to 1.8 Å. C. PPY- and TSA-bound *JbCDTCM, solved to 1.9 Å. D. Acetate-bound *MgCMCDT, solved to 1.8 Å. E. TSA-bound *MgCMCDT, solved to 2.4 Å. F. *DsCDTCM, solved to 2.5 Å. σ A-weighted mF_o-DF_c maps for the ligands are depicted at contour level at 3 σ , except for TSA bound to *JbCDTCM (2 σ .)

3.4 Model system: extracytoplasmic aromatic amino acid biosynthesis pathway enzymes from *Pseudomonas aeruginosa*

The *P. aeruginosa* genome encodes a cytoplasmic CMPDT bifunctional enzyme (PaeCMPDT, referred to as ‘P-protein’ in some publications), which catalyzes the conversion of chorismate to phenylpyruvate (Fischer *et al.* 1991, Zhao *et al.* 1993). In addition, the organism exports to the periplasmic space an AroQ γ -subclass monofunctional chorismate mutase, as active as its cytoplasmic counterpart (Calhoun *et al.* 2001). Remarkably, the periplasmic space of *P. aeruginosa* also hosts a cyclohexadienyl dehydratase, which carries prephenate dehydratase activity, and several aminotransferases, thus having all the components for the post-chorismate mutase biosynthesis of aromatic amino acids (Fiske *et al.* 1983, Zhao *et al.* 1992). Two independent studies on the subcellular localization of *PaeCM and *PaeCDT revealed that both enzymes are located at the polar regions of the periplasmic space (Zhao *et al.* 1993, Calhoun *et al.* 2001). Enzyme-enzyme interaction is a common

feature between enzymes catalyzing sequential reactions (Cascante *et al.* 1994); moreover, the occurrence of periplasmic *CMCDT and *CDTCM fusion enzymes suggests the possibility that the two enzymes could form a functional complex.

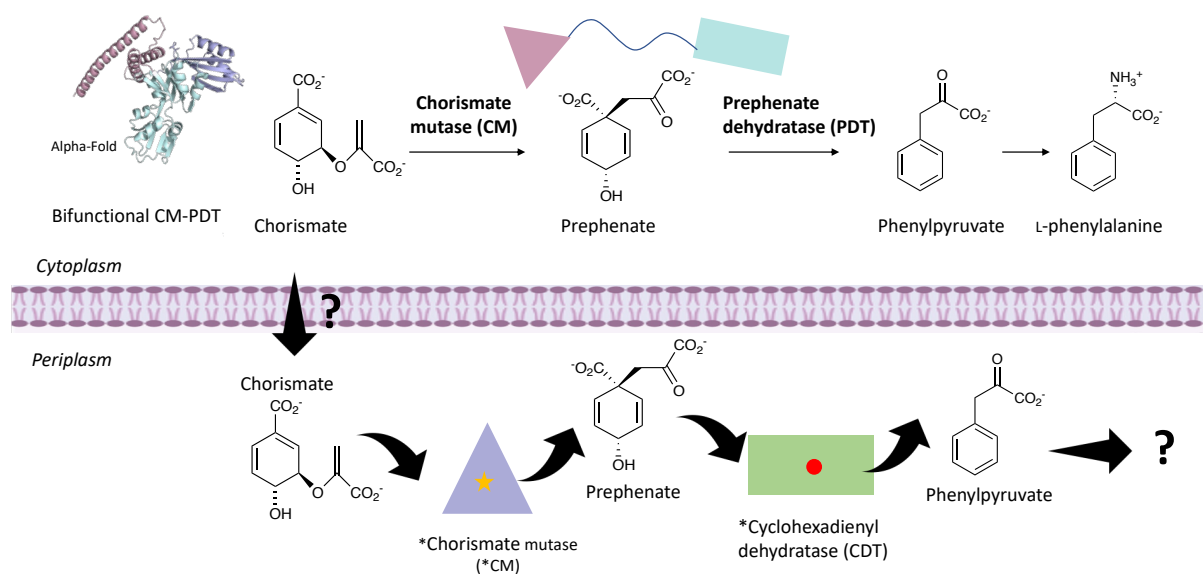


Figure 15. Hypothetical spillover pathway for the synthesis of L-Phe in *P. aeruginosa*. *P. aeruginosa* encodes cytoplasmic CM fused to PDT on its C-terminus (CM: pink triangle, CDT: cyan rectangular), and periplasmic monofunctional CM (purple triangle). In addition to CM, the organism also possesses periplasmic CDT (green triangle). As CM and CDT catalyze sequential reaction (conversion of chorismate to phenylpyruvate via prephenate), Jensen and coworker hypothesized a ‘spillover’ periplasmic pathway for the biosynthesis of L-Phe (Zhao *et al.* 1993, Calhoun *et al.* 2001).

In **Manuscript II**, we collected structural data on *PaeCM and *PaeCDT by X-ray crystallography and small-angle X-ray scattering and probed their potential interaction by size-exclusion chromatography.

3.4.1 Manuscript II, Summary

The crystal structure of *PaeCM was solved at 1.2 Å, with two subunits per asymmetric unit; SAXS data validated the dimer as the biological unit. Each protomer displays a typical AroQ₇-class CM fold, which consists of six α-helices folded into a pseudo-symmetric dimer of three α-helices (Figure 4 E) (Okvist *et al.* 2006). *PaeCDT crystallized as a hexamer (dimer of two trimers) and the structure was determined to 2.2 Å (Figure 16 A). The CDT domain displays a ‘periplasmic solute-binding protein-like’ fold, having two α/β subdomains (‘large’ and ‘small’), connected by a two-β-strands hinge. MD simulations performed by

Kaczmariski *et al.* on the *PaeCDT monomer revealed that the two β -strands mediate the opening and closing of the active site cleft (Kaczmariski *et al.* 2020) (Figure 16 C).

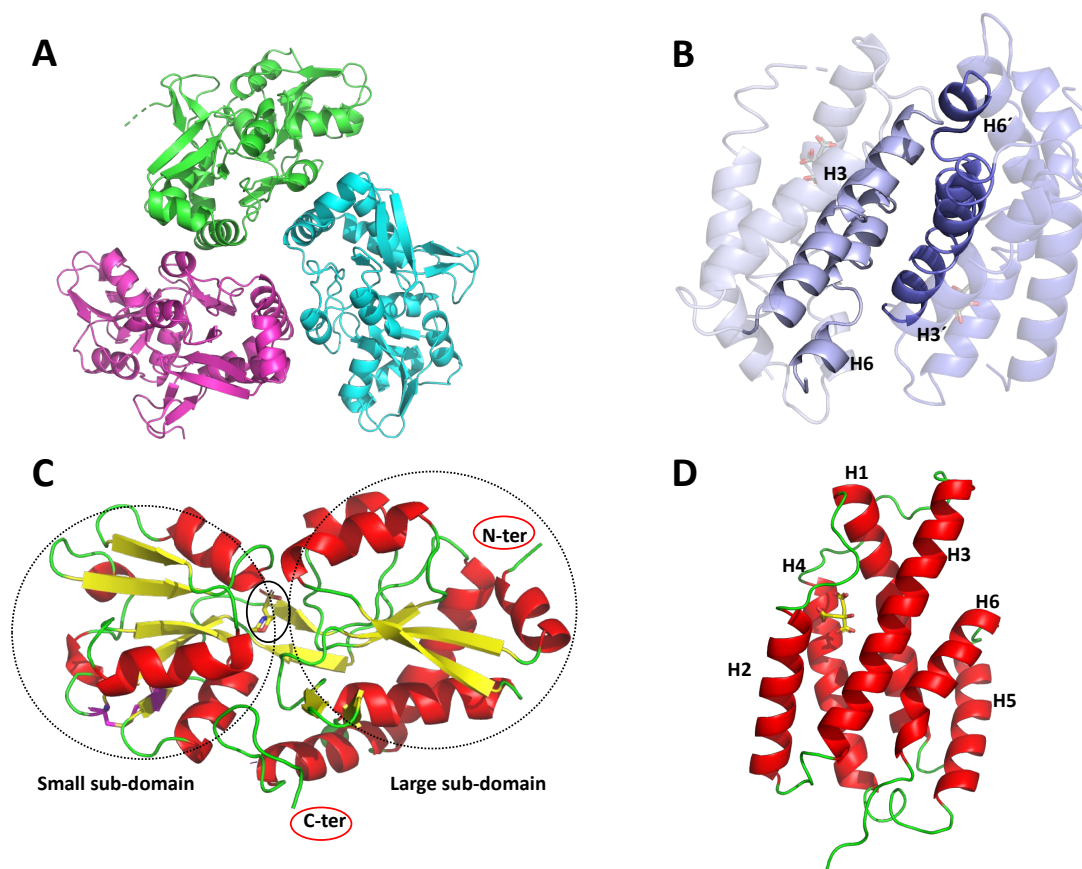


Figure 16. Structural information on *PaeCDT and *PaeCM. A. *PaeCDT trimer. B. *PaeCM dimer. C. Periplasmic solute-binding protein-like domain fold, *PaeCDT. D. AroQ γ subclass CM fold, *PaeCM.

A comparison of the citrate-bound *PaeCM with the TSA-bound *MtCM, its ortholog, revealed a significantly wider distance (4.1 Å) between the N-termini of H2 and H4, the two α -helices defining the opening to the catalytic cleft (Figure 17 A). Normal mode analysis (NMA) on *PaeCM displayed a large H2/H4 scissoring motion and conformational flexibility of the H1-H2 and H3-H4 loops (Figure 17 B). This is in good agreement with a previous observation of ‘open’ and ‘closed’ conformation for the AroQ γ subclass CM active site (Okvist *et al.* 2006). Based on our structural and computational data, we suggest the presence of a CM active site substrate gating mechanism coupled to the enzyme breathing motion.

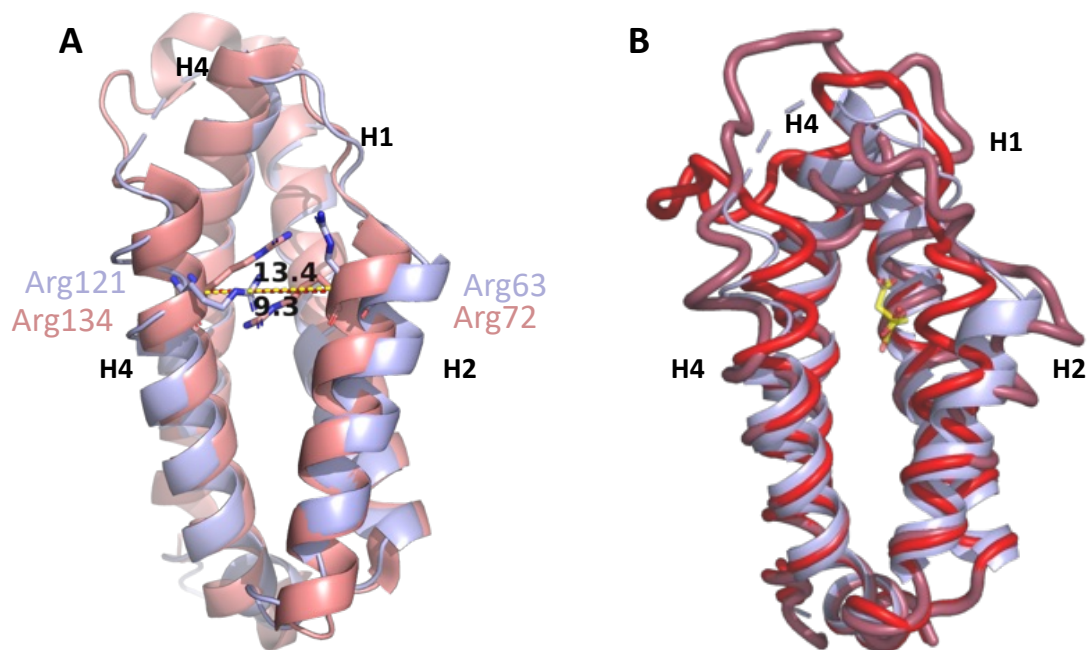


Figure 17. Substrate gating mechanism of AroQ₇ subclass CM. **A.** Distance measured between the N-termini of α -helices H2 and H4 in the citrate-bound *PaeCM structure, Chain A (Manuscript II; light blue) the TSA-bound *MtCM structure (PDB ID: 2FP2, Chain B, (Okvist *et al.* 2006); salmon). Wider active site opening was observed for TSA-bound *MtCM compared to citrate-bound *PaeCM. **B.** First and last frame of one of the lowest energy modes generated by NMA (red and raspberry, respectively) compared to *PaeCM, Chain A (Manuscript II; light blue), showing a displacement of H2 and H4.

The SAXS data confirms that the quaternary structures of *PaeCDT (trimer) and *PaeCM (dimer) captured in crystals are stable in solution. Normalized Kratky plot and $P(r)$ function suggest that *PaeCM is a well-folded globular protein (Figure 18 B, C). The crystal structure revealed that *PaeCDT trimer is formed by the interaction between the ‘large’ domains of the three protomers, while the ‘small’ domains are exposed to the solvent (Figure 16 A). This is reflected by the normalized Kratky plot and $P(r)$ function, showing features compatible with a modular protein with short linker (Figure 18 E, F).

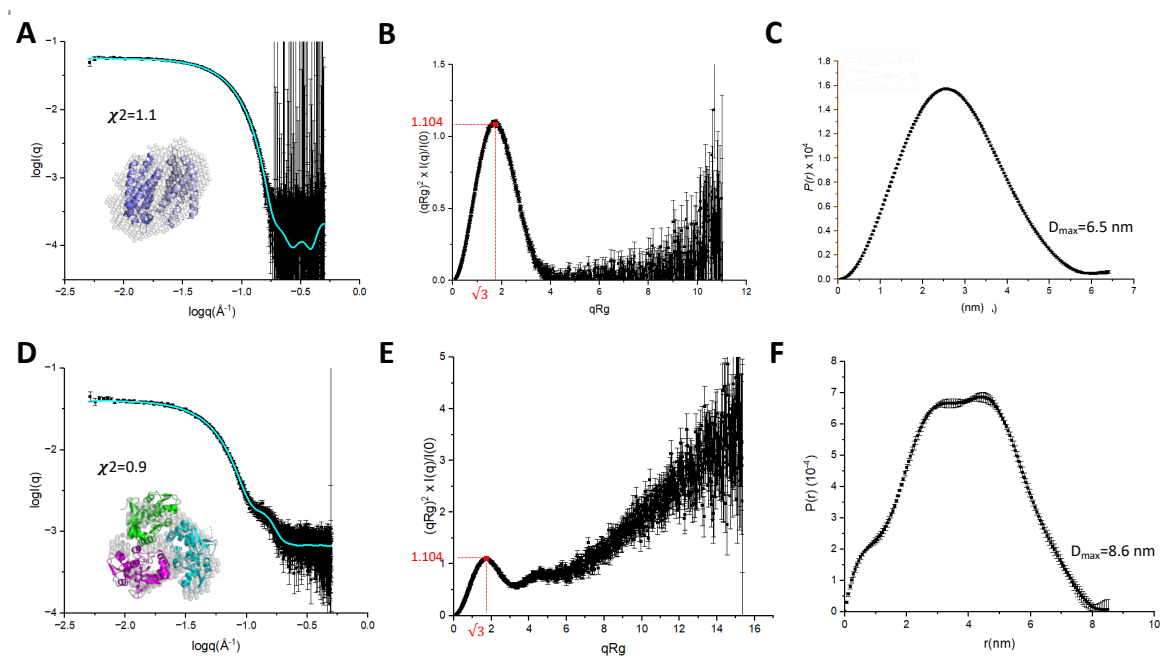


Figure 18. SAXS data for *PaeCM and *PaeCDT. **A.** Fit between experimental scattering curve (black) and theoretical (cyan) scattering curve of the *PaeCM calculated using *Pepsi-SAXS* from its dimer, with χ^2 value of 1.1. **B.** Normalized Kratky plot for *PaeCM, having maximum value of 1.104 for $qR_g = \sqrt{3}$. **C.** $P(r)$ function for *PaeCM, with a bell-shape **D.** Fit between experimental scattering curve (black) and theoretical (cyan) scattering curve of the *PaeCDT calculated using *Pepsi-SAXS* from its trimer with χ^2 value of 0.9. **E.** Normalized Kratky plot for *PaeCDT, having a peak maximum, 1.104 for $qR_g > \sqrt{3}$. **F.** $P(r)$ function for *PaeCDT, showing a shoulder peak.

We carried out the interaction study using size-exclusion chromatography on the two proteins mixed in equimolar ratio (Figure 19). The elution profiles were compared to the individual protein peaks, revealing the absence of peaks at lower retention volume compared to the single-enzymes, imply the absence of a complex between the two.

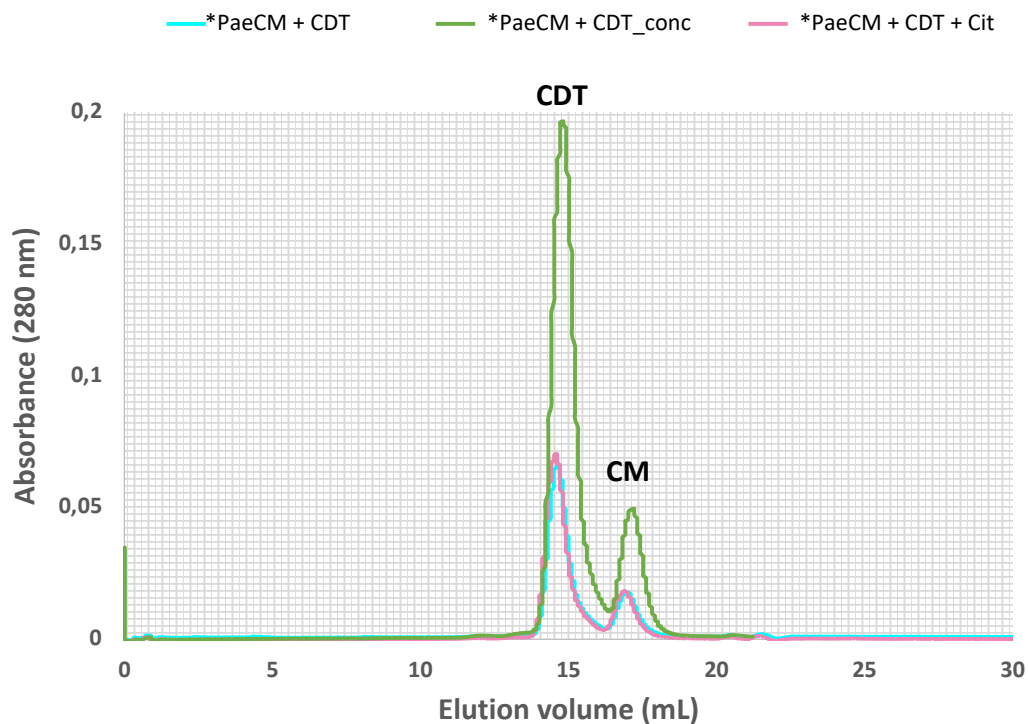


Figure 19. SEC analysis of *PaeCM, *PaeCDT and their equimolar mixture. A. Superposed chromatograms of equimolar mixtures of the *PaeCM and *PaeCDT without ligands (lower concentration mixture - cyan, higher concentration mixture - green) and with citrate added (pink).

Overall, SAXS and crystallographic data provide information on the multimerization state of *PaeCM and *PaeCDT, and suggest that the two enzymes do not form a stable complex. The lack of a complex formation suggests the absence of substrate channeling between the two, leaving substrate uptake/release as the most likely mechanism. Finally, we gathered insights into the dynamic behavior of *PaeCM and *PaeCDT. NMA on *PaeCM crystallographic data suggests the existence of an active site gating mechanism coupled to the enzyme's breathing motion. SAXS data collected on *PaeCDT confirmed the active opening mechanism predicted by past MD simulations, associating the bump in the scattering plot to the flexibility of the small subdomain (Kaczmariski *et al.* 2020).

3.5 Additional insights into CM and CDT domains

3.5.1 CDT domains

3.5.1.1 Hinge-mediated open and closed conformation of the active site of CDT domain

In contrast to the large number of publications studying CMs, less is known for CDTs. The CDT is a periplasmic protein/domain evolved from ancestral solute-binding proteins. The CDT domain consists of two subdomains, ‘large’ and ‘small’, connected by two β -strands. The active site is located at the cleft between the two (Figure 16 C).

So far, only the *Pseudomonas aeruginosa* CDT structure has been published (Clifton *et al.* 2018). Clifton *et al.*, identified the CDT active site residues by docking prephenate, a natural substrate of CDT, into the *PaeCDT crystal structure (PDB ID: 5HPQ) (Clifton *et al.* 2018) We obtained structures of the CDT domain from the *JbCDTCM bifunctional enzyme in complex with MES and phenylpyruvate (PPY). As phenylpyruvate is a product of CDT, this is the first ever structure of a CDT domain with its natural product. Molecular dynamics simulations run by Kaczmariski *et al.* on *PaeCDT showed that the protein can take either an open or a close active site conformation, hinging on the two β -strands connecting the small and large subdomains (hence referred to as ‘hinge region’) (Kaczmariski *et al.* 2020). Among all the CDT crystal structures, including the *CMCDT and *CDTCM bifunctional enzymes reported in this thesis, only the MES- or PPY-bound *JbCDTCM showed an solvent-accessible active site conformation (Figure 20 B). The ligand-free and acetate-bound CDTs took a closed active site conformation similar to the published *PaeCDT (PDB ID: 5HPQ, (Clifton *et al.* 2018).

In the MES- and PPY- bound *JbCDTCM structures, Trp87, an active site residue, was found in two different conformations (Figure 20 B). In the most frequently observed conformation, the side chain of Trp87 points inward, towards the substrate in the active site, interacting by π -stacking with the aromatic ring of the cyclohexadienyl substrate. However, in presence of PPY the Trp87 side chain was rotated almost 180° towards the solvent. As Trp87 lines the entrance to the CDT active catalytic cleft, its conformational shift might be associated with the active site opening and closing.

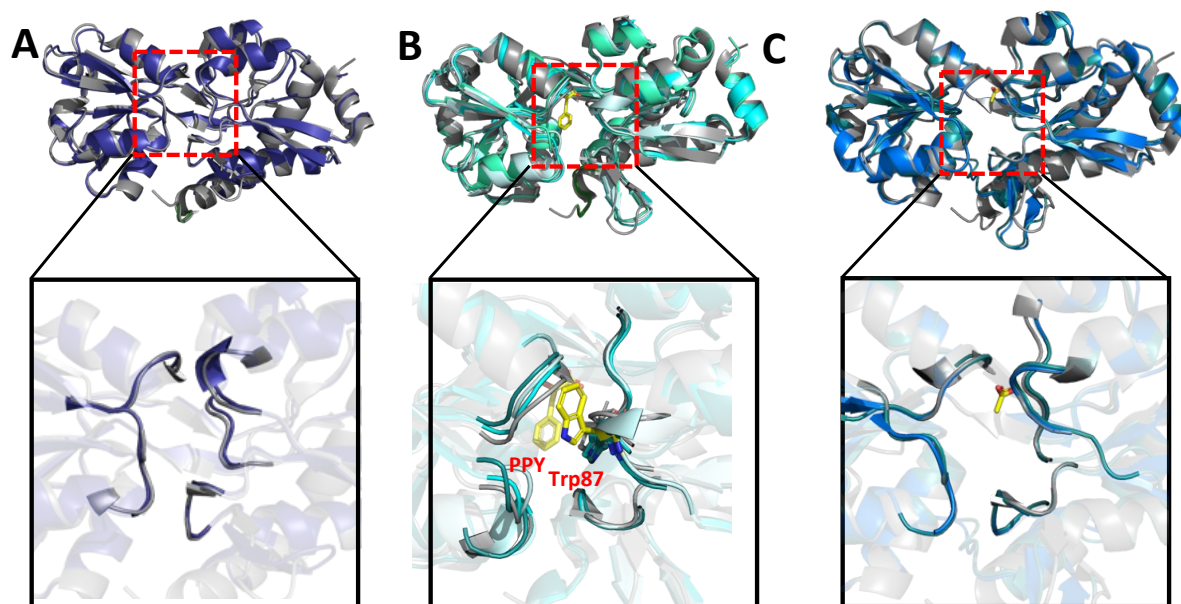


Figure 20 The active site opening of the CDT domains of the bifunctional enzymes compared to the *PaeCDT (grey). A. CDT domain of the ligand-free *DsCDTCM. B. CDT domain of *JbCDTCMs (MES-bound (cyan), PPY bound (green cyan and pale cyan)). The Trp87 is depicted as sticks in each structure, whose side chain flipped 180°outward (open) in PPY-bound *JbCDTCM (pale cyan) C. CDT domain of *AfCMCDT (acetate-bound (blue), ligand-free (teal)).

The equilibrium between the open and close conformations controls the substrate-binding affinity of solute-binding proteins (SBPs), and thus of CDT (Marvin and Hellinga 2001, Seo *et al.* 2014). Clifton *et al.* generated hypothetical ancestral solute-binding proteins (AncCDTs), taking milestones along the evolutionary path from the *Wolinella succinogene* amino acid binding protein (AABP) to *PaeCDT (Clifton *et al.* 2018). This paper described how ancestral SBPs gained a dehydratase activity and how remote mutations can affect the catalytic efficiency by changing global conformational sampling (Clifton *et al.* 2018).

Molecular dynamics and double electron-electron resonance studies revealed that AncCDTs predominantly sample catalytically unfavorable open conformations (‘wide open’ and ‘open’), while in dehydratases (*e.g.* *PaeCDT) the conformational equilibrium is shifted towards more compact states (‘open’ and ‘closed’) (Kaczmariski *et al.* 2020). According to their model, the change in conformational dynamics underlie the substrate-binding and catalytic activities of CDT (Kaczmariski *et al.* 2020). They proposed the emergence of three structural features affecting the conformational dynamics. First, unlike AABPs CDTs form homotrimers, with protomer-protomer contact stabilizing a closed conformation (East *et al.* 2022). Second, the presence of a flexible region neighboring the hinge strand, affecting the open/close transition (Figure 22). This flexible region has been detected as a mutational

hotspot along the evolutionary trajectory, and it is important for the ligand binding affinity in solute-binding proteins (Kaczmarek *et al.* 2020). Third, the C-terminus of CDT folds into a cleft between the large and small subdomains, holding them closer (Figure 23) (Kaczmarek *et al.* 2020, East *et al.* 2022). The last two features are analyzed in detail in the following paragraphs.

3.5.1.2 Intrinsically flexible region neighboring the β -strand (hinge region) of the CDT

To sample wide-open conformations, the flexible region next to the β -strands hinge needs to undergo a conformational change from kinked to an extended conformation in solute-binding proteins. Residues in this region have relatively small, hydrophobic side chains in solute-binding proteins (Kaczmarek *et al.* 2020). Starting from the VHLD sequence found in WsAABP, Leu is mutated to a Pro in *PaeCDT. This introduces a kinked conformation enabling the neighboring glutamine to form hydrogen bond with an arginine residue from the small domain (Arg186) to stabilize the kinked conformation (Figure 22 A; grey). In the bifunctional enzymes analyzed in this thesis, residues in the flexible region are well-conserved among *CDTCMs, while significantly different among *CMCDTs orthologs (Figure 21).

A	*PaeCDT	VHPQ	B	*PaeCDT	VHPQ	*TvCMCDT	TMPD
	*JbCDTCM	VHPD		*AfCMCDT	VLLD	*SbCMCDT	TMPN
	*DsCDTCM	VHPD		*ScCMCDT	ATSA	*SpCMCDT	AMPD
	*MpCDTCM	VHPE		*TaCMCDT	TMNG		

Figure 21. Sequence of the intrinsically flexible region in all nine bifunctional enzymes compared to *PaeCDT. A. *CDTCM orthologs. All three enzymes contain proline residue followed by a negatively charged residue (Asp/Glu). **B.** *CMCDT orthologs. Three bifunctional enzymes, *TvCMCDT, *SbCMCDT, and *SpCMCDT, contains proline residue followed by a hydrophilic residue (Asp/Asn).

The crystal structures revealed that this region adopts a kinked conformation in *JbCDTCM and *DsCDTCM, similar to the *PaeCDT (Figure 22 A), and an extended conformation in *AfCMCDT, similar to WsAABP (Figure 22 B, C). The Asp-Arg salt bridge stabilizing the kink in the flexible observed in *PaeCDT is conserved in *DsCDTCM and *JbCDTCM; however, it is missing from *AfCMCDT (Figure 22 A, B).

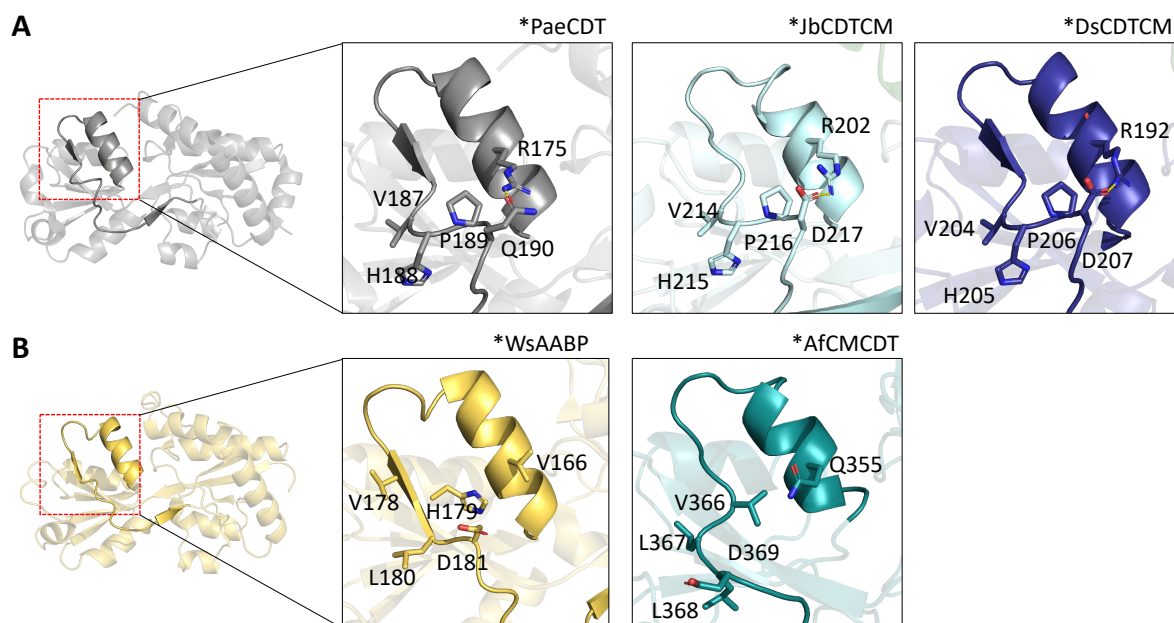


Figure 22. Conformations of the intrinsically flexible region in CDT domain of the bifunctional enzymes observed in crystal structures. **A.** Kinked conformation of the intrinsically flexible loop. *PaeCDT (PDB ID: 5HPQ, (Clifton *et al.* 2018); dark grey), residue ¹⁸⁷V-¹⁹⁰Q. The kinked conformation is fixed by a hydrogen-bond between Q190 and R175. *JbCDTCM, residue 214V-217D. *DsCDTCM, residue 204V-D207. The D217 from the *JbCDTCM and *D207 from the *DsCDTCM form a salt-bridge with R202 and R192, respectively. **B.** An extended conformation of the intrinsically flexible loop. *WsAABP (PDB ID: 3K4U; yellow), residue V178-D181. The neighboring helix misses arginine to establish a salt-bridge with D181. *AfcMCDT, residue V366-D369. The side chain of D369 faces away from the neighboring helix.

3.5.1.3 Extended C-terminus enabling interactions between the large and small subdomains

The C-terminus of *PaeCDT is wedged between the ‘large’ and ‘small’ subdomains, stabilizing the closed active site conformation (Figure 16 C) (Kaczmarek *et al.* 2020). The C-terminus of the CDT domain of bifunctional enzymes shows the same conformation observed in *PaeCDT, binding the large subdomain by polar interactions. In enzymes with a *CDTCM domain arrangement, the C terminus of the CDT domain is fused to the CM domain via a short linker region (Figure 23 A, B). The linker region stabilizes the small subdomain by forming hydrogen bonds (Figure 23 A, B). The C-terminus of *AfcMCDT is embedded into the large subdomain of CDT, engaged in intensive hydrogen bonding, without any interaction with the small subdomain (Figure 23 C).

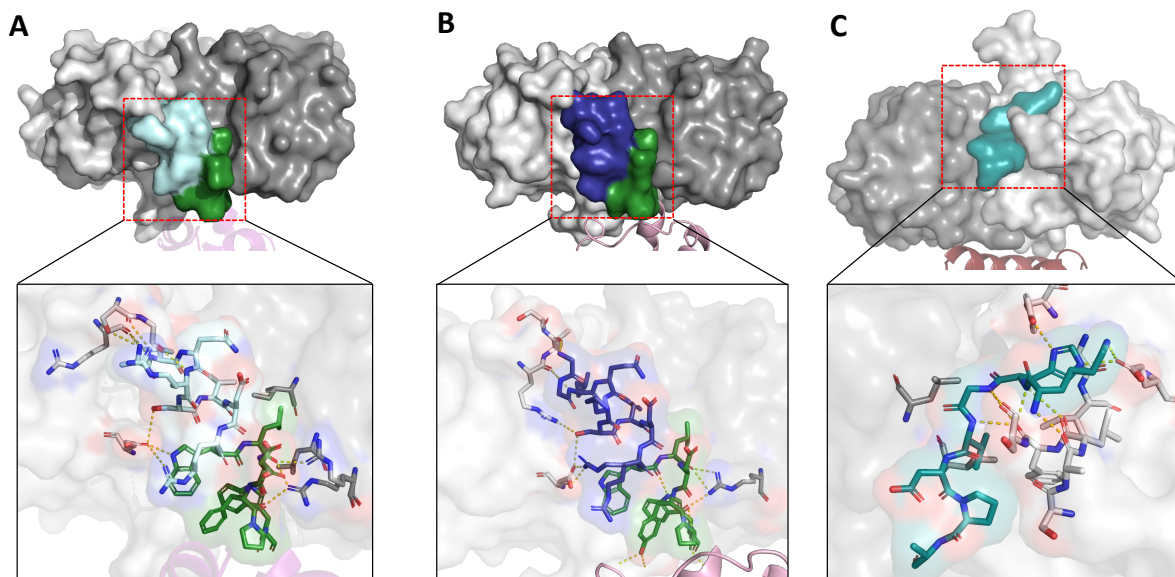


Figure 23. Interaction between C-terminus and two sub-domains of CDT domain of the bifunctional enzymes. **A.** The PPY-bound *JbCDTCM (pale cyan). **B.** The apo *DsCDTCM (purple). **C.** The apo *MgCMCDT (teal). Linker between CM and CDT domains are shown in green. The large and small subdomains of the CDT domains are depicted in light grey and dark grey, respectively. In both domain-swapped orthologs, the C-terminal region is engaged in several hydrogen bonding with the large subdomain. In *CDTCMs, the linker region established further interactions with the small subdomain. C-termini are colored in cyan, purple, and teal for A, B, and C, respectively.

The conformation of the intrinsically flexible region (mutational hotspot) and C-terminal region of CDT are important factors for its activity. The ability to sample a defined range of open/close conformations, which depends on the rigidity of the hinge region and the constraints enforced by the interaction with other protomers, will likely affect the kinetic parameters of CDT enzymes.

The tertiary structures of the domain-swapped bifunctional enzymes, show a significant difference in the intrinsically flexible regions, which may affect their catalytic performance. Kinetic study revealed that the CDT domain of the bifunctional enzymes exhibit kinetic parameters similar to those of the *PaeCDT (k_{cat} and K_m of 12-42 s^{-1} and 3-40 μM vs. 18.4 s^{-1} and 18.7 μM , respectively). Among all wild-type bifunctional enzymes investigated in **Manuscript I**, *AfCMCDT holds that highest K_m value (40 μM) for prephenate, while CDT from *SbCMCDT exhibits K_m value of 3.6 μM which is even lower than those of extracytoplasmic *CDTCMs (7.3 - 18.1 μM). *SbCMCDT contains a proline residue followed by asparagine in its intrinsically flexible region, thereby it may adopt the kinked conformation observed in *CDTCMs and *PaeCDT. Hence, the two *CMCDT orthologs

possibly have a different conformation sampling regarding the hinge-mediated active-site opening reflected by different substrate affinity.

3.5.2 CM domains

3.5.2.1 Binding affinity for chorismate for the CM domains of extracytoplasmic CM/CDT fusion enzymes

Helices H2 and H4 in AroQ_γ-subclass CMs define the opening to the catalytic cleft. The change in the relative distance between the N-termini of H2 and H4, visible comparing the TSA-bound vs. apo *MtCM, exposed the presence of ‘open’ and ‘closed’ active site conformations (Okvist *et al.* 2006). The scissoring motion involving helices H2 and H4, revealed by NMA, provides gating to the active site, and thus mediates substrate uptake and release (**Manuscript II**; Figure 24 A). The distance between the N-termini of helices H2 and H4 was measured for the bifunctional enzymes. The ‘closed’, TSA-bound CM domains from bifunctional enzymes displayed only a slightly wider active site opening compared to their ‘open’ unliganded forms (1.1 Å and 0.8 Å for *JbCDTCM and *AfCMCDT, respectively; Figure 24 B, C).

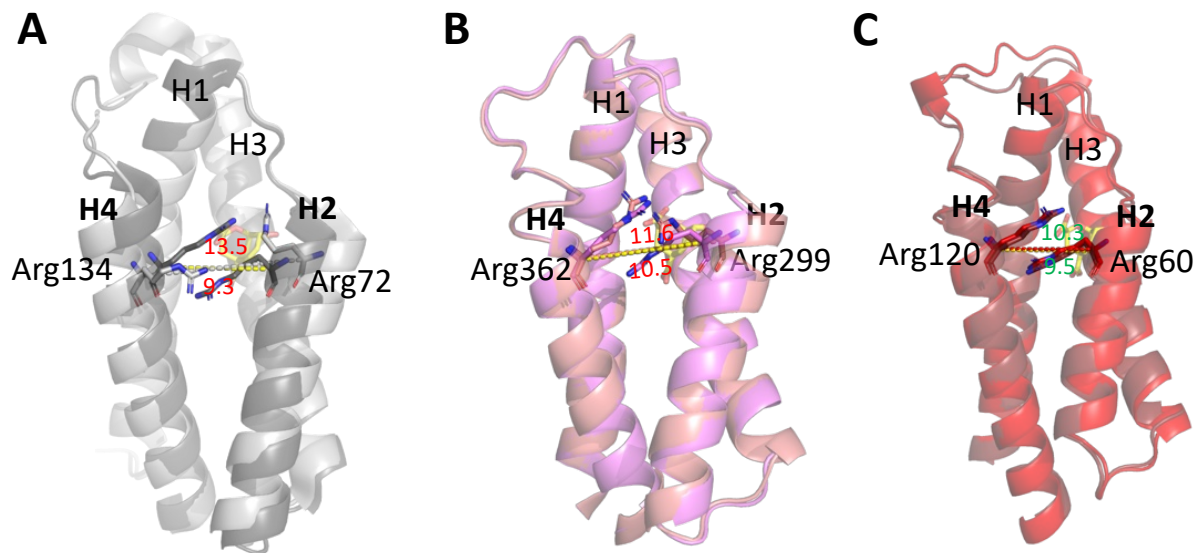


Figure 24. Substrate gating of the CM domain of bifunctional enzymes. The distance between N-termini of the H2 and H4 helices is measured for **A.** *MtCM, unliganded (white grey) and TSA-bound (dark grey) chains (PDB ID: 2FP2, (Okvist *et al.* 2006)). **B.** *JbCDTCM, unliganded (salmon) and TSA-bound (pink). **C.** *AfCMCDT, unliganded (red) and TSA-bound (raspberry). The CM domain of the bifunctional enzymes show a tighter substrate gating compared to *MtCM (PDB ID: 2FP2, (Okvist *et al.* 2006))

Normalized *B*-factor analysis (Johnson *et al.* 2018) of the bifunctional enzymes revealed that the H3-H4 loop is the most mobile region in *JbCDTCM and *AfCMCDT, followed by the

H1-H2 loop and H2 helix (Figure 25). This reflects the positional shift for the same loops observed from NMA on *PaeCM, suggesting that bifunctional and monofunctional AroQ₇-subclass CMs share the same substrate gating mechanism.

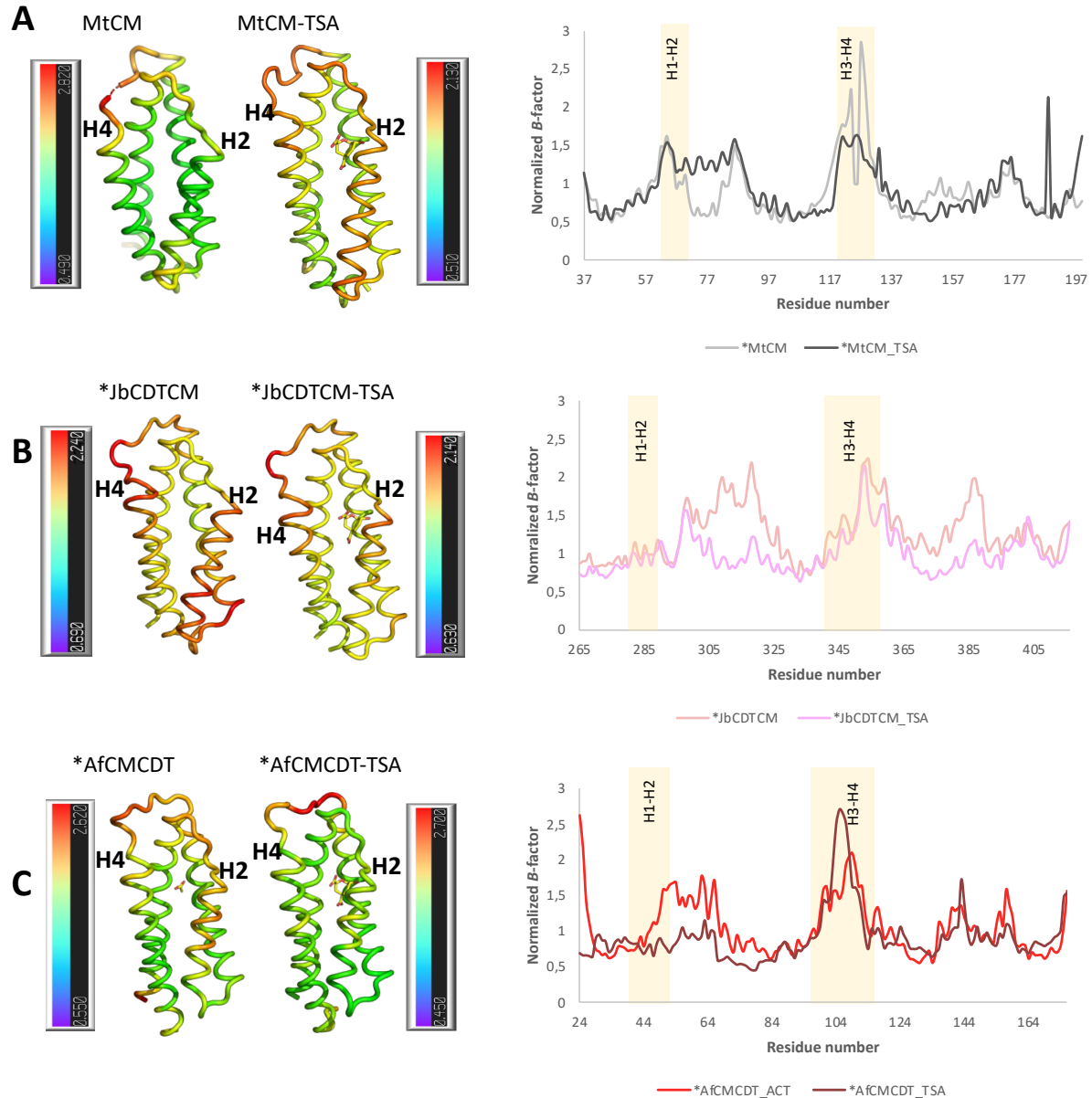


Figure 25. Normalized B -factor analysis of the crystal structures of CM domain of TSA-bound and apo bifunctional enzymes compared to the monofunctional *MtCM. Left panel: cartoon presentation of helices H1 to H4 of the CMs, colored by B -factors (C_{α}). Right panel: Normalized B -factor vs. residue graphs. **A.** *MtCM (PDB ID: 2FP2, (Okvist *et al.* 2006)). **B.** *JbCDTCM. **C.** *AfcMCDT. Residues covering the loops connecting neighboring helices are highlighted in yellow box. Relatively high B -factors are observed for H1-H2 loop, H3-H4 loop, and H2 for all CMs.

In **Manuscript I**, key residues in the active site of the CM domain of the *JbCDTCM, *DsCDTCM and *AfcMCDT enzymes were identified by aligning them with the TSA-bound *MtCM structure (PDB ID: 2FP2, (Okvist *et al.* 2006)). The CM domain from

*CDTCM and *CMCDT fusion enzymes differ from *MtCM by four active site residues: Q76V and T105S or T105A (in *JbCDTCM or *AfCMCDT, respectively), E106K, and E109Q (Figure 26A). This difference in amino acid active site composition was pointed as the possible origin for the high binding affinity ($K_m < 7$) towards chorismate.

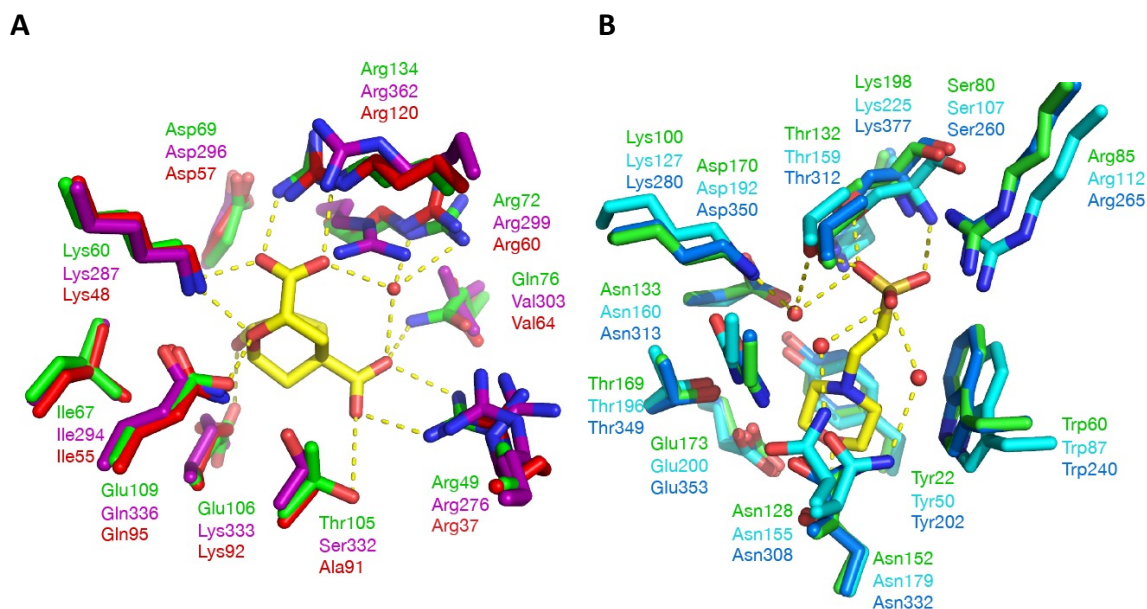


Figure 26. Superimposition of active site residues in CM and CDT. **A.** Superimposition of CM active sites of *AfCMCDT with acetate (red) and *JbCDTCM (magenta) with *MtCM (PDB: 2FP2, (Okvist *et al.* 2006); green), in complex with the *endo*-oxabicyclic TSA; yellow; CM active sites of *AfCMCDT and *JbCDTCM are not occupied). **B.** Superimposition of CDT active sites of *AfCMCDT (blue) and *JbCDTCM (cyan) with *PacCDT (PDB ID: 5HPQ, (Clifton *et al.* 2018) ; green). *JbCDTCM contains a MES molecule from the buffer (yellow) in the CDT active site. Hydrogen-bonding interactions between ligand and protein are depicted as yellow dashed lines. Water molecules are displayed as red spheres.

However, similar to the CDT domain, which lost a wide-open conformation sampling during its evolution from ancestral solute-binding protein, the very low K_m values of the CM domain of bifunctional enzymes compared to monofunctional enzyme could be a consequence of altered conformational dynamics. Split CM domains from the *CM/CDT bifunctional enzymes exhibited similarly low K_m values as their fused forms, suggesting that fusion does not affect the substrate affinity. However, CM domains from bifunctional enzymes lack a disulfide bond between helices H5 and H6, which is conserved in monofunctional CMs and stabilizes their fold. As an alternative explanation to the active site differences, the missing disulfide bridge could be responsible for altered conformational sampling of the CM domain, and explain their different affinity for chorismate.

4 Jamming the engine - understanding fine details of enzyme function (allostery, new inhibitors)

4.1 Allostery in enzymes

Proteins are dynamic molecules, sampling various conformational states on the free energy landscape. This conformational freedom is essential for their function (Papaleo *et al.* 2016, Lisi and Loria 2017). The conformational shift between different states can be induced by environmental factors such as pH, ion concentration, temperature, and ligand binding (Motlagh *et al.* 2014, Nussinov 2016). Due to the dynamic nature of proteins, structural changes at one site can affect a distal site, often with consequences on the function (Motlagh *et al.* 2014). The mechanism by which proteins transmit the effect of binding at one site to a distal one is called *allostery* (Motlagh *et al.* 2014, Lisi and Loria 2017). The degree of the allosteric effect can vary, from as small as side-chain rotations to as large as the conformational change of large regions within the protein (Lisi and Loria 2017). Allosteric effectors can be either activators, enhancing the protein activity, or inhibitors, reducing the protein activity (Nussinov 2016). Effector molecules can be any protein-binding molecules such as DNA, lipids, nucleic acids, small molecule ligands or even other proteins (Motlagh *et al.* 2014). Disordered regions of the proteins, such as surface loops and linkers, are important for allostery as they can easily change conformation and relay allosteric stimuli (Nussinov 2016, Papaleo *et al.* 2016). Allosteric regulation is important to keep cellular homeostasis by modulating enzyme activity in metabolic pathways (Munack *et al.* 2016, Sweetlove and Fernie 2018). In multienzyme complexes, many proteins undergo allosteric regulation triggered by partner enzymes (Munack *et al.* 2016, Schupfner *et al.* 2020). Examples can be found in the shikimate pathway, in the bifunctional CM-PDT (as described previously), the complex of CM and DS from *M. tuberculosis* and the bifunctional CM-DS from, *Listeria monocytogenes*, or *Bacillus subtilis* (Light *et al.* 2012, Munack *et al.* 2016, Pratap *et al.* 2017, Bai *et al.* 2019, Bai and Parker 2021). The naturally sluggish MtCM gets 100-fold activated upon binding to MtDS, forming a non-covalent complex (Munack *et al.* 2016). In the *Prevotella nigrescens* bifunctional CM-DS, the enzymatic activity of each domain is dependent on the other (Bai *et al.* 2019, Bai and Parker 2021). Splitting the two domains dramatically decreases the catalytic activity of both. (Bai and Parker 2021).

4.2 Model system: CM and DAHP synthase from *Mycobacterium tuberculosis*

MtCM, a housekeeping enzyme in *Mycobacterium tuberculosis*, belongs to the AroQ δ subclass, representing naturally mediocre chorismate mutases (Sasso *et al.* 2009). The tertiary structure of MtCM consists of three α -helices (labeled H1, H2 and H3) connected by short loops (Sasso *et al.* 2009). In the biologically active unit, two protomers form an intertwined homodimer (Sasso *et al.* 2009). The dimeric MtCM contains two active sites, each nested within a four-helix bundle formed by helices H1 to H3 from one protomer and H1' from the other protomer (Sasso *et al.* 2009).

4.2.1 CM-DS inter-enzyme allostery

CM needs to work efficiently to satisfy the metabolic need for aromatic amino acids, while MtCM has a low catalytic efficiency (Sasso *et al.* 2009). The catalytic enhancement of MtCM is prompted by another enzyme, DAHP synthase (DS), by forming a non-covalent complex (Sasso *et al.* 2009). Upon the complex formation, MtCM gains >100-fold catalytic efficiency (Sasso *et al.* 2009). *M. tuberculosis* genome encodes for a single DAHP synthase (DS), a 50 kDa enzyme exhibiting a (α/β)₈-triose fold (Rizzi *et al.* 2005). The biological form of DS is homotetrameric, a dimer of two tightly bound dimers, and its activity is regulated by all three proteinogenic aromatic amino acids, Tyr, Trp, and Phe (Sasso *et al.* 2009, Munack *et al.* 2016) (Figure 2 D).

The CM-DS complex is a heterooctamer, where the homotetrameric DS forms the core and two CM dimers dock onto it from two opposite sites (Figure 27 A) (Sasso *et al.* 2009, Munack *et al.* 2016). Each DS protomer interacts with one CM protomer. Upon complex formation, DS exhibits only minor conformational changes at the CM docking site; however, it induces a kinked conformation in the C-terminus of CM (Figure 27 C) (Munack *et al.* 2016). This, in turn, enables the interaction of the CM C-terminus with the H1-H2 loop of the same CM protomer, involving a hydrophobic contact between Leu54 and Leu88, and a salt-bridge formed by Asp53 and the C-terminal carboxylate (Figure 27 C).

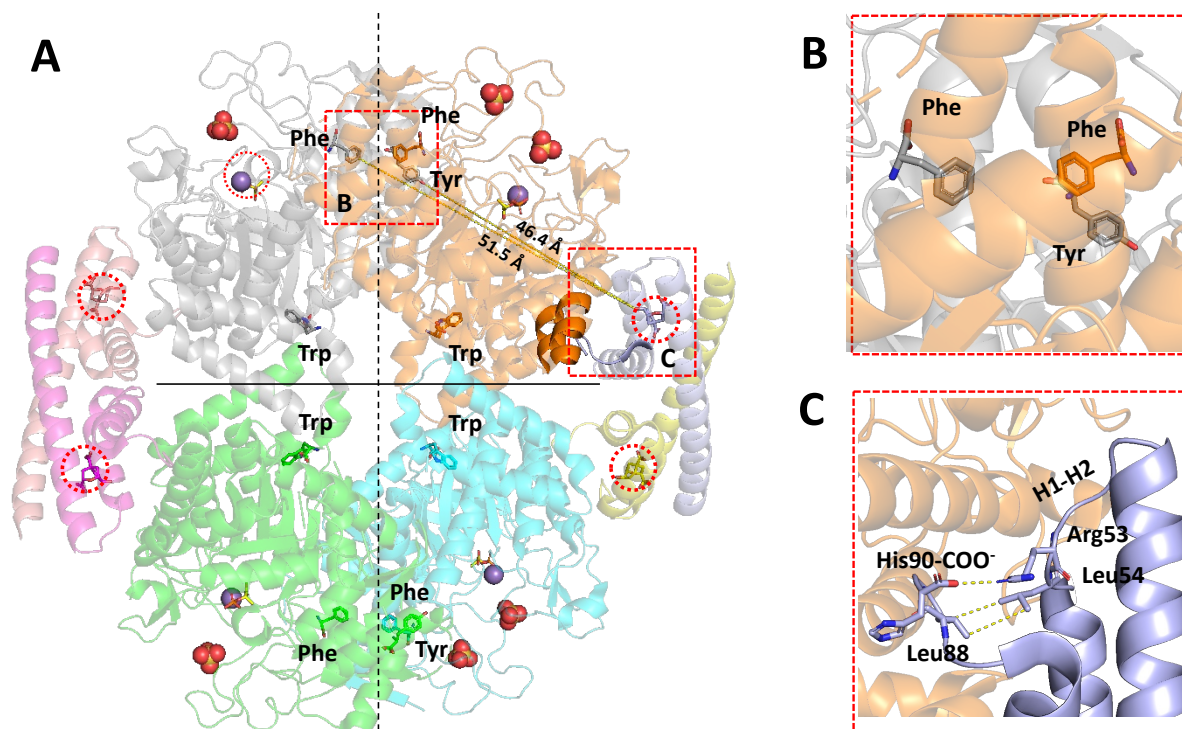


Figure 27. Allosteric regulation in aromatic amino acids biosynthesis pathway. **A.** A non-covalent heterooctameric complex CM-DS for *M. tuberculosis* (PDB ID: 5CKX, (Munack *et al.* 2016)). Dimerization interface is shown as dash line, and tetramerization interface is shown as line. Allosteric regulators, Phe, Tyr, and Trp, bound to the DS are depicted as sticks at the dimerization and tetramerization interfaces. The active site of CM and DS are highlighted by circle (red). The distance between Tyr and Phe binding sites and closest CM active site is shown as a dash line (yellow). **B.** Binding of Phe and Tyr. The cavity formed at the dimerization interface accommodates two Phe and a Tyr. **C.** The interaction between CM and DS which enables interaction between C-terminus of CM with a kinked H1-H2 loop.

4.2.2 Allosteric CM-DS feedback regulation by aromatic amino acids

MtDS exhibits two types of allosteric effects on MtCM: it rearranges the CM active site residues to adopt a catalytically favorable conformation, and it modulates the MtCM activity in the MtCM-DS complex via a yet unknown allosteric signaling route (Munack *et al.* 2016, Fahrig-Kamarauskait *et al.* 2020). Binding of aromatic amino acids to DS allosterically modulates the enzymatic activity of both DS and CM, and constitutes an example of inter-enzyme allostery (Munack *et al.* 2016). The coupled inhibitory effect of the aromatic amino acids was reported in several studies (Munack *et al.* 2016, Burschowsky *et al.* 2018). The binding of Phe, Tyr and Trp synergistically inhibit MtDS activity, thus shutting down the entire shikimate pathway (Munack *et al.* 2016). Phe and Tyr jointly bind 55 Å and 45 Å away, respectively, from the closest CM active sites, in a cavity formed at the DS ‘tight’ dimerization interface (Figure 27 B). Their binding induces the dissociation of the CM-DS complex (Munack *et al.* 2016). Despite knowing the exact position of the allosteric binding

MtCM activated by MtDS adopts a kinked H1-H2 loop conformation (PDB ID: 2W1A and 2W19 (Sasso *et al.* 2009), (Figure 28 A, C). The same feature was observed in the crystal structure of the super-active MtCM^V variant (Fahrig-Kamarauskait *et al.* 2020), opposed to the extended conformation seen in wild-type, low-activity MtCM (PDB ID: 2VKL, 2QBV (Kim *et al.* 2008, Sasso *et al.* 2009)) (Figure 28 B) In the active MtCM the side chain of Arg46, a catalytically important residue, adopts a catalytically favorable conformation, while in wild-type apo MtCM this residue points away from the substrate, adopting a catalytically unfavorable conformation (Figure 28 C).

Therefore, the kinked H1-H2 loop conformation and its interaction with C-terminus was believed to be important for gaining catalytic efficiency. Substitutions introduced in the H1-H2 loop, T52P and V55D, improved k_{cat}/K_m of MtCM by 6-fold and 12-fold, respectively, and 22-fold when combined. The structural effect of each replacement could not be evaluated from the crystallographic model due to the presence of extensive crystal contacts (Figure 29 A).

In **Manuscript III** we aimed to gain a deeper understanding of the functional importance of critical replacements introduced in the super-active MtCM^V using enzyme kinetics, crystal structures and molecular dynamics simulation.

Table 4. Enzyme kinetics assay carried out in Manuscript III

Variant	Residue changes	k_{cat} (s ⁻¹) ^a	K_m (μM) ^a	k_{cat}/K_m (M ⁻¹ s ⁻¹) ^a
MtCM ^V (PDAM)	PD/PDAM ^b , V62I, D72V, V11L, D15V, K40Q	9.4 ± 1.3	22 ± 2	430,000 ± 30,000
MtCM PNAM	MtCM ^V , D88N	7.6 ± 0.2	45 ± 7	170,000 ± 20,000
MtCM PLAM	MtCM ^V , D88L	6.0 ± 0.4	38 ± 1	160,000 ± 10,000
MtCM wt	–	1.7 ± 0.2	980 ± 80	1,700 ± 300
MtCM L88D	L88D	3.0 ± 0.1	1110 ± 70	2,700 ± 300
MtCM 3p3	T52P V55D	8.5 ± 0.0	450 ± 70	19,000 ± 3,000
MtCM Triple	T52P V55D L88D	11 ± 2	510 ± 130	22,000 ± 1,000

PD/PDAM indicates amino acid substitutions T52P, V55D, R87P, L88D, G89A, and H90M.

PNAM indicates amino acid substitutions R87P, L88N, G89A, and H90M

PLAM indicates amino acid substitutions R87P, G89A, and H90M

4.2.3 Manuscript III, Summary

Enzymes kinetics assay was carried out for C-terminal MtCM variants. Single amino acid substitution, L88D, caused 1.6-fold catalytic activity, revealing the importance of negatively charged residue in the C-terminus (Table 4). The importance of the L88D substitution was studied by comparing its effect on catalytic activity of the wild-type, super-active, and double mutant (V55D, T52P) variants. The L88D substitution, however, on top of double mutant (T52P and V55D), Asp88 did not lead to a significant increase compared to the double mutant (Table 4). Hence, the L88D substitution is less important in the presence of H1-H2 loop mutations, but still beneficial in context of other selected mutations.

The crystal structures of two single MtCM variants, MtCM^{T52P} and MtCM^{V55D}, were determined to 1.6 Å and 2.0 Å, respectively. The T52P replacement was thought to be responsible for a kinked H1-H2 loop conformation in the super-active MtCM^V variant, however, in our crystal structure of MtCM^{T52P} variant, the H1-H2 loop adopted an extended conformation similar to wild-type MtCM (Figure 29 B). A partially refined model of MtCM^{V55D} variant revealed a kinked H1-H2 loop conformation, however, the loop is involved in extensive crystal contacts (Figure 29 C, D). V55D substitution introduced a carboxylate group in the N-terminal of helix H2, which forms a salt-bridge to Arg46.

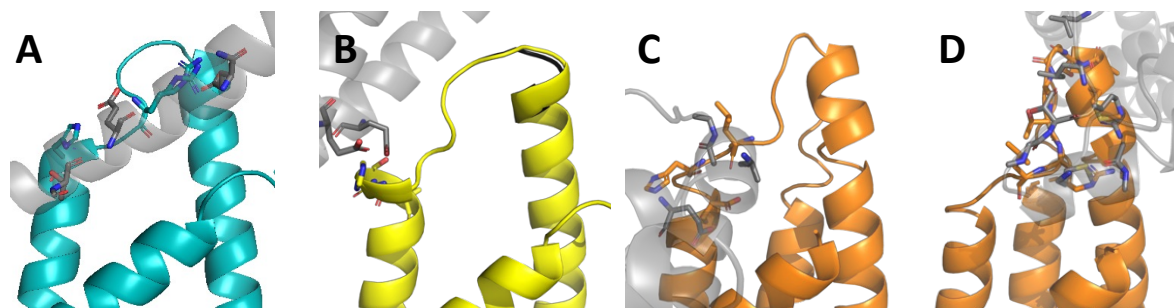


Figure 29. Crystal contacts in the H1-H2 loop of MtCM variants. **A.** Super-active MtCM^V, where kinked H1-H2 loop is involved in extensive crystal contacts. **B.** MtCM^{T52P}, where H1-H2 loop adopted an extended conformation like wild-type MtCM. **C.** MtCM^{V55D}, protomer A. **D.** MtCM^{V55D}, protomer B. In MtCM^{V55D} variant, the H1-H2 loop in both protomer adopted a kinked conformation and is involved in crystal contacts.

Therefore, we performed a microsecond molecular dynamics simulation on the wild-type apo MtCM (PDB ID: 2VKL (Sasso *et al.* 2009)) and the super-active MtCM^V variant (PDB ID: 5MPV, (Fahrig-Kamarauskait *et al.* 2020)). To study the effect of ligand, we also included TSA-bound MtCM^{LC} from MtCM-MtDS complex (PDB ID: 2W1A, (Sasso *et al.* 2009)).

Effect of ligand binding: Wild-type, apo MtCM revealed a significant structural flexibility compared to the TSA-bound MtCM. Two protomers making up the biological dimer broke their symmetry, independently exploring different conformations. In the absence of a ligand, positively charged active site residues (Arg18', Arg35, Arg46 and Arg 58) (Figure 3B) undergo conformational fluctuations caused by electrostatic repulsion. Binding of a negatively charged ligand, TSA, significantly rigidifies the structure and reorganizes the active site residues.

Kinked H1-H2 loop: Conformational landscape of H1-H2 loop was investigated during the simulation using Arg53 as a reported residue. The H1-H2 loop in MtCM^V and TSA-bound MtCM^{LC} retained its kinked conformation throughout the simulation, while in wild-type, apo MtCM, it was preferentially found in extended conformation. Asp55 substitution enables a salt-bridge formation to Asp 46 and Arg18', promoting preorganization of the active site region (Figure 30).

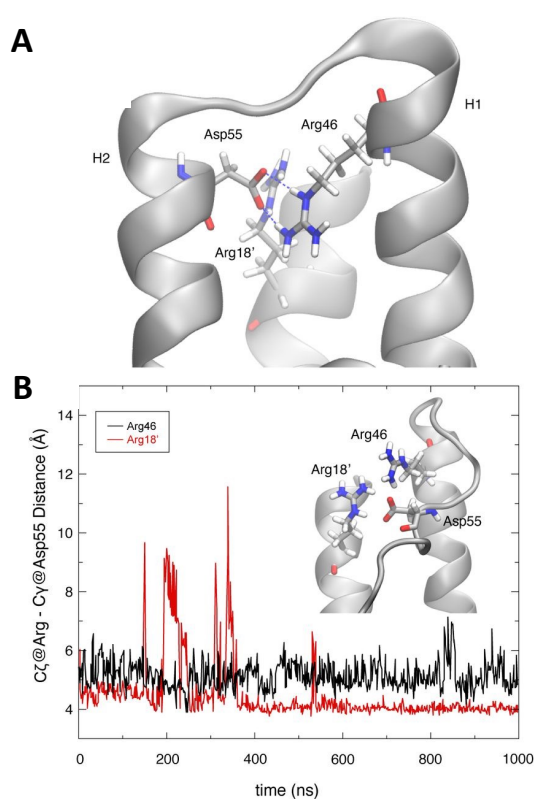


Figure 30. Role of MtCM^V residue Asp55 in positioning active site residues. A. Extension of H2 and stabilization of the H1-H2 loop by residue Asp55. Substitution of Val55 by Asp stabilizes helix H2 through interactions with Arg18' and Arg46 across the active site (the image shows the structure of V55D after 1 μ s of MD simulations). B. Distance plotted between MtCM^V Arg46 (black, chain A) or Arg18' (red, from chain B) and Asp55 (chain A) observed during the simulation. In both cases, the distance measured is between Asp C γ and Arg C ζ , using PDB nomenclature. Figure adapted from **Manuscript III**, prepared by Michele Cascella.

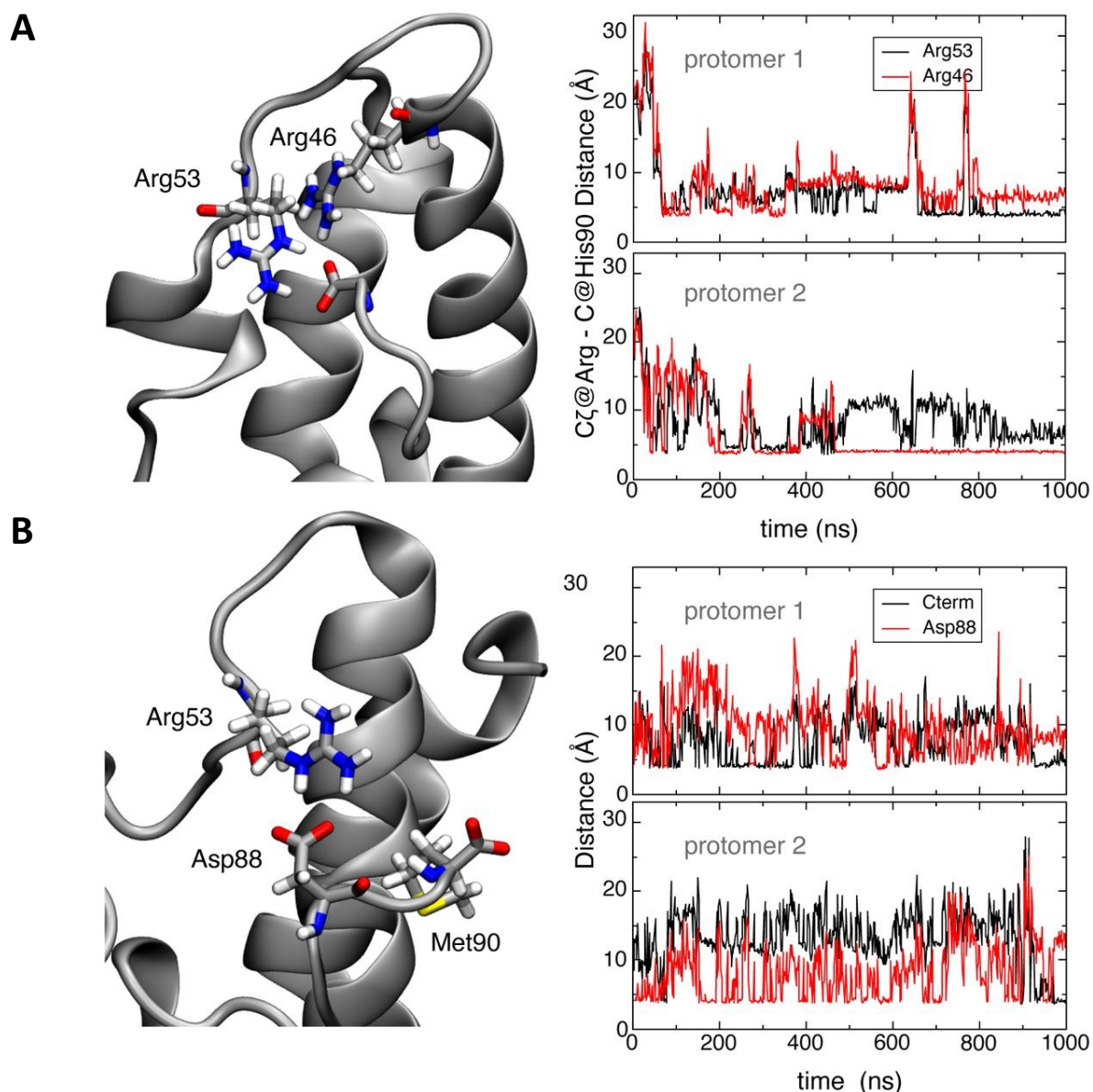


Figure 31. Interaction between C-terminal carboxylate of MtCM and H1-H2 loop. A. The right panels show the distance between the C_{ζ} carbon of arginine residues 53 or 46 and the carboxylate carbon of the C-terminus of MtCM in the two protomers (upper and lower panels). Formation of a steady contact ($< 5 \text{ \AA}$) with Arg46 (bottom panel) corresponds to stabilization of the catalytically productive H1-H2 loop conformation, which allows for stabilization of the transition state of the chorismate to prephenate rearrangement. **B.** Interaction between C-terminal residues and the H1-H2 loop in MtCM^V. Salt-bridge contacts between Arg53 and the carboxyl groups of Asp88 (red line) and Met90 (C-terminus; black line) in MtCM^V in the two protomers. The top and bottom panels on the right show the evolution of the distances over time between Arg53's C_{ζ} and the corresponding carboxylate carbons for each of the two protomers of MtCM^V. Figure adapted from **Manuscript III**, prepared by Michele Cascella.

In MtCM^V, a newly introduced Asp88 offer an alternative mode of salt-bridge with Arg53, thus, strengthening the H1-H2 loop and C-terminal interaction throughout the simulation. The kinked H1-H2 loop conformation in In MtCM^V imposes a steric block preventing Arg46,

catalytically important residue, from stretching to the C-terminal carboxylate. In wild-type, apo MtCM, C-terminal carboxylate interchangeably interacts with Arg46 and Arg53, thus, forcing Arg46 in catalytically unfavorable conformation (Figure 31). The effect of Pro52 seems to be small on structural basis, however, the substitution increased melting temperature of MtCM by 2°C and catalytic activity by 6-fold. We assume that Pro52 pre-stabilizes the H1-H2 loop, thereby, reducing the entropic costs for substrate binding.

4.3 Fragment-screening for MtCM (preliminary results from FragMAX and DSF)

Ligand screening is an alternative (or parallel) method to searching for allosteric binding sites by structural analysis or molecular dynamics (Klebe 2006, Zhou *et al.* 2021). Ligand screening is carried out by testing the protein of interest against a library of chemical compounds through different experimental techniques. In case of fragment screening, the library is formed by small organic molecules, or *fragments*, which correspond to common building blocks of more complex drug molecules (Rees *et al.* 2004, Schultes *et al.* 2010).

We attempted to run fragment screening on the MtCM-MtDS complex, with the aim to discover new ligands for known and previously undiscovered binding sites. In particular, we searched for bound fragments providing leads to active site inhibitors, inhibitors binding to known and novel allosteric binding sites or allosteric activation enhancers. The latter would lead to depleting the metabolic resources of the host organism by locking the shikimate pathway into a hyper- active state.

We used X-ray crystallography as the primary screening method, as it has a higher sensitivity for low affinity molecules and provides direct structural information on protein-ligand interactions. As an orthogonal validation method, a fluorescent-based thermal shift assay was employed (Differential Scanning Fluorimetry – DSF, also known as ‘Thermofluor’ (Zhang and Monsma 2010, Senisterra *et al.* 2012). Unfortunately, our initial efforts to establish a robust crystallization system, generating reproducible, diffraction quality MtCM-DS crystals failed, so we switched out target to MtCM alone. We screened 136 compounds from the FragMAXlib library by X-ray crystallography, and 171 compounds by DSF (Lima *et al.* 2020).

Using PanDDA, a crystallographic event map-based software for ligand identification at the FragMAX facility (Pearce *et al.* 2017, Lima *et al.* 2020) (MAX IV, Lund, Sweden), we identified a single hit (G5, FragMAXlib-1 library). A close inspection at the event map revealed a positive electron density blob at the active site of MtCM, which was not visible on the averaged ground-state map; however, the compound could not be fit into the residual density. Additionally, the critical active site residue Arg46 adopted a non-favorable conformation for catalysis as observed in free, wild-type MtCM, pointing to the hit being a false positive.

The melting temperature of wild-type MtCM was measured by DSF to be $70.5 \pm 0.6^\circ\text{C}$. Three compounds shifted the melting temperature positively in a DSF screen of the FragMAX library (FragMAXlib-1-B1, FragMAXlib-2-A7 and FragMAXlib-2-D11; Table 5). As second iteration step, MtCM crystals were resoaked in FragMAXlib-1-G5 and FragMAXlib-2-D11, prepared in the DSF buffer (100 mM MOPS pH 7.0, and 150 mM NaCl). Although we collected multiple diffraction data, no residual density for the resoaked fragments could be observed in the electron density maps.

Ligand binding can either stabilize or destabilize the protein fold (Cimpmperman *et al.* 2008, Kabir *et al.* 2016). The number of fragments reducing melting temperature by more than 2°C was considerably higher than the stabilizing one, leading to 40 more possible candidates (Table 6). Due to time constraints, we did not look further into the destabilizing fragments identified by DSF, and they will be the subject of future investigations. Future efforts will also try to expand the screening to incorporate virtual docking and test different fragment libraries.

Table 5. Fragments positively shifted the melting temperature of MtCM

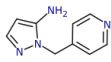
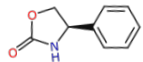
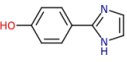
FragMAXlib-1			
Well	ΔT_m ($^\circ\text{C}$)	Smiles	Structure
B1	1.2	<chem>NC1=CC=NN1CC1=CC=NC=C1</chem>	
FragMAXlib-2			
Well	ΔT_m ($^\circ\text{C}$)	Smiles	Structure
A7	0.8 $^\circ\text{C}$	<chem>O=C1N[C@@H](CO1)C1=CC=CC=C1</chem>	
D11	2.7	<chem>C1=CC(=CC=C1C2=NC=CN2)O</chem>	

Table 6. Fragments negatively shifting the melting temperature of MtCM

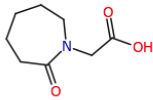
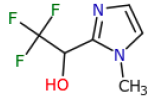
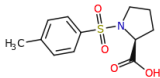
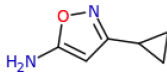
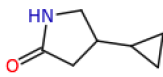
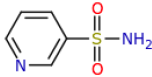
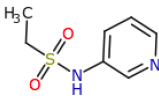
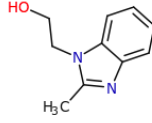
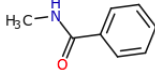
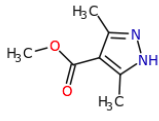
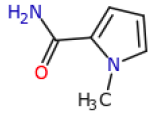
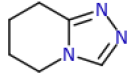
FragMAXlib-1			
Well	ΔT_m (°C)	Smiles	Structure
A6	3.3	<chem>OC(=O)CN1CCCCC1=O</chem>	
B8	3.0	<chem>CN1C=CN=C1C(O)C(F)(F)F</chem>	
C2	2.6	<chem>CC1=CC=C(C=C1)S(=O)(=O)N1CCC[C@H]1C(=O)O</chem>	
C7	2.6	<chem>NC1=CC(=NO1)C1CC1</chem>	
C9	3.0	<chem>O=C1CC(CN1)C1CC1</chem>	
E4	3.2	<chem>NS(=O)(=O)C1=CN=CC=C1</chem>	
E8	2.6	<chem>CCS(=O)(=O)NC1=CC=CN=C1</chem>	
E10	3.5	<chem>CC1=NC2=CC=CC=C2N1CCO</chem>	
F1	3.5	<chem>CNC(=O)C1=CC=CC=C1</chem>	
F9	5.0	<chem>COC(=O)C1=C(C)NN=C1C</chem>	
F10	4.2	<chem>CN1C=CC=C1C(N)=O</chem>	
F12	2.9	<chem>C1CCC2=NN=CN2C1</chem>	

Table 6 (continued). Fragments negatively shifting the melting temperature of MtCM

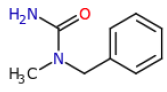
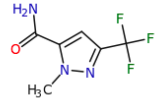
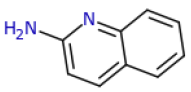

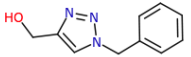
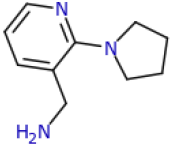
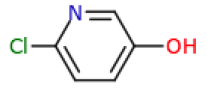
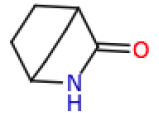
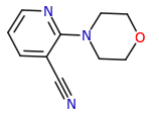
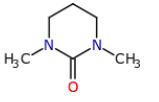
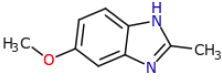
FragMAXlib-1			
Well	ΔT_m (°C)	Smiles	Structure
G1	4.0	<chem>CN(CC1=CC=CC=C1)C(N)=O</chem>	
G2	2.9	<chem>CN1N=C(C=C1C(N)=O)C(F)(F)F</chem>	
G3	2.8	<chem>NC1=CC=C2C=CC=CC2=N1</chem>	
G5	3.8	<chem>CN(C)S(=O)(=O)C1=CC=CC=C1F</chem>	
G11	3.3	<chem>OCC1=CN(CC2=CC=CC=C2)N=N1</chem>	
H1	4.9	<chem>NCC1=CC=CN=C1N1CCCC1</chem>	
H11	2.8	<chem>OC1=CC=C(Cl)N=C1</chem>	
FragMAXlib-2			
Well	ΔT_m (°C)	Smiles	Structure
A8	3.4	<chem>O=C1NC2CCC1CC2</chem>	
B6	2.4	<chem>N#CC1=CC=CN=C1N1CCOCC1</chem>	
B8	6.2	<chem>CN1CCCN(C)C1=O</chem>	
B9	5.2	<chem>COC1=CC=C2NC(C)=NC2=C1</chem>	

Table 6 (continued). Fragments negatively shifting the melting temperature of MtCM

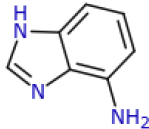
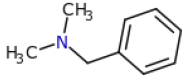
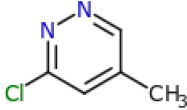
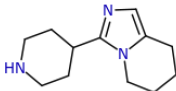
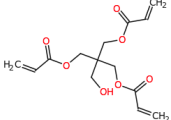
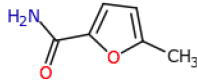
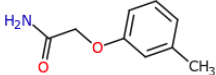
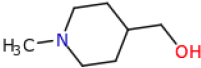
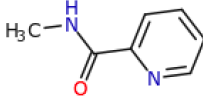
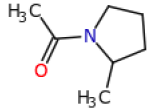
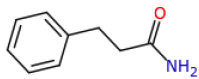
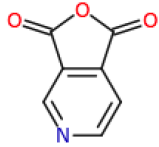
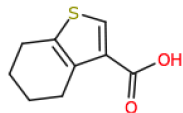
FragMAXlib-1			
Well	ΔT_m (°C)	Smiles	Structure
B10	3.2	<chem>NC1=C2N=CNC2=CC=C1</chem>	
C6	3.4	<chem>CN(C)CC1=CC=CC=C1</chem>	
C8	2.8	<chem>Cc1cc(Cl)nnc1</chem>	
D1	3.5	<chem>CN(C)C(=O)C1=CC=CN1</chem>	
D10	3.7	<chem>C1CCN2C(=CN=C2C3CCNCC3)C1</chem>	
E2	2.8	<chem>C=CC(=O)OCC(CO)(COC(=O)C=C)CO</chem> <chem>C(=O)C=C</chem>	
E7	4.3	<chem>CC1=CC=C(O1)C(=O)N</chem>	
E9	2.7	<chem>CC1=CC(=CC=C1)OCC(=O)N</chem>	
E10	3.4	<chem>CN1CCC(CC1)CO</chem>	
E11	4.4	<chem>CNC(=O)C1=CC=CC=N1</chem>	
E12	4.4	<chem>CC1CCCN1C(=O)C</chem>	

Table 6 (continued). Fragments negatively shifting the melting temperature of MtCM

FragMAXlib-1			
Well	ΔT_m (°C)	Smiles	Structure
F4	2.9	<chem>C1=CC=C(C=C1)CCC(=O)N</chem>	
F6	3.5	<chem>C1=CN=CC2=C1C(=O)OC2=O</chem>	
F10	2.7	<chem>C1CCC2=C(C1)C(=CS2)C(=O)O</chem>	

5 Conclusions and future prospects

This PhD thesis was aimed at the structure-functional study of chorismate mutase (CM), an enzyme involved in the aromatic amino acids biosynthesis pathway in bacteria, plant, and fungi. The pathway is a very promising source of novel drug targets, as it is unique to these organisms, many of them involved in unwanted processes, such as human infection, food spoilage and plant parasitosis. CM is a branch-point enzyme, directing the pathway towards the synthesis of L-Phe or L-Tyr. It catalyzes the conversion of chorismate to prephenate through a pericyclic reaction, one of the few natural enzymes doing so. Several CM classes have been identified, carrying different folds and a highly conserved catalytic machinery and reaction mechanism. Among them, the cytoplasmic AroQ_δ-subclass CMs are naturally mediocre enzymes, requiring allosteric activation. On the contrary, extracytoplasmic AroQ_γ-subclass CMs, periplasmic or secreted, do not require activation, although their role is very poorly understood. We tried to lay the foundation for the future discovery of new drug targets by gaining precise understanding on its allosteric regulation and its mode of interaction with associated enzymes.

In **Manuscript I** we analyzed periplasmic enzymes isolated from the genome of non-pathogenic γ - and β -proteobacteria, carrying a AroQ_γ-subclass CM domain fused to a cyclohexadienyl dehydratase (CDT) domain in two different topologies: *CMCDT and *CDTCM. CM and CDT catalyze sequential reactions in the terminal pathway leading to the synthesis of L-Phe. Postulating that they could provide a paradigm for communication between *fused* sequential enzymes, we aimed to understand the evolutionary reason for the genetic fusion of CM and CDT. We probed the presence of a catalytic advantage granted by the fusion of the two domains by using enzyme kinetics. The relative orientation of the CM and CDT active sites and existence of substrate-channeling was investigated by X-ray crystallography. Overall, no catalytic benefit for the fusion of sequential enzymes was apparent, reflected by the lack of a catalytic rate enhancement in presence or absence of the other domain. Furthermore, we could not detect the formation of multimers of any of the enzymes, nor any indication to the presence of substrate channeling, either intra or intermolecular. Based on the extremely high affinity of the CM domain for its substrate ($K_m < 7 \mu\text{M}$), chorismate, and the presence of genes encoding transporters and substrate binding proteins in the same operon, we hypothesize that these bifunctional fusion enzymes play a

role in signaling, rather than the biosynthesis of aromatic amino acids. In addition to the results reported in Manuscript I, we collected crystallographic and SAXS data on other CM/CDTs orthologs, which confirm the lack of multimerization and substrate channeling for the fusion enzyme tested in this thesis.

In **Manuscript II** we conducted a parallel investigation on the quaternary structure and association of exported monomolecular AroQ_γ-subclass CM and CDT enzymes found in the periplasm of *Pseudomonas aeruginosa* (*PaeCM and *PaeCDT), which have a fused CM-PDT cytoplasmic counterpart (PaeCMPDT). *P. aeruginosa* is an opportunistic pathogen targeting immunocompromised patients (*e.g.* cancer or HIV patients). The structure of *PaeCM and *PaeCDT was studied by X-ray crystallography and small-angle X-ray scattering (SAXS), while their association was probed by size-exclusion chromatography (SEC). The collected data suggest that *PaeCDT and *PaeCM do not form a stable complex.

Both in case of the fusion bifunctional enzymes and the periplasmic *P. aeruginosa* monofunctional enzymes, the original hypothesis was not confirmed by experimental results. *P. aeruginosa* periplasmic monofunctional enzymes do not seem to emulate their fusion counterpart in the cytoplasm by forming a stable complex; periplasmic bifunctional enzymes do not gain any catalytic advantage, except their colocalization, nor regulation from the fusion of the CM and CDT domains. End-products inhibition assay and genomic analysis suggest that bifunctional enzymes are not even involved in the biosynthesis of aromatic amino acids; they rather transform or synthesize metabolites destined to intercellular signaling. This might very well be the case for *PaeCM and *PaeCDT, and can constitute a new paradigm to interfere with the bacterial physiology, by *e.g.* feeding it the ‘wrong’ molecular signal, or inhibiting cellular communication with the infection host and other bacteria.

However, we did gain insights into the dynamic behavior of the CDT and the AroQ_γ-subclass CM domains. In case of the latter, past observations revealed the existence of an open and close state for the catalytic cleft, a feature confirmed by the structural analysis on the CM domain of bifunctional enzymes. Normal mode analysis (NMA) performed on *PaeCM points to a coupling between the breathing motions of the enzyme and active site gating, mediated by the scissoring motion of the two α -helices (H2, H4 in the AroQ_γ-subclass) that define the entrance to the catalytic cleft. Running molecular dynamic simulation on different

CM classes could lead to identify dynamic hotspots that can be targeted for inhibition or hyperactivation.

On the other end, the active site gating mechanism of CDT is well known, and it relies on the hinge movement of the CDT small and large subdomains, held together by a double- β -strand joint. This was confirmed by the SAXS data we collected on *PaeCDT, showing the flexibility of the small domain. The CDT activity derives from ancestral solute-binding proteins (SBPs), which progressively lost their conformational freedom by accumulating mutations in the hinge region and forming stable trimers (Kaczmarek *et al.* 2020). The need to rigidify the CDT fold could be the underlying reason for the formation of a fusion enzyme with CM, and might be a general mechanism used by enzymes to tune their activity. Quite interestingly, the CM domain of periplasmic bifunctional enzymes shows a higher substrate affinity than the CM from *Mycobacterium tuberculosis*, a monomolecular AroQ γ -subclass CM ($K_m < 7 \mu\text{M}$ vs. $180 \mu\text{M}$). While the fusion with CDT might not affect its catalytic efficiency, it might modulate the dynamic properties of the CM domain to bind to its substrate in the extracellular environment. Targeted molecular dynamics simulations and mutational assays to swap the CM/CDT active sites to match that of their monomolecular orthologs might shed further light on this hypothesis.

AroQ δ -subclass CMs undergo inter-enzyme allosteric activation by forming a complex with DAHP synthase (DS), the first enzyme of the aromatic amino acid biosynthesis pathway. We took as model system CM from *Mycobacterium tuberculosis* (MtCM), an AroQ δ -subclass CM. As *M. tuberculosis* is the causative agent of the tuberculosis disease, understanding in deep detail the MtCM activation can lead to the design of effective drugs. In **Manuscript III** we aimed to elucidate the structure-functional determinants underlying the allosteric activation of MtCM. We carried out structural studies and enzyme kinetics analysis targeting the wild-type MtCM (MtCM^{WT}), a laboratory-evolved super-active variant (MtCM^V) and two variants boosting the catalytic activity by a factor of six (MtCM^{T52P}) and twelve (MtCM^{V55D}). We investigated the structural importance of mutations introduced by directed evolution by comparing the conformation sampling of three MtCMs: wild-type (mediocre), highly active (lab-evolved), and naturally activated (in complex with DS) using molecular dynamics simulation. Our results revealed that the T52P and V55D substitutions in H1-H2 loop covering the active site help to rigidify an active-site loop that carries catalytically crucial residues, getting them pre-organized for catalysis. This does not fully explain the 100x

increase in the catalytic efficiency observed upon the formation of the CM-CS heterooctameric complex, nor the CM allosteric regulation provided by aromatic amino acids binding to DS. Further simulations on the full CM-DS complex might provide further understanding of the inter-enzyme allosteric process.

Finally, we aimed to discover potent small-molecule compounds providing leads to inhibitors and allosteric modulators for MtCM by fragment-based screening using X-ray crystallography and differential scanning fluorimetry (DSF). The original plan to screen the MtCM-DS complex could not be carried out, as crystals of the complex proved to be poorly reproducible. However, MtCM alone easily yielded well-diffracting crystals, and was tested with 171 fragments from the FragMAX-1 and FragMAX-2 libraries, developed at the FragMAX facility (MAX IV synchrotron, Lund, Sweden). X-ray diffraction data collected on crystals soaked with the fragments reported one hit, while three more fragments led to a stabilizing effect in DSF measurement. Unfortunately, the hit from crystal screening proved to be a false positive, while none of the DSF hits provided structural evidence of binding when soaked into MtCM crystals. This could be traced to a number of reasons, which include MtCM being locked into a non-binding conformation within the crystal, the presence of crystal contacts preventing the binding of fragment in allosteric hotspots or the use of the wrong fragment library. Future efforts would focus on either testing different ligand libraries, or re-testing FragMAXlib on a different crystal form. As an alternative, the target could be shifted to the superactive variant, which represents a pre-activated state, or thorough screening could lead to a reproducible and well-diffracting crystal form for the MtCM-DS complex, which could be finally used for fragment screening.

6 Materials and Methods

6.1 Recombinant gene expression and protein purification

6.1.1 Novel and exported bifunctional *CMCDT and *CDTCMs

Design and construction of the expression plasmids for the bifunctional fusion enzymes using the pKTCTET backbone (Roderer *et al.* 2014) was carried out by Christian Stocker, at the Peter Kast lab (ETH, Zürich, Switzerland). This includes construct with and without the N-terminal leader sequences, as well as the active site KO and single domain variants of *AfCMCDT and *JbCDTCM. Gene expression was carried out in *E. coli* KA29 cells (Sasso *et al.* 2005) by Christian Stocker at the same lab the details of gene expression and purification are summarized in Table 7 and 8, and explained in detail in **Manuscript I**.

6.1.2 Untagged and His₆-tagged *PaeCDT and *PaeCM

Design and construction of pKTCTET expression plasmids for untagged and His₆-tagged *PaeCM and *PaeCDT were carried out by Kathrin Würth-Roderer and Luca Bressan, respectively, at the Peter Kast lab (ETH, Zürich, Switzerland). Purified *PaeCM-His₆ was obtained from Kathrin-Würth-Roderer. Protein expression and production of untagged and *PaeCM-His₆ was carried out in *E. coli* KA12/pT7POLTS (Roderer *et al.* 2014). The untagged and *PaeCDT-His₆ were expressed and produced in *E. coli* KA12 (Roderer *et al.* 2014) and *E. coli* NEBExpress I^q (New England Biolabs), respectively, both carrying the pT7POLTS auxiliary vector. Production and purification steps are summarized in Table 7 and 8 and reported in detail in **Manuscript II**.

Table 7. Production of monofunctional and bifunctional chorisimate mutases

	*CMCDTs *CDTCMs	*PaeCM	*PaeCM-His6	*PaeCDT	*PaeCDT-His6	MtCM ^{V55D}	MtCM ^{T52P}	MtCM ^{wt}
Vector	pKTCTET	pKTCTET	pKTCTET	pKTCTET	pKTCTET	pKTCTET	pKTCTET	pKTCTET
Expression cell	<i>E. coli</i> KA29	<i>E. coli</i> KA12/ + pT7POLTS	<i>E. coli</i> KA12/ + pT7POLTS	<i>E. coli</i> KA12/ + pT7POLTS	<i>E. coli</i> NEB Express ^{IV} + pT7POLTS	<i>E. coli</i> KA13	<i>E. coli</i> KA13	<i>E. coli</i> KA13
Media	LB + Amp ¹⁰⁰ Kan ⁵⁰	LB + Amp ¹⁵⁰ Cam ³⁰	LB + Amp ¹⁵⁰ Cam ³⁰	LB + Amp ¹⁰⁰ Cam ²⁵	LB + Amp ¹⁰⁰ Cam ²⁵	LB + Amp ¹⁰⁰	LB + Amp ¹⁵⁰	LB + Amp ¹⁵⁰
Growth temp (°C)	20	30	30	30	30	30	30	30
Growth duration	2.5 days	OD ₆₀₀ =0.6-08	OD ₆₀₀ =0.6	OD ₆₀₀ =0.6	OD ₆₀₀ =0.6	OD ₆₀₀ =0.5	OD ₆₀₀ =0.5	OD ₆₀₀ =0.3-0.5
Induction	basal	Tet ²	Tet ¹	Tet ¹	Tet ¹	IPTG ^{0.5}	IPTG ^{0.5}	IPTG ^{0.5}
Expression duration	24 h	overnight	overnight	overnight	overnight	overnight	overnight	overnight
Expression temp (°C)	20	30	20	20	20	30	30	25
Harvest	4000 g, 15 min, 4°C	4000g, 20 min, 4°C	4000g, 20 min, 4°C	4000g, 20 min, 4°C	5000g, 20 min, 4°C	6500 g, 20 min, 4°C	6500 g, 20 min, 4°C	4000 g, 20 min, 4°C
Lysis buffer	20 mM Tris-HCl, pH 8, 150 mM NaCl + 1 mg/mL lysozyme	50 mM K-phosphate, pH5.0 7.2, 150 mM NaCl + 1 mg/mL polymyxin mm NaCl	50 mM K-phosphate, pH5.0 7.2, 150 mM NaCl + 1 mg/mL polymyxin mm NaCl	50 mM K-phosphate, pH2.0 7.5, 300 mM NaCl	50 mM K-phosphate, pH2.0 7.5, 300 mM NaCl	50 mM Na-acetic acid, pH 5.2+ DNase (I), 150 µM PMSF, protease inhibitor	50 mM Na-MES, pH 6.5 + DNase (I), 150 µM PMSF, protease inhibitor	20 mM piperazine-HCl 9.8+ protease inhibitor tablet
Lysis	Sonicator, 3 x (2 min on and 2 min off)	Sonicator 8 min (30-s on and 45-s off)	Sonicator 8 min (30-s on and 45-s off)	Sonicator, 4 x (30-s on and 60-s off)	Sonicator, 4 x (30-s on and 60-s off)	BeadBeater, 4 x (30-s on and 60-s pulse off)	BeadBeater, 4 x (30-s on and 60-s pulse off)	Sonicator, 3 x 30 sec (3-s on and 7- s off)
Lysate clarification	100% amplitude 20000 g, 20 min, 4°C	50% amplitude 20000 g, 15 min, 4°C	50% amplitude 30000 g, 15 min, 4°C	50% amplitude 20000 g, 20 min, 4°C	50% amplitude 20000 g, 20 min, 4°C	48000g, 30 min, 4°C	48000g, 30 min, 4°C	20% amplitude 10500g, 30 min, 4°C
Personally expressed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Amp¹⁵⁰: sodium ampicillin 150 µg/mL, Amp¹⁰⁰: sodium ampicillin 100 µg/mL

Kan⁵⁰: kanamycin 50 µg/mL

Cam²⁵: chloramphenicol 25 µg/mL, Cam³⁰: chloramphenicol 30 µg/mL

Tet¹: tetracycline 1 µg/mL; Tet²: 2 µg/mL

IPTG^{0.5}: 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG)

Table 8. Purification of monofunctional and bifunctional chorismate mutases

	*CMCDT/ *CDTCMs	*PaeCM	*PaeCM-His ₆	*PaeCDT	*PaeCDT-His ₆	MiCM ^{SSD}	MiCM ^{TSP}	MiCM ^W
Step 1	AC	CEX	AC	AEX	AC	CEX	CEX	AEX
Technique								
Column	gravity-flow Ni-NTA MonoS HR 10/10	gravity-flow Ni-NTA Mono Q HR 10/10	gravity-flow Ni-NTA Mono Q HR 10/10	gravity-flow Ni-NTA Mono Q HR 10/10	gravity-flow Ni-NTA HiTrap XL SP	HiTrap XL SP	HiTrap XL SP	HiTrap XL Q 5ml
Elution buffer	50 mM Tris-HCl, pH 8, +250-500 mM imidazole	20 mM Na-MES, pH 6.5 + 1 M NaCl	20 mM Tris-HCl, pH 8.0 + 1 M NaCl	20 mM K-phosphate, pH 7.5, 300 mM NaClacid, + 250 mM imidazole	50 mM Na-acetic acid, pH 5.2 + 0.5 M NaCl	50 mM Na-MES, pH 6.5 + 0.5 M NaCl	20 mM Na-MES, pH 7.5, 20 mM BTP, 150 mM NaCl	20 mM piperazine-HCl pH 9.8 + 0.5 M NaCl
Technique	Dialysis	SEC	Dialysis	SEC	SEC	SEC	SEC	SEC
Column		Superdex 75 Increase 10/300 GL		Superdex 75 Increase 10/300 GL	Superdex 200 Increase 10/300 GL	Superdex 75 10/300 GL	Superdex 75 10/300 GL	Superdex 75 Inc 10/300 GL
Buffer	20 mM Tris-HCl, pH 8, NaCl	20 mM K-phosphate, pH 7.2, 150 mM NaCl	20 mM K-phosphate, pH 6.5, 150 mM NaCl	20 mM K-phosphate, pH 7.2, 150 mM NaCl	20 mM K-phosphate, pH 7.5, 150 mM NaCl	20 mM BTP, pH 7.5, 150 mM NaCl	20 mM BTP, pH 7.5, 150 mM NaCl	20 mM BTP, pH 7.5, 150 mM NaCl
Personally purified	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

AC: Affinity chromatography

AEX: anion-exchange chromatography

CEX: cation-exchange chromatography

SEC: size-exclusion chromatography

6.1.3 *Mycobacterium tuberculosis* CM wild type expression and purification

Glycerol stocks of *E. coli* KA13 transformed with the vector pKTCMM-H carrying the gene for wild type MtCM were inoculated into 50 mL LB media containing 150 $\mu\text{g/mL}$ sodium ampicillin (Amp¹⁵⁰). The 50 mL pre-cultures were incubated overnight at 37°C under vigorous shaking, before being transferred into 500 mL of LB medium containing Amp¹⁵⁰. Production cultures were further incubated at 30°C in a shaker until OD₆₀₀ reached 0.3-0.5. Gene expression was induced overnight at 25 °C, by addition of IPTG to a final concentration of 0.5 mM. Cells were harvested by centrifugation (4000g, 4°C, 20 min) and resuspended in 20 mM piperazine-HCl pH 9.8 containing cOmplete protease inhibitor cocktail (Roche). Cells were lysed by ultrasonication, using a 1/4" (6mm) probe with 20% amplitude, three cycles of 30" (3" on, 7" off). Protein purification was performed by anion-exchange on a HiTRAP Q XL column (GE Life Sciences) with a 0-100% gradient elution of low (20 mM piperazine-HCl pH 9.8) to high salt buffer (20 mM piperazine-HCl pH 9.8 containing 0.5 M NaCl). Fractions containing MtCM were further purified by size-exclusion chromatography (SEC) on a Superdex 75 Increase 10/300 GL (GE Life Sciences) gel filtration column. Fractions containing pure MtCM were collected and concentrated to 2.84 mg/mL using a VIVASPIN 20 PES membrane 5 000 MWCO concentrator.

6.2 *Differential scanning fluorimetry (DSF)*:

6.2.1 *Introduction to the method*

Differential scanning fluorimetry (DSF)/Thermofluor is a fluorescence-based thermal shift assay, which quantifies thermally induced protein denaturation (Zhang and Monsma 2010, Senisterra *et al.* 2012). The denatured protein is tracked through the use of a fluorescent probes, such as a fluorescent dye (*e.g.* Sypro Orange), or relying on the natural fluorescence of aromatic amino acids present in the protein sequence (Senisterra *et al.* 2012). Applying a temperature gradient, proteins unfold; when using a fluorescent solvatochromic dye, the percentage of unfolded protein can be estimated by a fluorescence signal emitted by the dye interacting with exposed hydrophobic amino acids from the protein core (Senisterra *et al.* 2012). The melting temperature of the protein (T_m) is defined as the temperature at which 50% of the protein is unfolded (Vieille and Zeikus 2001). Binding of ligands can affect the

protein stability, shifting the T_m , in both directions: positive, stabilization, or negative, destabilization (Cimpmperman *et al.* 2008, Kabir *et al.* 2016).

6.2.2 Experimental procedure

DSF experiments were run in 96-well plates on a LightCycler 480 Real-Time PCR machine (Roche), using 25 μ l samples and the Sypro Orange dye (Sigma Aldrich). Fluorescent signal was recorded at 580 nm, varying the temperature between 20 and 95 °C and using an excitation wavelength of 465 nm. The data were analyzed using a modified version of the Microsoft Excel-based DSF Analysis tool developed by Frank Niesen (Structural Genomics Consortium, Oxford, United Kingdom). The melting temperature was calculated by non-linear fitting the data with a Boltzmann sigmoid function through the software GraphPad Prism, version 5 (GraphPad Software Inc., La Jolla, CA, USA). Before the stability actual experiments, preliminary screenings to establish the optimal experimental conditions were performed for all targets, using different combinations of protein concentrations (0.05-0.5 mg/ml) and Sypro Orange dilutions (Sigma Aldrich, 100-1000-fold diluted).

6.2.3 Mycobacterium tuberculosis chorismate mutase

Thermal stability of non-tagged MtCM variants was assessed by differential scanning fluorimetry using a protein concentration of 0.5 mg/ml (wild-type: MtCM^{WT}, T52P and V55D single mutants: MtCM^{T52P} and MtCM^{V55D}, super-active mutant: MtCM^V) or 0.4 mg/ml (T52P+V55D double mutant: MtCM^{DP}) in presence of Sypro Orange diluted to a final 1:100 volume ratio, 100 mM bicine pH 9, 150 mM NaCl. For each protein, the melting temperature was calculated from three replicates.

Table 9. Experimental conditions for DSF experiments on MtCM variants

	MtCM variants				
	MtCM ^{WT} wild-type	MtCM ^{T52P} T52P	MtCM ^{V55D} V55D	MtCM ^V superactive	MtCM ^{3P3} T52P+V55D
Protein [C] (mg/ml)	0.5 mg/ml	0.5 mg/ml	0.5 mg/ml	0.5 mg/ml	0.4 mg/ml
Sypro Orange dilution			1:100		
Buffer			100 mM bicine pH 9		
NaCl			150 mM NaCl		

6.2.4 Bifunctional *CDTCM and *CMCDT

The best buffer for DSF experiments on bifunctional *CDTCM and *CMCDT variants was selected by pre-screening buffers from the Rubic Additive kit (Molecular Dimensions Ltd.). The buffer pre-screening experiment was carried out using 0.5 mg/mL bifunctional proteins resuspended in their storage buffer and Sypro Orange diluted to a final 1:200 volume ratio. The choice fell on a buffer formulation containing 100 mM Tris-HCl, pH 8.5 and 150 mM NaCl, which gave a clear transition in the melting curve without being too close to the pI of any of the bifunctional enzymes (5.8-7.1). Subsequent protein stability experiments were carried out at final protein concentration of 0.5 mg/mL, in a 100 mM Tris-HCl, pH 8.5, 150 mM NaCl buffer, 1:200 Sypro Orange dilution. The melting temperature was calculated from five replicates .

Table 10. Experimental conditions for DSF experiments on bifunctional enzymes

	Bifunctional enzymes				
	*AfCMCDT	*ScCMCDT	*JbCDTCM	*DsCDTCM	*MpCDTCM
	Buffer pre-screening				
Protein [C] (mg/ml)	0.5 mg/ml				
Sypro Orange dilution	1:200				
Rubic buffers	100 mM (various)				
NaCl	150 mM NaCl				
	Stability screening				
Protein [C] (mg/ml)	0.5 mg/ml				
Sypro Orange dilution	1:100				
Buffer	100 mM Tris-HCl pH 8.5				
NaCl	150 mM NaCl				

6.3 Protein X-ray crystallography

6.3.1 Overview of the method

X-ray crystallography is the most widely used technique to determine the three-dimensional structures of proteins (Wilson 2018), as it yielded 85% of protein structures deposited on the Protein Data Bank (PDB) (Berman *et al.* 2002). The technique utilizes high intensity X-ray and diffraction-quality protein crystal. The wavelength of X-rays (~1Å) is compatible with the interatomic spacing in molecules, so diffracted X-rays provide structural information on molecules at the atomic level (Parker 2003, Wilson 2018).

The electrons of atoms scatter the X-ray beam. The scattered waves with the same phase interact with each other constructively, resulting in diffracted radiation. Out-of-phase scattered waves give destructive interference, and low to null diffracted X-ray intensity (Parker 2003, Wilson 2018) (Figure 32). A protein crystal is a regular array of identically oriented individual molecules; when hit by an X-ray beam, the repeated scattering from molecules sitting at regular intervals on the crystal lattice amplifies the scattering signal (McPherson and Gavira 2004, Gawas *et al.* 2019). Constructive interference is produced by diffracted X-rays which satisfy the Bragg condition: $2 \cdot d \cdot \sin\theta = n \cdot \lambda$, where n is an integer, λ is the wavelength of the incoming radiation, θ is the incident beam angle, and d is the spacing for a defined family of ‘crystal planes’, a theoretical construction that simplify the understanding of diffraction by a crystal lattice (Figure 32). The scattered X-rays are recorded on a detector, where they form a diffraction figure, or pattern (Figure 33) (Parker 2003, Gawas *et al.* 2019).

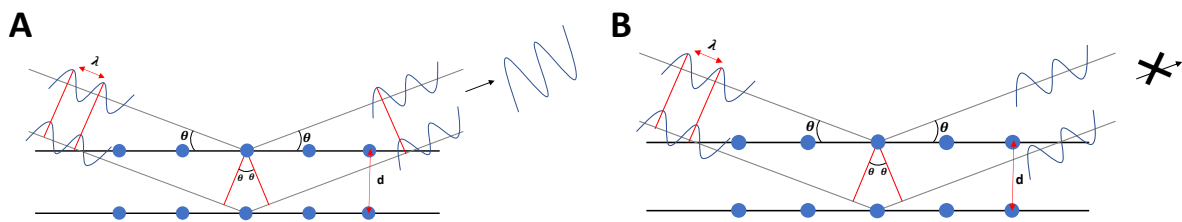


Figure 32. Bragg diffraction. A. Constructive interference, two waves in phase interfere constructively. B. Destructive interference, two waves out of phase cancel each other.

The relation between the electron density distribution within the crystal unit cell and the diffraction figure is described by the Fourier transform (Rupp 2013) (Figure 33). The amplitudes of the Fourier terms correspond to the amplitude of the diffracted X-rays, and they can be calculated from the intensity of diffraction spots; their phase is lost during the measurement process, and thus is unknown (Rupp 2013). Without a complete set of amplitudes and phases it is not possible to perform a Fourier synthesis (the inverse of a Fourier transform), thus it is not possible to derive the shape of the electron-density function: this is known as the ‘phase problem’ (Parker 2003, Taylor 2003). Different methods were developed to solve the phase problem such as MAD, SAD, and molecular replacement (Taylor 2003). The latter uses the phase information calculated by placing in the unit cell a homologous protein with known structure, matching the orientation of the original one (Parker 2003). Once an electron density map is obtained, a model is built by fitting the

coordinates of the amino acids from the protein sequence into the density. The model is then refined iteratively, minimizing the difference between the calculated scattering figure from the model and the experimentally-recorded intensities (Parker 2003, Rupp 2013).

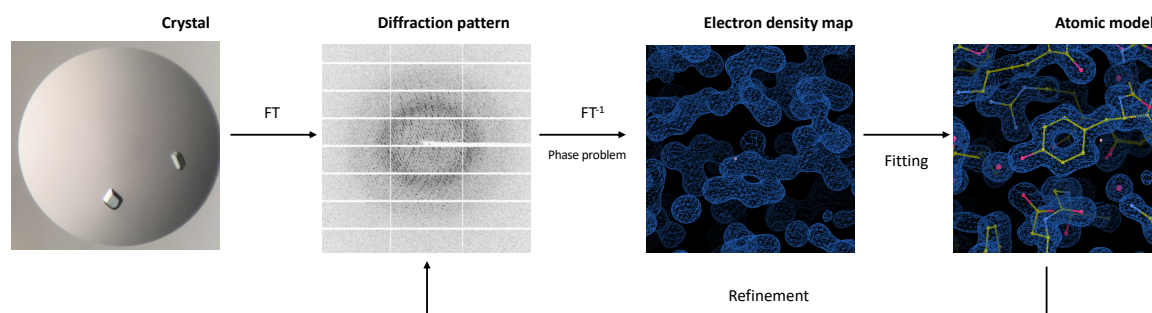


Figure 33. Flow chart showing the main steps in protein X-ray crystallography. From crystal to atomic model.

6.3.2 Protein crystallization

6.3.2.1 Introduction to the method

Protein crystallization is a mandatory step and the biggest bottleneck in X-ray crystallography. A protein crystal is a regular array of individual molecules packed into a crystal lattice. Crystallization is a form of ordered precipitation, a physicochemical process with a phase transition from a disordered state (the protein in solution) to a highly ordered solid state (the protein crystal). The number of variables influencing crystallization is high and include temperature, pH and the presence in solution of a variety of chemicals, one of them acting as a precipitant (McPherson and Gavira 2004). Protein crystallization can be illustrated through the phase diagram for a simplified system formed by a solution of the protein and the precipitant (Figure 33). The diagram is divided into three areas, each describing a different state of the system in terms of the relative concentration of protein and precipitant: undersaturated, supersaturated, and amorphous precipitation. The supersaturated zone is further divided into two: metastable and nucleation. Protein crystallization requires the system to reach supersaturation conditions. When the system reaches the nucleation zone, it forms crystal nuclei (microscopic, completely ordered assemblies), which further grow in size in the metastable zone.

Supersaturation and the formation/growth of crystal nuclei can be promoted by manipulating a wide array of physicochemical variables. Protein crystallization is affected by sample purity, concentration, pH, additives such as cofactors, ligands, ionic strength of the solution, type of precipitant, intrinsic flexibility of the sample, temperatures and the crystallization experiment geometry (McPherson and Gavira 2004). However, whether and how the crystallization occurs (and within which time frame) is rather unpredictable; thus, obtaining a protein crystal is a trial-and-error process, which usually requires testing different combination of reagents, buffers and pH at different temperatures.

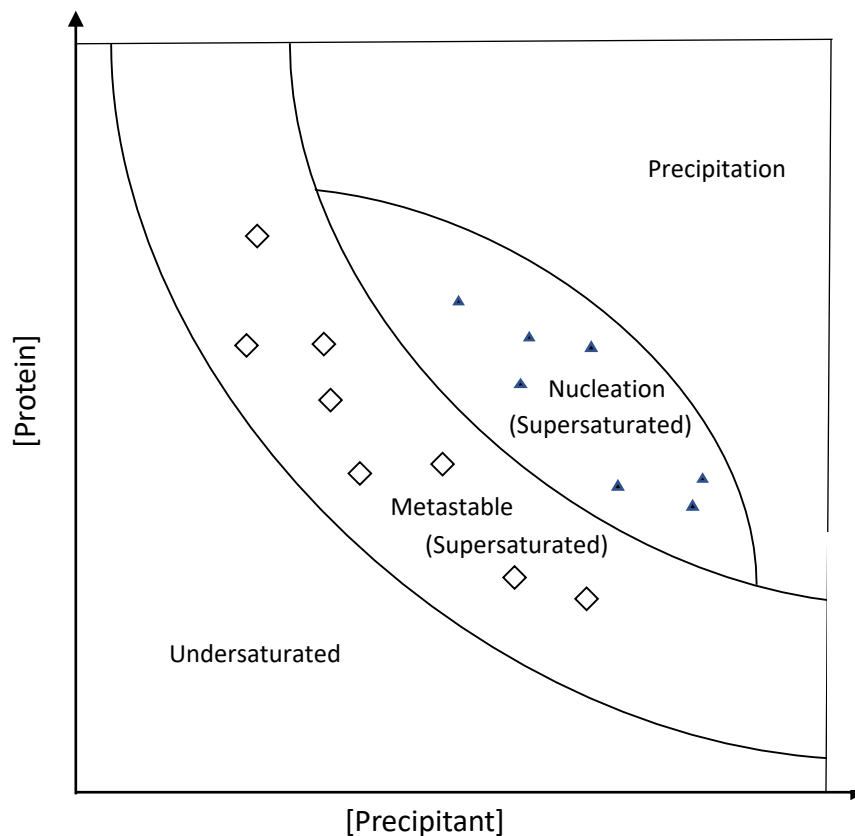


Figure 34. Phase diagram of protein crystallization. Crystals form when the protein is supersaturated. Nucleation occurs in nucleation zone and crystals grow in metastable zone.

6.3.2.2 Crystallization experiments

Diffraction quality crystals were obtained by either hanging-drop or sitting-drop vapor diffusion experiments. For the sitting-drop setup, we used 96-well UVXPO 2 Lens or 3 Lens crystallization plates (SWISSCI Laboratory Solutions) with 50 μL and 40 μL reservoir

volumes, respectively. The volume of the crystallization drops varied between 400 and 600 nL; protein stocks were mixed in a 1:1 or 1:2 volume ratio with crystallization solutions. Hanging-drop vapor diffusion experiments were performed using 24-well tissue culture plates (TPP Techno Plastic Products AG) and 18 mm circular siliconized cover slides (Molecular Dimensions Ltd.), on which 1 μ L crystallization drops consisting of 0.5 μ L protein and 0.5 μ L reservoir solution were equilibrated against 1 mL reservoirs filled with crystallization solutions. The hanging drop experiments were pipetted manually and stored in a temperature-controlled room. The crystallization plate for sitting drop experiments were prepared by a Oryx4 protein drop-setting robot (Douglas Instruments). The plates were stored and imaged in a RockImager RI-182 crystallization plate hotel (Formulatrix).

Initial crystallization condition screenings were carried out using commercial formulations: *PACT premier*, *JCSG+*, *PGA*, and *Morpheus* all sold by Molecular Dimensions Ltd. The crystallization conditions that yield three-dimensional crystals, either microcrystals, needles or plates, are defined as 'hits'. When needed, hits were optimized by varying the concentration and type of chemical components of the crystallization mixture to obtain diffraction-quality crystals.

In this thesis, I presented the crystal structures of three CM/CDT fusion proteins (*AfCMCDT, *JbCDTCM, *DsCDTCM) with and without ligands, two monofunctional periplasmic enzymes from *P. aeruginosa* (*PaeCDT and *PaeCM) and the intracellular chorismate mutase from *M. tuberculosis* (MtCM). Bifunctional enzymes and *PaeCM were never crystallized before, while the crystal structure of *PaeCDT and MtCM were already known (*PaeCDT: PDB ID: 5HPQ, (Clifton *et al.* 2018), MtCM: PDB ID: 2VKL, (Sasso *et al.* 2009) and 2QBV (Kim *et al.* 2008))

The latter was used for the fragment screening described in the next chapter. A summary of all the targets crystallized and their crystallization conditions is reported in Table 11.

6.3.2.3 Ligand-bound complexes crystallization

The ligand-bound structures were obtained by either soaking the crystals in solutions containing the ligand or co-crystallizing the protein in its presence. In case of the latter, the ligand was either dissolved in the mother liquor or added to it in solid form. The transition state analog (TSA)-bound *AfCMCDT structure was obtained by co-crystallization in presence of TSA. Few flakes of TSA were added to the protein solution and incubated on ice

for 30 min. Insoluble TSA was removed by spinning down at 21 000 g for 15 min at 4°C. To prevent diffusion of the ligand out of crystal during cryoprotection, the ligands were also added to the cryo-protectant drops.

We also got TSA- and sodium phenylpyruvate (Na-PPY)-bound structures from soaking. As TSA has a very high binding affinity towards CMs, adding a few flakes of TSA to the mother liquor drop with the crystals was enough to get TSA-bound protein.

Table 11. Crystallization experiments

Samples	Conc ¹ (mg/mL)	Crystallization plate	Crystallization condition	V _{res} ² (μL)	V _{drop} ³ (μL)	P:R ⁴	Ligand soak
JbCDTCM_MES	3.5	96-well SWISSCI UVXPO 2 Lens	30% w/v PEG 550M_20K 0.09 M NPS 0.1 M MES/imidazole pH 6.5	50	0.4	1:1	-
JbCDTCM_MES/PPY	3.5	96-well SWISSCI UVXPO 2 Lens	30% w/v PEG500MME_20K 0.09 M NPS 0.05 M MB1 pH 6.5	1000	2	1:1	1:100 PPY ⁵
JbCDTCM_PPY	3.5	24-well tissue culture plate	30% w/v PEG500M_20K 0.09 M NPS 0.1 M imidazole-HCl pH 6.3	1000	2	1:1	1:50 PPY
JbCDTCM_PPY_TSA	3.5	24-well tissue culture plate	28% w/v PEG500M_20K 0.07 M NPS 0.1 M imidazole-HCl + 0.08 M PPY, pH 6.5	1000	2	1:1	1:50 PPY Few flakes TSA
DsCDTCM	5.6	96-well SWISSCI UVXPO 2 Lens	3 M NaCl 0.1 M Na-Bis-Tris pH 5.5	50	0.4	1:1	-
AfCMCDT_ACT	5	96-well SWISSCI UVXPO 3 Lens	20% w/v PEG 3350 0.2 M Na-acetate 0.1 M BTP pH7.5	40	0.6	1:3	-
AfCMCDT_TSA (co-crystallized)	5	24-well tissue culture plate	20% PEG 3350 0.2 M KSCN 0.1 M BTP pH 8.5	1000	2	1:1	1:100 PPY Few flakes of TSA
PaeCDT	11	96-well SWISSCI UVXPO 3 Lens	0.28 M Na- malonate 20 w/v % PEG 3350 0.1 M Na-acetate pH 4.7	40	0.6	1:1	-
PaeCM	8		20% w/v PEG3350 0.2 M NH ₄ -citrate dibasic	50	0.6	1:3	-

¹ Concentration

² Volume of the reservoir

³ Volume of the crystallization drop

⁴ Protein to reservoir volume ratio

⁵ Excess molar concentration of PPY compared to the protein

Stock solutions of Na-PPY were prepared in milliQ water in molar concentration which is 50 or 100 -fold excessive compared to the protein concentration. Na-PPY soaking drops were prepared by mixing stock solutions of Na-PPY with a reservoir solution in a 1:1 volume ratio. Na-PPY soaking drops were equilibrated overnight against the reservoir in their corresponding crystallization wells.

6.3.3 *Data collection*

Crystals were cryoprotected by dipping them into a drop of reservoir solution diluted either with ethylene glycol or glycerol, to a final concentration of 20-25% v/v of ethylene glycol or glycerol. For the Na-PPY co-crystals, the crystallization solution was complemented with

cryoprotectant, then diluted with the Na-PPY stock solutions to a 1:1 volume ratio. The cryoprotectant+ligand drops were spotted next to the crystallization drops and were equilibrated against the reservoir solutions overnight before crystal soaking, to yield a final 1:50 or 1:100 protein:Na-PPY molar ratio. Soaked crystals were then flash-cooled in liquid nitrogen and stored until data collection.

Data collection was performed by the rotation method under cryocooling (100K), at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) and at the MAX IV laboratory synchrotron (Lund, Sweden). The X-ray wavelengths used in the experiments varied between 0.873 and 0.984 Å at different beamlines (Table 12). Either 180° or 360° total oscillation data sets were collected, with oscillation ranges varying between 0.05° and 0.2°. All the beamlines were equipped with pixel detectors.

Table 12. Data collection parameters

Crystals	Cryoprotectant	Beamline	Detector	Total oscillation	Oscillation range	X-ray wavelength, Å	Detector edge, Å
JbCDTCM_MES	-	MAX IV BioMax	Eiger X 16M	360°	0.1°	0.976	1.47
JbCDTCM_MES/PPY	20 % glycerol + 1:25 PPY	MAX IV BioMax	Eiger X 16M	360°	0.1	0.976	
JbCDTCM_PPY	20 % glycerol + 1:12 PPY	ESRF ID30A-1	PILATUS3 2M	180°	0.2°	0.965	1.71
JbCDTCM_PPY_TSA	25 % glycerol + 1:12 PPY + few flakes TSA	ESRF ID30B	EIGER2 X 9M	165°	0.05°	0.976	1.51
DsCDTCM	20 % ethylene glycol	MAX IV BioMax	EIGER 16M	360°	0.1°	0.984	1.78
AfCMCDT	20 % glycerol	ESRF ID30A-1	PILATUS3 2M	360°	0.2°	0.965	1.03
AfCMCDT_TSA	25 % glycerol + 1:25 PPY + few flakes TSA	ESRF ID30B	EIGER2 X 9M	180°	0.1°	0.976	2.35
PaeCDT	20 % glycerol	ESRF ID23-2	PILATUS3 X 2M	360°	0.1°	0.873	1.99
PaeCM	20 % glycerol	ESRF ID23-2	PILATUS3 X 2M	360°	0.1°	0.873	1.11

6.3.4 *Data processing and the structure determination*

Diffraction images were processed and scaled by various automated pipelines (*autoPROC*, *XDSAPP*, and *GrenADES*) at the synchrotrons (Krug *et al.* 2012, Monaco *et al.* 2013). Data merging was performed with *AIMLESS* from the CPP4 software package (Evans 2006, Evans and Murshudov 2013). When needed, anisotropic data sets were corrected using the STARANISO server (Vonrhein *et al.* 2018). The structures were solved by molecular replacement using *Phaser* from the CCP4 software package (McCoy *et al.* 2007). The search models were found running a BLASTP search of the target protein sequences restricted to proteins with a PDB entry (Johnson *et al.* 2008). Search models used to perform molecular replacement on different data sets are listed in the Table 13. The structures underwent alternate rounds of maximum-likelihood-restrained refinement with

REFMAC5 (Kovalevskiy *et al.* 2018) and real-space refinement using *Coot* (Emsley *et al.* 2010). The NCS (non-crystallographic symmetry) refinement was applied for *DsCDTCM and *PaeCDT datasets. As the quality of the electron density maps improved, water molecules were added to the model and, finally, alternative side-chain conformations and ligands, based on σ_A -weighted *F_o-F_c* difference map peaks and chemical surroundings. As a final step for all structures, occupancy refinement was carried out with *phenix.refine*, a tool of the *PHENIX* software suite (Afonine *et al.* 2012).

Table 13. Data processing, structure determination, and refinement software matrix

Crystals	Integration and scaling	Merging and truncation	Anisotropy correction	Phasing	MR model	Refinement
JbCDTCM_MES	<i>XDS (autoPROC)</i>	<i>AIMLESS (Staraniso)</i>	<i>Staraniso</i>	<i>PHASER</i>	BtCM (PDB ID: 6CNZ) PaeCDT (PDB ID: 5HPQ)	<i>REFMAC</i> <i>COOT</i> <i>Phenix.refine</i>
JbCDTCM_PPY	<i>XDS (XDSAPP)</i>	<i>AIMLESS</i>	-	<i>PHASER</i>	JbCDTCM_MES	<i>REFMAC</i> <i>COOT</i> <i>Phenix.refine</i>
JbCDTCM_PPY_TSA	<i>XDS (autoPROC)</i>	<i>AIMLESS</i>	-	<i>PHASER</i>	JbCDTCM_MES	<i>REFMAC</i> <i>COOT</i> <i>Phenix.refine</i>
DsCDTCM	<i>XDS (autoPROC)</i>	<i>AIMLESS (Staraniso)</i>	<i>Staraniso</i>	<i>PHASER</i>	JbCDTCM_MES	<i>REFMAC</i> <i>COOT</i> <i>Phenix.refine</i>
AfCMCDT	<i>XDS (autoPROC)</i>	<i>AIMLESS (Staraniso)</i>	<i>Staraniso</i>	<i>PHASER</i>	BtCM (PDB ID: 6CNZ) PaeCDT (PDB ID: 5HPQ)	<i>REFMAC</i> <i>COOT</i> <i>Phenix.refine</i>
AfCMCDT_TSA	<i>XDS (GrenADES)</i>	<i>AIMLESS</i>	-	<i>PHASER</i>	BtCM (PDB ID: 6CNZ) PaeCDT (PDB ID: 5HPQ)	<i>REFMAC</i> <i>COOT</i> <i>Phenix.refine</i>
PaeCDT	<i>GrenADES</i>	<i>AIMLESS</i>	-	<i>PHASER</i>	PaeCDT (PDB ID: 5HPQ)	<i>REFMAC</i> <i>COOT</i> <i>Phenix.refine</i>
PaeCM	<i>XDS (XDSAPP)</i>	<i>AIMLESS</i>	-	<i>PHASER MR</i>	BtCM PDB ID: 6CNZ	<i>REFMAC</i> <i>COOT</i> <i>Phenix.refine</i>
MtCMT52P	<i>XDS</i>	<i>AIMLESS (Staraniso)</i>	<i>Staraniso</i>		PDB ID: 5MPV	<i>REFMAC</i> <i>COOT</i> <i>Phenix.refine</i>
MtCMV55D	<i>XDS</i>	<i>AIMLESS (Staraniso)</i>	<i>Staraniso</i>	<i>PHASER MR</i>	MtCM PDB ID: 2W1A	<i>REFMAC</i> <i>COOT</i> <i>Phenix.refine</i>

6.4 *Fragment-based screening*

6.4.1 *Introduction to the method*

Fragment-based screening, or fragment screening, is a method to identify small molecules ('fragments') binding weakly to the active site and/or to allosteric sites of a protein target, with the aim of subsequently developing them into drug candidates (Rees *et al.* 2004). Fragments are often designed to follow the "rule of three": less than 300 Da molecular weight, less than three hydrogen-bond donors/acceptors, and whose clogP is less than three (Congreve *et al.* 2003, Kirsch *et al.* 2019). The small size of the fragments allows for high efficiency of sampling, giving a high number of hits and a high degree of freedom to further evolve the fragment while limiting unfavorable interactions (Congreve *et al.* 2003).

Identifying *hits*, small molecules which specifically bind to the target, is quite challenging as the fragments have a low binding affinity (micromolar to millimolar range) (Schultes 2010, Silvestre *et al.* 2013). A number of biophysical and biochemical methods is employed to select hits. Some of them produce lots of false hits, due to the use of relatively high concentrations of fragments, which is needed to improve the sensitivity of the methods for lower affinity molecules (Silvestre *et al.* 2013, Kirsch *et al.* 2019). Therefore, it is important to use orthogonal methods, such as surface plasmon resonance (SPR), thermal shift assay/differential scanning fluorimetry (DSF), isothermal titration calorimetry (ITC), NMR, and X-ray crystallography, for the detector of bound fragment (Kirsch *et al.* 2019).

Hit molecules which are further developed ('grown') from fragments are called 'drug-like molecules'. Subsequent optimization steps focus on increasing their affinity and specificity by employing different growing strategies (Silvestre *et al.* 2013, Kirsch *et al.* 2019, Li 2020). In later stages of drug development, physico-chemical properties such as solubility and surface activity, are considered (Schultes 2010, Silvestre *et al.* 2013, Kirsch *et al.* 2019).

6.4.2 *Fragment-based screening using X-ray crystallography*

X-ray crystallography is a powerful method for fragment-based screening, providing direct structural information on the protein-fragment complex (Congreve *et al.* 2003, Schiebel *et al.* 2016). Compared to other methods, such as DSF, SPR and ITC, X-ray crystallography is much more sensitive for low affinity molecules, thus making it the best choice for primary hit screening (Blundell *et al.* 2002, Carr and Jhoti 2002). The main challenge is to produce

suitable crystals for fragment screening. Ideally, crystals for fragment screening should be robust, of large size (>100 μm) (Collins *et al.* 2018), diffracting better than 2 Å resolution and, most importantly, easily reproducible (Chilingaryan *et al.* 2012, Collins *et al.* 2018).

The general workflow of X-ray crystallography-based fragment screening includes the crystallization of the target protein, the preparation of a protein-fragment co-crystal (either by soaking or co-crystallization), data collection, data processing and analysis of bound ligands (Patel *et al.* 2014, Lima *et al.* 2020). Modern high-energy synchrotron beamlines enable screening of many protein crystals in a relatively short time and provide automated pipelines for data processing and ligand fitting. Thus, screening large number of fragment libraries by X-ray crystallography is not as time-consuming as it used to be in the past (Schiebel *et al.* 2016).

6.4.2.1 Crystallization and soaking

The bottleneck of screening ligands by X-ray crystallography is establishing a crystallization system that yields highly reproducible, large and well-diffracting crystals (>2Å). Furthermore, while protein often crystallize under multiple conditions, many of the crystallization cocktails would not be suitable for fragment screening, due to the presence of chemicals binding to the active site and competing with the fragment molecules.

Initial crystallization trials on MtCM were carried out at 20°C by the sitting drop vapour diffusion geometry in 96-well UVXPO 2 Lens plates (SWISSCI Laboratory Solutions), using the commercial crystallization formulations *PACT premier* and *JCSG+* (Molecular Dimensions Ltd). Multiple conditions yielded 3D crystals. Crystallization conditions containing compounds that are able to bind to the active site of MtCM, for example citrate and malonate, were not tested further. Diffraction quality and stability of crystals obtained from other crystallization conditions were tested over the course of several synchrotron trips (Figure 35).

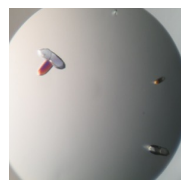
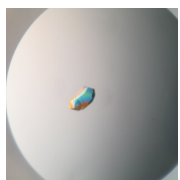



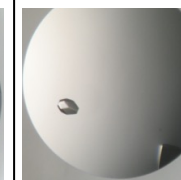

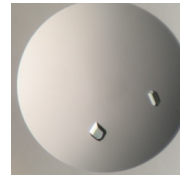
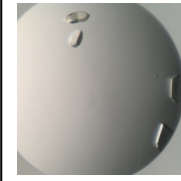
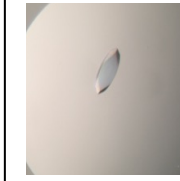
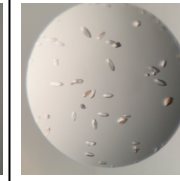
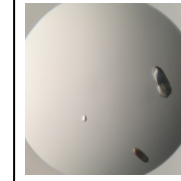
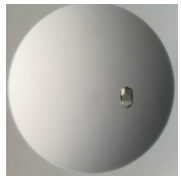
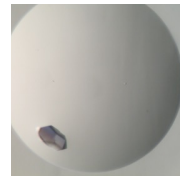
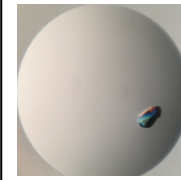
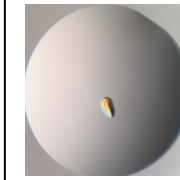
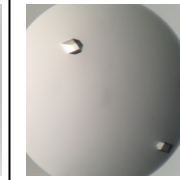
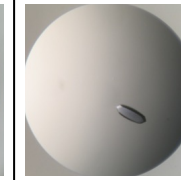
					
<i>JCSG+</i> A4 0.1 M SPG buffer 7.0 25% w/v PEG 1500	<i>JCSG+</i> A5 0.1 M SPG buffer 8.0 25% w/v PEG 1500	<i>JCSG+</i> B6 0.1 M MIB buffer 9.0 25% w/v PEG 1500	<i>JCSG+</i> D2 0.2 M Magnesium chloride x 7H ₂ O 0.1 M Sodium HEPES 7.5 30% v/v PEG 400	<i>JCSG+</i> D3 0.2 M Sodium chloride 0.1 M Sodium/potassium phosphate 6.2 50% v/v PEG 200	<i>JCSG+</i> E12 0.2 M Na-malonate 20% w/v PEG 3350
					
<i>JCSG+</i> F11 0.2 M Na-citrate 0.1 M Bis Tris propane 6.5 20% w/v PEG 3350	<i>JCSG+</i> G8 0.2 M Na-sulfate 0.1 M Bis Tris propane 7.5 20% w/v PEG 3350	<i>JCSG+</i> G11 0.2 M Na-citrate 0.1 M Bis Tris propane 7.5 20% w/v PEG 3350	<i>JCSG+</i> H8 0.2 M Na-sulfate 0.1 M Bis Tris propane 8.5 20% w/v PEG 3350	<i>JCSG+</i> H11 0.2 M Na-citrate 0.1 M Bis Tris propane 8.5 20% w/v PEG 3350	<i>JCSG+</i> H12 0.2 M Na-malonate 0.1 M Bis Tris propane 8.5 20% w/v PEG 3350
					
<i>PACT</i> C10 0.2 M magnesium chloride 0.1 M HEPES 7.0 20 % w/v PEG 6000	<i>PACT</i> D10 0.2 M magnesium chloride 0.1 M TRIS 8.0 20 % w/v PEG 6000	<i>PACT</i> F8 0.2 M sodium sulfate 0.1 M Bis Tris propane 7.5 20% w/v PEG3350	<i>PACT</i> G2 0.2 M sodium bromide 0.1 M Bis Tris propane 8.5 20% w/v PEG3350	<i>PACT</i> G7 0.2 M sodium acetate 0.1 M Bis Tris propane 7.5 20% w/v PEG 3350	<i>PACT</i> H6 0.2 M sodium formate 0.1 M Bis Tris propane 8.5 20% w/v PEG 3350

Figure 35. Diffracting quality crystals of MtCM obtained from commercial crystallization screens *JCSG+* and *PACT premier*. Crystals grew from 1:1 volume ratio of protein and crystallization solution in sitting drop crystallization plates.

Diffracting-quality crystals were test-soaked for few hours to overnight in their respective reservoir solution diluted with either DMSO or ethylene glycol to a final 5-20% v/v concentration. Soaked crystals were then flash-cooled in liquid nitrogen and stored until data collection. Crystals that best withstood test-soaking were obtained from condition A5 from *JCSG+* and D10 from *PACT premier*. The crystals still diffracted to 2.0 Å after soaking for 2 h in 5% v/v DMSO or overnight in 10% v/v ethylene glycol. No additional cryo-protectant was required for either crystal forms. Finally, reproducibility of the two conditions was checked. We accidentally exchanged PEG 6000 with PEG 600 for PACT-D10; however, even the new condition yielded well-diffracting and highly reproducible crystals, thus it was used for fragment screening (Table 14).

Table 14. Summary for crystallization of MtCM for fragment screening

Samples	Crystallization plate	Crystallization condition	V _{res} ¹ (μ L)	V _{drop} ² (μ L)	P:R ³	Temp ($^{\circ}$ C)	Fragments [100 mM]
2.8 mg/mL MtCM in 20 mM Tris-HCl pH 7.5	96-well SWISSCI UVXPO 2 Lens	0.2 M MgCl 20% w/v PEG 600 100 mM Tris-HCl pH 8.0	50	0.5	1:1	20	FragMAXlib-1 (A1-H12) FragMAXlib-2 (A1-D6)

¹ Volume of the reservoir² Volume of the crystallization drop³ Protein to reservoir volume ratio

For screening we used FragMAXlib, a library developed at the FragMAX facility (MAX IV synchrotron, Lund, Sweden) and composed of 171 fragments (Lima *et al.* 2020). Fragments were delivered as a dry powder; we dissolved each fragments in 5 μ L of pure ethylene glycol, which resulted in \sim 1 M fragment solutions. Soaking solutions were prepared by diluting 1 μ L of each fragment solution in 9 μ L of 1.1X stock crystallization solution. Crystals were soaked overnight in 0.5 μ L of soaking solution. Soaked crystals were then flash-cooled in liquid nitrogen and stored until data collection.

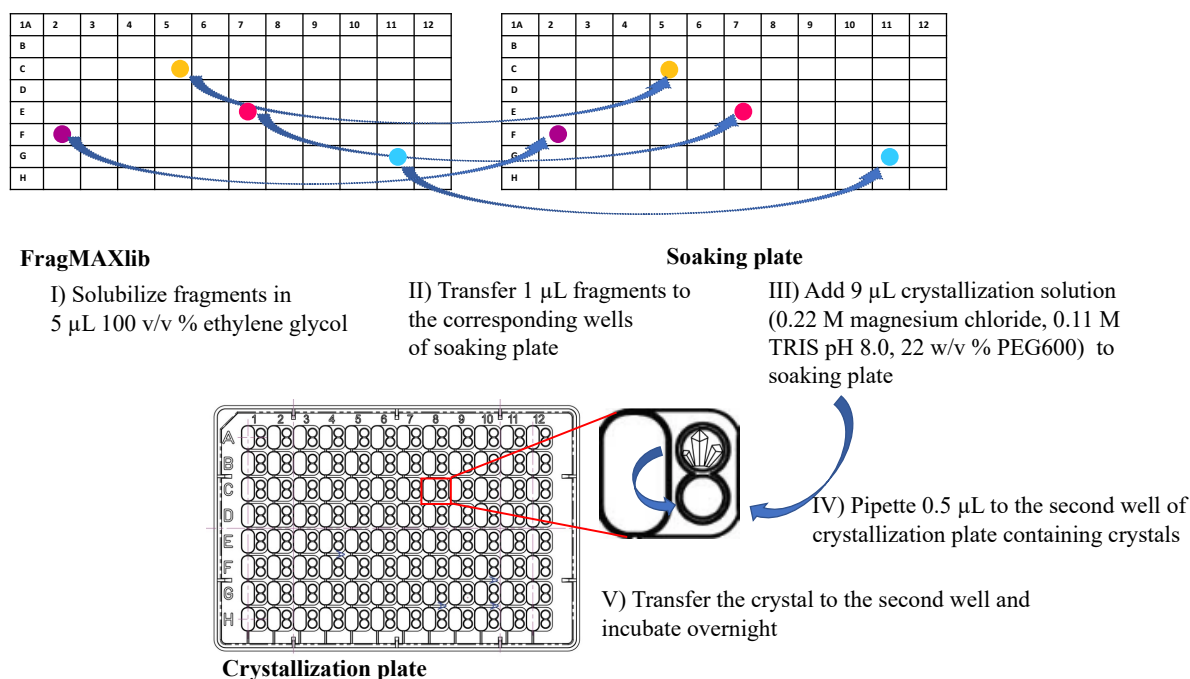


Figure 36. Workflow of fragment soaking. The FragMaxlib compound library was solubilized in pure ethylene glycol. The soaking plate was prepared by adding 1 μ L fragment to 9 μ L crystallization solution. MtCM crystals were soaked in 0.5 μ L drops from the soaking plate and incubated overnight.

6.4.2.2 Data collection and processing

Data collection, processing and analysis was performed in collaboration with Dr. Tobias Krojer at the FragMAX facility (MAX IV laboratory, Lund, Sweden). Data collection took place at cryogenic temperature (100K) at the BioMAX beamline, MAX-IV synchrotron light source (Lund, Sweden). Diffraction data for all crystals were collected with a 0.1° oscillation range covering a 180° total oscillation. All data sets were processed by multiple automated pipelines at the FragMAX facility (XDSAPP, xia2/DIALS, xia2/XDS, *fastdp*, *EDNA_proc*) (Lima *et al.* 2021). Crystals diffracted on average to a 2.0 Å resolution, with the highest resolution range varying between 1.6 and 3 Å. 78 structures were solved and partially refined by the *dimple* automated pipeline (Lima *et al.* 2021). After *dimple* processing, the PanDDA software (Pan-Dataset Density Analysis) (Pearce *et al.* 2017) was used to reveal potential binding events. The software identifies ‘crystallographic events’, such as a binding ligand, by subtracting average of dozens of ground-states (‘unbound’) data sets from a data set containing a changed-state (‘bound’); this yields a residual partial-difference map, called an *event map*. Event maps identify regions that are significantly different from the ground-state (Pearce *et al.* 2017).

6.4.3 Orthogonal validation using thermal shift assay

Due to its simple experimental setup and high-throughput option (96 and 384 wells), Thermofluor/Differential Scanning Fluorimetry (DSF) is an attractive method for fragment-screening (Kirsch *et al.* 2019). Most drug-like molecules stabilize the protein, while some destabilize it by specifically binding to the unfolded protein or accelerating the unfolding process (Cimpmperman *et al.* 2008). Due to poor sensitivity for low affinity molecules, a relatively high concentration of fragments is needed to elicit a significant shift in DSF assays, which in turn increases the number of false positive hits (Silvestre *et al.* 2013).

6.4.3.1 Experimental work

Thermal shift assays were performed as described in chapter 6.2. A pre-screen of buffers from the Rubic Additive Screen (Molecular Dimension Ltd.) was carried out, to choose a buffer that yields a clear transition in the melting curve without interfering with the fragment binding. The final choice fell on MOPS pH 7.0. Fragments from the FragMAXlib-1 and -2

libraries were screened preparing samples with a final concentration of 0.2 mg/ml MtCM, 40 mM fragment molecule, 100 mM MOPS pH 7.0, 150 mM NaCl, and 1:100 Sypro Orange in a final volume of 25 μ L per well. Samples were prepared from 1M stock solutions of the fragments dissolved in pure ethylene glycol, leading to a final concentration of 4% v/v ethylene glycol. A MtCM stock solution, initially stored in 20 mM BTP, 150 mM NaCl pH 7.5, was used in sample preparation. Sypro Orange was diluted in water from a stock solution of the dye stored in pure DMSO. The melting temperature of native MtCM was calculated from five independent measurements, while for the fragments a single measurement was done.

Table 15 A – Data collection and refinement statistics (CM/CDT bifunctional enzymes)

<i>Data collection statistics</i>		JbCDTCM MES	JbCDTCM PPY TSA	JbCDTCM PPY TSA	DsCDTCM	AfCMDT ACT	AfCMDT TSA
Crystals	BioMAX/MAXIV	ID30A-1/ESRF	ID30B/ESRF	BioMAX/MAXIV	ID30A-1/ESRF	ID30B/ESRF	ID30B/ESRF
Wavelength (Å)	0.976	0.965	0.976	0.984	0.965	0.976	0.976
Space group	<i>P2₁2₁2₁</i>	<i>P12₁1</i>	<i>P2₁2₁2₁</i>	<i>P2₁2₁2₁</i>	<i>P2₁</i>	<i>P12₁1</i>	<i>P12₁1</i>
Unit cell dimension							
<i>a</i> , <i>b</i> , <i>c</i> (Å)	39.4, 73.3, 130.5	40.5, 128.6, 50.0	41.5, 74.3, 129.4	92.1, 111.9, 221.4	36.4 98.8 55.3	48.6 105.3 55.8	48.6 105.3 55.8
α , β , γ (°)	90.0, 90.0, 90.0	90.0 111.5 90.0	90.0 90.0 90.0	90.0, 90.0, 90.0	90.0, 0 108.5, 90.0, 0	90.0 112.5 90.0	90.0 112.5 90.0
Number of protein chains in the a.s.u.	1	1	1	4	1	1	1
Matthew's coefficient	2.2	2.7	2.2	3.2	2.1	2.3	2.3
Resolution (Å) ^a	65.6-1.66 (1.84-1.65) ^a	46.6-1.80 (1.86-1.80)	41.5-1.90 (1.94-1.90)	99.9-2.44 (2.66-2.44)	52.5-1.55 (1.73-1.55)	46.3-2.36 (2.44-2.36)	46.3-2.36 (2.44-2.36)
CC _{1/2} (%) ^a	99.86 (0.5436)	99.6 (35.8)	99.6 (36.4)	99.4 (0.12)	99.7 (52.9)	95.9 (47.5)	95.9 (47.5)
Mean I/ σ (I) ^a	18.41 (1.60.8)	8.4 (0.8)	10.6 (0.8)	7.8 (1.5)	9.3 (1.6)	4.1 (1.1)	4.1 (1.1)
Completeness (%) ^a							
Spherical	72.1 (13.5)	99.3 (100)	97.8 (95.8)	60.3 (13.5)	61.7 (14.7)	97.2 (86.3)	97.2 (86.3)
Ellipsoidal	93.0 (54.6)			94.0 (72.7)	88.9 (69.7)		
Number of unique reflections ^a	43762 (2632)	43727 (2632)	31515 (1955)	51859 (2594)	33225 (2216)	20917 (1867)	20917 (1867)
Multiplicity ^a	12.7 (9.9)	3.3 (3.4)	3.1 (2.5)	13.2 (12.9)	6.5 (6.9)	3.1 (2.7)	3.1 (2.7)
Wilson B-factor (Å ²)	-	22.4	36.6	-	-	34.2	34.2
Refinement statistics							
Resolution range (Å)	65.6-1.65	46.6-1.80	41.5-1.90	99.9-2.44	52.4-1.55	46.3-2.36	46.3-2.36
<i>R</i> _{free} / <i>R</i> _{work} (%) ^b	24.3/19.9	24.7/22.0	25.9/21.8	25.8/23.0	24.8/20.5	28.1/22.5	28.1/22.5
Average <i>B</i> -factors (Å ²)	30.3	36.1	43.6	59.5	26.2	40.3	40.3
Number of non-hydrogen atoms	3587	3367	3209	12318	3269	3329	3329
Protein	3421	3156	3129	12234	3139	3282	3282
Water	154	188	52	71	117	21	21
Ligands	12	23	28	10	13	26	26
r.m.s.d. from ideal geometry							
Bond lengths (Å)	0.010	0.012	0.010	0.007	0.010	0.009	0.009
Bond angles (deg.)	1.74	1.66	1.54	1.11	1.64	1.76	1.76
Ramachandran plot							
Favored (%)	97.7	96.1	97.1	96.4	95.8	95.1	95.1
Allowed	2.3	3.4	2.87	3.3	3.9	3.6	3.6
Outliers (%)	0	0.5	0	0.3	0.3	1.3	1.3
PDB code	8CQ4	8CQ6	8CQ6	8CQ6	8CQ3	8CQ3	8CQ3

^a Values in parentheses refer to highest resolution shell.

^b *R*_{free} was calculated from 5% of randomly selected reflections for each data set.

Table 15 B – Staraniso report (CM/CDT bifunctional enzymes)

Crystals	*JbCDTCM, MES	*DsCDTCM	*AfCMCDT
Resolution range			
Diffraction limit #1 (Å)	1.95	3.71	2.08
Principal axes (orthogonal basis)	1.0000, 0.0000, 1.0000	1.0000, 0.0000, 0.0000	0.9777, 0.0000, -0.21000
Principal axes (reciprocal lattice)	a*	a*	0.784 * - 0.621 c*
Diffraction limit #2 (Å)	1.81	2.54	1.67
Principal axes (orthogonal basis)	0.0000, 1.0000, 0.0000	0.0000, 1.0000, 0.0000	0.0000, 1.0000, 0.0000
Principal axes (reciprocal lattice)	b*	b*	b*
Diffraction limit #3 (Å)	1.65	2.44	1.55
Principal axes (orthogonal basis)	0.0000, 0.0000, 1.0000	0.0000, 0.0000, 1.0000	0.2100, 0.0000, 0.97777
Principal axes (reciprocal lattice)	c*	c*	0.159 a* + 0.987 c*
Wilson B-factors (Å²)			
Eigenvalue #1 (Å)	35.3	135.41	46.6
Principal axes (orthogonal basis)	1.0000, 0.0000, 0.0000	1.0000, 0.0000, 0.0000	0.9900, 0.0000, -0.1408
Principal axes (reciprocal lattice)	a*	a*	0.824 a* - 0.567 c*
Eigenvalue #2 (Å)	31.2	52.8	18.3
Principal axes (orthogonal basis)	0.0000, 1.0000, 0.0000	0.0000, 1.0000, 0.0000	0.0000, 1.0000, 0.0000
Principal axes (reciprocal lattice)	b*	b*	b*
Eigenvalue #3 (Å)	23.6	47.1	17.2
Principal axes (orthogonal basis)	0.0000, 0.0000, 1.0000	0.0000, 0.0000, 1.0000	0.1408, 0.0000, 0.9900
Principal axes (reciprocal lattice)	c*	c*	0.103 a* + 0.995 c*

Diffraction limits and eigenvalues of overall anisotropy tensor on |F|s are displayed alongside the corresponding principal axes of the ellipsoid fitted to the diffraction cut-off surface as direction cosines in the orthogonal basis and in terms of reciprocal unit-cell vectors.

Table 16 A – Data collection and refinement statistics (monofunctional CM and CDT enzymes)

<i>Data collection statistics</i>				
Crystals	*PaeCM	*PaeCDT	MtCM T52P	MtCM V55D
Beamline	ID23-2/ESRF	ID23-2/ESRF	ID3-A30/ ESRF	ID29/ESRF
Wavelength (Å)	0.873	0.873	0.9677	0.9753
Space group	$P3_2$	$P1$	$P4_32_12$	$P2_1 2_1$
Cell parameters - a, b, c (Å)	48.4, 48.4, 113.4	77.1 80.4 81.0	59.6, 59.6, 46.6	32.2, 59.7, 72.1
α, β, γ (Å)	90 90 12.0	99.1 106.0 106.8	90, 90, 90	90, 90, 90
Number of protein chains/a.s.u	2	6	1	2
Matthew's coefficient	2.1	2.8	2.0	1.7
Resolution (Å) ^a	39.4-1.2 (1.23-1.21)	75.2-2.2 (2.24-2.20)	36.7-1.64 (1.72-1.64)	45.9-2.06 (2.25-2.06)
CC _{1/2} (%) ^{ab}	99.9 (50.7)	99.4 (49.6)	99.5 (45.1)	99.9 (47.6)
Mean I/σ(I)	15.0 (1.5)	8.4 (1.8)	15.7 (1.1)	14.6 (1.1)
Completeness (%) ^a				
Spherical	97.4 (79.8)	97.9 (97.8)	88.1 (33.4)	82.4 (22.3)
Ellipsoidal			94.0 (57.0)	92.0 (43.3)
Number of unique reflections ^a	88399 (3688)	85843 (4542)	9570 (479)	7445 (438)
Multiplicity ^a	9.8 (7.8)	3.4 (3.5)	8.7 (9.8)	6.0 (6.1)
Wilson B-factor (Å ²)	14.6	31.1	22.2	57.8
<i>Refinement</i>				
Resolution range (Å)	39.4-1.2	75.2-2.2	29.8-1.64	45.9-2.06
R_{free}/R_{work} (%) ^a	20.6/19.3	23.0/18.6	24.0/26.5	
Average B-factors (Å ²)	20.5	35.1	32.6	
Number of non-hydrogen atoms				
Protein	2689	11889	674	
Water	143	172	18	
Ligands	26	0		
r.m.s.d. from ideal geometry				
Bond lengths (Å)	0.016	0.009	0.010	
Bond angles (deg.)	1.96	1.54	1.20	
Ramachandran plot				
Favored (%)	99.0	96.6	93.1	
Allowed	1.0	3.2	5.2	
Outliers (%)	0.0	0.2	0	
PDB code			6YGT	

^a Values in parentheses refer to highest resolution shell.

^b R_{free} was calculated from 5% of randomly selected reflections for each data set.

Diffraction limits and eigenvalues of overall anisotropy tensor on |F|s are displayed alongside the corresponding principal axes of the ellipsoid fitted to the diffraction cut-off surface as direction cosines in the orthogonal basis and in terms of reciprocal unit-cell vectors.

Table 16 B – Staraniso report (monofunctional CM and CDT enzymes)

Crystals	MtCM T52P	MtCM V55D
Resolution range		
Diffraction limit #1 (Å)	1.62	2.00
Principal axes (orthogonal basis)	1.0000, 0.0000, 0.0000	1.0000, 0.0000, 0.0000
Principal axes (reciprocal lattice)	a*	a*
Diffraction limit #2 (Å)	1.62	2.22
Principal axes (orthogonal basis)	0.0000, 1.0000, 0.0000	0.0000, 1.0000, 0.0000
Principal axes (reciprocal lattice)	b*	b*
Diffraction limit #3 (Å)	1.78	2.20
Principal axes (orthogonal basis)	0.0000, 0.0000, 1.0000	0.0000, 0.0000, 1.0000
Principal axes (reciprocal lattice)	c*	c*
Wilson B-factors (Å²)		
Eigenvalue #1 (Å)	24.4	52.4
Principal axes (orthogonal basis)	1.0000, 0.0000, 0.0000	1.0000, 0.0000, 0.0000
Principal axes (reciprocal lattice)	a*	a*
Eigenvalue #2 (Å)	24.4	63.9
Principal axes (orthogonal basis)	0.0000, 0.0000, 0.0000	0.0000, 1.0000, 0.0000
Principal axes (reciprocal lattice)	b*	b*
Eigenvalue #3 (Å)	37.6	62.9
Principal axes (orthogonal basis)	0.0000, 0.0000, 1.0000	0.0000, 0.0000, 1.0000
Principal axes (reciprocal lattice)	c*	c*

Diffraction limits and eigenvalues of overall anisotropy tensor on |F|s are displayed alongside the corresponding principal axes of the ellipsoid fitted to the diffraction cut-off surface as direction cosines in the orthogonal basis and in terms of reciprocal unit-cell vectors.

6.5 Small-angle X-ray scattering

6.5.1 Principles of the method

Small-angle X-ray scattering (SAXS) is a low-resolution structural technique (below the nm-range accuracy), providing an estimate of the size and shape of X-ray-scattering particles in solution (Mertens and Svergun 2010). The shape and size of the particles can be derived from the scattering curve, which is a plot of the scattering intensity as a function of the amplitude of the scattering vector:

$$q = \frac{4\pi \sin\theta}{\lambda},$$

where λ is the wavelength and 2θ is the scattering angle (Da Vela and Svergun 2020) (Receveur-Bréchet and Durand 2012).

Scattering data can provide the size and a low-resolution shape of a protein in solution (Grant *et al.* 2015, Da Vela and Svergun 2020). Its flexibility and quaternary structure can be derived from SAXS data by comparing them to the theoretical scattering calculated from an atomistic structure of the protein, when available (Rambo and Tainer 2013, Korasick and Tanner 2018). Basic parameters to evaluate protein folding and flexibility from SAXS data are the radius of gyration (R_g), Kratky plot, and $P(r)$ function (Receveur-Bréchet and Durand 2012). The radius of gyration (R_g) provides an average size of a protein in solution. It is determined using the Guinier equation:

$$I(q) = I(0) \exp\left(-\frac{q^2 R_g^2}{3}\right)$$

where $I(q)$ is a scattering intensity at scattering vector q and $I(0)$ is a forward scattering intensity. R_g and $I(0)$ can be obtained from the slope and y-axis interception of the linear region of a Guinier plot, ($\ln[I(q)]$ versus q^2) (Mertens and Svergun 2010).

The $P(r)$ function, which is a histogram of distances between pair of atoms within a protein, gives an estimation of the size of the protein along with its overall shape. Globular proteins display a bell-shaped $P(r)$ function, which gradually goes to 0 at D_{\max} , while the $P(r)$ function of unfolded proteins (or proteins with a rod-like shape) shows an extended tail, making it difficult to estimate the D_{\max} . Multidomain proteins have multiple peaks on the $P(r)$ function. Size and shape of the enzymes are interpreted from Kratky plots (qR_g vs $(qR_g)^2 \times I(q)/I(0)$)

and $P(r)$ functions calculated from SAXS diffractograms. A well-folded, globular protein would show a maximum in the Kratky plot at $\sqrt{3}$ vs 1.104 (Receveur-Bréchet and Durand 2012).

In this thesis, SAXS was used to analyze the quaternary structure of the bifunctional (*ScCMCDT, *JbCDTCM) and monofunctional enzymes (*PaeCM, *PaeCDT).

6.5.2 Sample preparation

As all particles in solution contribute to scattering, sample monodispersity is of utmost importance in SAXS experiment. All samples were purified by size-exclusion chromatography (SEC) to high purity and homogeneity and stored at -80°C . Before the shipment, the proteins were thawed on ice and spun down at 21 000g, 4°C for 15 min. The supernatants were filtered through a sterile filter with $0.2\ \mu\text{m}$ pore size to remove aggregates. The fusion enzyme samples were refrozen and transported to the ESRF (Grenoble, France) on dry ice. The monofunctional *PaeCDT and *PaeCM were shipped on ice blocks without refreezing.

6.5.3 Data collection and processing

All SAXS data was collected at the BM29 beamline (BioSAXS) at the ESRF (Grenoble, France) (Kieffer *et al.* 2022). The data collection was performed at 20°C in the q-range $0.0039\text{--}0.5210\ \text{\AA}^{-1}$, with an X-ray wavelength of $0.992\ \text{\AA}$. 10 frames of 1 s were collected for each sample analyzed in this thesis, including buffers; integration of the frames and buffer subtraction were performed by the SAXS beamline autoprocessing software (Kieffer *et al.* 2022). The data was further processed using the ATSAS software package (Mertens and Svergun 2010, Manalastas-Cantos *et al.* 2021).

Table 17 Sample details

	*PaeCDT	*PaeCM+TSA	*ScCMCDT	*JbCDTCM
MM^* from sequence (Da)	27826	18891	45695	43995
Protein concentration (g/cm^3)	0.8	2	1.5	1
\bar{v} ($\text{cm}^3\ \text{g}^{-1}$)	0.73	0.73	0.73	0.73
ρ (cm^{-2})	2.95	2.97	2.89	2.89
Solvent	20 mM K-phosphate pH 7.3, 150 mM NaCl		20 mM Tris-HCl pH 7.2, 150 mM NaCl/v glycerol (no salt for ScCMCDT)	

Table 18 SAXS data collection parameters

	*PaeCDT	*PaeCM+TSA	*ScCMCDT	*JbCDTCM
Instrument	BM29 with Pilatus2M detector at ESRF			
Wavelength (Å)	0.992			
Beam size (µm)	200H x 100V			
Sample to detector distance (m)	2.813			
q range (Å ⁻¹)	0.0051-0.5155	0.00514-0.5160	0.0039-0.5180	0.0044-0.5205
Absolute scaling	Milli-Q water standard measurement			
Normalization	Transmitted intensities through semi-transparent beam-stop			
Exposure time	1 s/frame, 10 frames in total			
Radiation damage monitorization	Frame to frame evaluation (correlation mapping)			
Capillary size (mm)	1			
Collection temperature (°C)	4	20	20	20

Table 19 Softwares used for data processing and analysis

	*PaeCDT	*PaeCM+TSA	*ScCMCDT	*DsCDTCM	*JbCDTCM
SAXS data reduction	Integrate multi-frame and sample chager pipeline at the BM29				
Calculation of contrast and specific volume	Neutron activation and scattering calculator				
Basic analysis (Guinier, Kratky, $P(r)$)	<i>PRIMUS (ATSAS)</i>				
Shape reconstruction	<i>DAMMIF, DAMAVER, DAMMIN (ATSAS)</i>				
Atomic structure modeling	<i>Pepsi-SAXS</i>				
Shape superimposition	<i>SUPCOMB, CIFSUP (ATSAS)</i>				
Representation	<i>PyMOL</i>				

Table 20 Structural parameters

	*PaeCDT	*PaeCM+TSA	*ScCMCDT	*JbCDTCM
Guinier analysis				
I(0) (cm ⁻¹)	0.041±0.00	0.08±0.00	0.022±0.00	0.016±0.00
R _g (Å)	29.8±0.01	21.5±0.00	30.5±0.05	24.2±0.05
q _{min} (Å ⁻¹)	0.0204	0.0246	0.0106	0.0096
qR _{gmax}	1.30	1.30	1.30	1.30
Fidelity	0.83	0.86	0.95	0.55
MM from I(0) (kDa) ratio to predicted	70988/2.6	37153/2.0	20122/0.4	21951/0.5
P(r) analysis				
I(0) (cm ⁻¹)	0.041±0.00	0.057±0.00	0.022±0.00	0.016±0.00
R _g (Å)	29.6±0.01	21.1±0.00	32.0±0.00	25.3±0.04
D _{max} (Å)	86	65	101	80
q range (Å ⁻¹)	0.0205-0.268	0.0097-0.373	0.0142-0.242	0.0116-0.330
Fidelity	1.0	0.98	0.83	0.95
MM from I(0) (kDa) (ratio to predicted)	70988/2.6	36512/1.9	20122/0.4	21951/0.5

Table 21 Shape model-fitting results

	*PaeCDT	*PaeCM+TSA	*ScCMCDT	*JbCDTCM
--	---------	------------	----------	----------

<i>DAMMIF</i>				
Symmetry, anisotropy assumptions	P1, unknown	P1, unknown	P1, unknown	P1, unknown
<i>DAMAVER/DAMMIN</i>				
Symmetry, anisotropy assumptions	P1, unknown	P1, unknown	P1	P1, unknown
χ^2	0.7	1.1	1.1	0.9

Table 22 Atomistic modelling

	*PacCDT	*PacCM+TSA	*ScCMCDT	*JbCDTCM
<i>Pepsi-SAXS</i>				
Model	*PacCDT trimer	*PacCM dimer	Alpha-Fold model	*JbCDTCM MES
Flexibility optimization	enabled	enabled	enabled	enabled
Rigid blocks, residues	13-100	2-29	1-32	2-86
	117-191	39-52	40-57	108-180
	217-250	59-81	61-90	205-231
		95-119	103-159	238-261
		124-139	177-255	270-285
		143-160	266-350	291-315
			362:406	333-359
			363:388	
χ^2	0.9	1.1	1.0	1.0

7 References

- Aalbers, F.S. and Fraaije, M.W. (2019). "Enzyme fusions in biocatalysis: coupling reactions by pairing enzymes." ChemBioChem **20**(1): 20-28.
- Adler, P., *et al.* (2009). "Mining for coexpression across hundreds of datasets using novel rank aggregation and visualization methods." Genome Biol. **10**(12): 1-11.
- Afonine, P.V., *et al.* (2012). "Towards automated crystallographic structure refinement with phenix.refine." Acta Crystallogr. D Biol. Crystallogr. **68**(Pt 4): 352-367.
- Akowski, J.P. and Bauerle, R. (1997). "Steady-state kinetics and inhibitor binding of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase (tryptophan sensitive) from *Escherichia coli*." Biochemistry **36**(50): 15817-15822.
- Assis, R.A.B., *et al.* (2021). "A Secreted chorismate mutase from *Xanthomonas arboricola* pv. *juglandis* attenuates virulence and walnut blight symptoms." Int. J. Mol. Sci. **22**(19).
- Averesch, N.J.H. and Kromer, J.O. (2018). "Metabolic engineering of the shikimate pathway for production of aromatics and derived compounds-present and future strain construction strategies." Front. Bioeng. Biotechnol. **6**: 32.
- Bai, Y., *et al.* (2019). "Domain cross-talk within a bifunctional enzyme provides catalytic and allosteric functionality in the biosynthesis of aromatic amino acids." J. Biol. Chem. **294**(13): 4828-4842.
- Bai, Y. and Parker, E.J. (2021). "Reciprocal allostery arising from a bienzyme assembly controls aromatic amino acid biosynthesis in *Prevotella nigrescens*." J. Biol. Chem. **297**(3).
- Bakheet, T.M. and Doig, A.J. (2009). "Properties and identification of human protein drug targets." Bioinformatics **25**(4): 451-457.
- Bartlett, P.A. and Johnson, C.R. (1985). "An inhibitor of chorismate mutase resembling the transition-state conformation." J. Am. Chem. Soc. **107**(25): 7792-7793.
- Bassetti, M., *et al.* (2018). "How to manage *Pseudomonas aeruginosa* infections." Drugs in context **7**.
- Batool, M., *et al.* (2019). "A Structure-based drug discovery paradigm." Int. J. Mol. Sci. **20**(11).
- Bauler, P., *et al.* (2010). "Channeling by proximity: the catalytic advantages of active site colocalization using Brownian dynamics." J. Phys. Chem. Lett. **1**(9): 1332-1335.
- Bauters, L., *et al.* (2020). "Chorismate mutase and isochorismatase, two potential effectors of the migratory nematode *Hirschmanniella oryzae*, increase host susceptibility by manipulating secondary metabolite content of rice." Mol. Plant. Pathol. **21**(12): 1634-1646.
- Becerril, J.M., *et al.* (1989). "Glyphosate effects on shikimate pathway products in leaves and flowers of velvetleaf." Phytochemistry **28**(3): 695-699.
- Bentley, R. and Haslam, E. (1990). "The shikimate pathway—a metabolic tree with many branches." Crit. Rev. Biochem. Mol. Biol. **25**(5): 307-384.
- Berman, H.M., *et al.* (2002). "The Protein Data Bank." Acta Crystallogr. D Biol. Crystallogr. **58**(Pt 6 No 1): 899-907.
- Bird, A.F. and Bird, J. (2012). The structure of nematodes, Academic Press.

- Blundell, T.L., *et al.* (2002). "High-throughput crystallography for lead discovery in drug design." Nat. Rev. Drug Discov. **1**(1): 45-54.
- Brandon, M. (2004). "Bifunctional and moonlighting enzymes: lighting the way to regulatory control." Trends Plant Sci. **9**(5): 221-228.
- Brown, K.D. and Doy, C.H. (1966). "Control of three isoenzymic 7-phospho-2-oxo-3-deoxy-D-arabino-heptonate-D-erythrose-4-phosphate lyases of *Escherichia coli W* and derived mutants by repressive and "inductive" effects of the aromatic amino acids." Biochim. Biophys. Acta - Gen. Subj. **118**(1): 157-172.
- Bueno, P.S., *et al.* (2019). "New inhibitors of chorismate synthase present antifungal activity against *Paracoccidioides brasiliensis*." Future Microbiol. **14**: 969-980.
- Burschowsky, D., *et al.* (2018). "Inter-enzyme allosteric regulation of chorismate mutase in *Corynebacterium glutamicum*: structural basis of feedback activation by Trp." Biochemistry **57**(5): 557-573.
- Burschowsky, D., *et al.* (2014). "Electrostatic transition state stabilization rather than reactant destabilization provides the chemical basis for efficient chorismate mutase catalysis." Proc. Natl. Acad. Sci. USA **111**(49): 17516-17521.
- Calhoun, D.H., *et al.* (2001). "The emerging periplasm-localized subclass of AroQ chorismate mutases, exemplified by those from *Salmonella typhimurium* and *Pseudomonas aeruginosa*." Genome Biol. **2**(8): research0030.0001.
- Carr, R. and Jhoti, H. (2002). "Structure-based screening of low-affinity compounds." Drug Discov. Today **7**(9): 522-527.
- Cascante, M., *et al.* (1994). "Enzyme-enzyme interactions and metabolite channelling: alternative mechanisms and their evolutionary significance." Biochem J. **298** (Pt 2): 313-320.
- Chilingaryan, Z., *et al.* (2012). "Fragment-based screening by protein crystallography: successes and pitfalls." Int. J. Mol. Sci. **13**(10): 12857-12879.
- Chiu, H.J., *et al.* (2010). "The structure of *Haemophilus influenzae* prephenate dehydrogenase suggests unique features of bifunctional TyrA enzymes." Acta Crystallogr. F:Struct. Biol. Commun. **66**(10): 1317-1325.
- Chook, Y.M., *et al.* (1994). "The monofunctional chorismate mutase from *Bacillus subtilis*: structure determination of chorismate mutase and its complexes with a transition state analog and prephenate, and implications for the mechanism of the enzymatic reaction." J. Mol. Biol. **240**(5): 476-500.
- Chook, Y.M., *et al.* (1993). "Crystal structures of the monofunctional chorismate mutase from *Bacillus subtilis* and its complex with a transition state analog." Proc. Natl. Acad. Sci. U.S.A. **90**(18): 8600-8603.
- Cimmperman, P., *et al.* (2008). "A quantitative model of thermal stabilization and destabilization of proteins by ligands." Biophys. J. **95**(7): 3222-3231.
- Claeysens, F., *et al.* (2011). "Analysis of chorismate mutase catalysis by QM/MM modelling of enzyme-catalysed and uncatalysed reactions." Org. Biomol. Chem. **9**(5): 1578-1590.
- Clifton, B.E., *et al.* (2018). "Evolution of cyclohexadienyl dehydratase from an ancestral solute-binding protein." Nat. Chem. Biol. **14**(6): 542-547.

- Collins, P.M., *et al.* (2018). "Achieving a good crystal system for crystallographic X-ray fragment screening." Methods Enzymol. **610**: 251-264.
- Congreve, M., *et al.* (2003). "A 'rule of three' for fragment-based lead discovery?" Drug Discov. Today **8**(19): 876-877.
- Da Vela, S. and Svergun, D.I. (2020). "Methods, development and applications of small-angle X-ray scattering to characterize biological macromolecules in solution." Curr. Res. Struct. Biol. **2**: 164-170.
- Darby, E.M., *et al.* (2022). "Molecular mechanisms of antibiotic resistance revisited." Nat. Rev. Microbiol.
- Davis, B.W., *et al.* (2015). "Colocalization and sequential enzyme activity in aqueous biphasic systems: experiments and modeling." Biophys. J **109**(10): 2182-2194.
- DeClue, M.S., *et al.* (2005). "Isochorismate pyruvate lyase: a pericyclic reaction mechanism?" J. Am. Chem. Soc. **127**(43): 15002-15003.
- Della-Cioppa, G., *et al.* (1986). "Translocation of the precursor of 5-enol pyruvylshikimate-3-phosphate synthase into chloroplasts of higher plants in vitro." Proc. Natl. Acad. Sci. U.S.A. **83**(18): 6873-6877.
- Djamei, A. and Kahmann, R. (2012). "Ustilago maydis: Dissecting the Molecular Interface between Pathogen and Plant." Plos Pathogens **8**(11).
- Djamei, A., *et al.* (2011). "Metabolic priming by a secreted fungal effector." Nature **478**(7369): 395-398.
- Dopheide, T.A., *et al.* (1972). "Chorismate mutase-prephenate dehydratase from *Escherichia coli* K-12. II. Kinetic properties." J. Biol. Chem. **247**(14): 4447-4452.
- Doron, S. and Gorbach, S. (2008). "Bacterial infections: overview." Int. Public Health J.: 273.
- Doy, C.H. and Brown, K.D. (1965). "Control of aromatic biosynthesis: The multiplicity of 7-phospho-2-oxo-3-deoxy-D-arabino-heptonate D-erythrose-4-phosphate-lyase (pyruvate-phosphorylating) in *Escherichia coli*." Biochim. Biophys. Acta - Gen. Subj. **104**(2): 377-389.
- Dyer, W.E., *et al.* (1989). "Wounding induces the first enzyme of the shikimate pathway in *Solanaceae*." Proc. Natl. Acad. Sci. U.S.A. **86**(19): 7370-7373.
- Dyer, W.E., *et al.* (1990). "A cDNA encoding 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase from *Solanum tuberosum* L." J. Biol. Chem. **265**(3): 1608-1614.
- East, N.J., *et al.* (2022). "The role of oligomerization in the optimization of cyclohexadienyl dehydratase conformational dynamics and catalytic activity." Protein Sci. **31**(12): e4510.
- Elleuche, S. (2015). "Bringing functions together with fusion enzymes—from nature's inventions to biotechnological applications." Appl. Microbiol. Biotechnol. **99**(4): 1545-1556.
- Emsley, P., *et al.* (2010). "Features and development of Coot." Acta Crystallogr. D Biol. Crystallogr. **66**(Pt 4): 486-501.
- Evans, P. (2006). "Scaling and assessment of data quality." Acta Crystallogr D Biol Crystallogr **62**(Pt 1): 72-82.
- Evans, P.R. and Murshudov, G.N. (2013). "How good are my data and what is the resolution?" Acta Crystallogr. D Biol. Cryst. **69**: 1204-1214.

- Fahrig-Kamarauskait, J., *et al.* (2020). "Evolving the naturally compromised chorismate mutase from *Mycobacterium tuberculosis* to top performance." J. Biol. Chem. **295**(51): 17514-17534.
- Fichera, M.E. and Roos, D.S. (1997). "A plastid organelle as a drug target in apicomplexan parasites." Nature **390**(6658): 407-409.
- Fischer, R.S., *et al.* (1991). "Cloning, sequencing, and expression of the P-protein gene (*pheA*) of *Pseudomonas stutzeri* in *Escherichia coli*: implications for evolutionary relationships in phenylalanine biosynthesis." Microbiology **137**(6): 1293-1301.
- Fiske, M.J., *et al.* (1983). "Hidden overflow pathway to L-phenylalanine in *Pseudomonas aeruginosa*." J. Bacteriol. **154**(2): 623-631.
- Gad, S. and Ayakar, S. (2021). "Protein scaffolds: A tool for multi-enzyme assembly." Biotechnol. Rep. **32**: e00670.
- Gawas, U.B., *et al.* (2019). Structural analysis of proteins using X-ray diffraction technique. Adv. Biol. Res., Elsevier: 69-84.
- Gething, M.J. and Davidson, B.E. (1976). "Chorismate mutase/prephenate dehydratase from *Escherichia coli* K12. 2. Evidence for identical subunits catalysing the two activities." Eur. J. Biochem. **71**(2): 327-336.
- Gilchrist, D.G. and Kosuge, T. (1980). 13 - Aromatic amino acid biosynthesis and Its regulation. Amino acids and derivatives. B. J. Mifflin, Academic Press: 507-531.
- Gokhale, R.S. and Khosla, C. (2000). "Role of linkers in communication between protein modules." Curr. Opin. Chem. Biol. **4**(1): 22-27.
- Grant, T.D., *et al.* (2015). "The accurate assessment of small-angle X-ray scattering data." Acta Crystallogr. D Biol. Crystallogr. **71**(Pt 1): 45-56.
- Greer, J., *et al.* (1994). "Application of the three-dimensional structures of protein target molecules in structure-based drug design." J. Med. Chem. **37**(8): 1035-1054.
- Grudinin, S., *et al.* (2017). "Pepsi-SAXS: an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles." Acta Crystallogr. D Struct. Biol. **73**(Pt 5): 449-464.
- Han, J.J., *et al.* (2004). "Evidence for dynamically organized modularity in the yeast protein-protein interaction network." Nature **430**(6995): 88-93.
- Henstrand, J.M., *et al.* (1995). "Only the mature form of the plastidic chorismate synthase is enzymatically active." Plant physiol. **108**(3): 1127-1132.
- Herrmann, K.M. and Weaver, L.M. (1999). "The Shikimate Pathway." Annu. Rev. Plant. Physiol. Plant. Mol. Biol. **50**: 473-503.
- Hughes, J.P., *et al.* (2011). "Principles of early drug discovery." Br. J. Pharmacol. **162**(6): 1239-1249.
- Hutchings, M.I., *et al.* (2019). "Antibiotics: past, present and future." Curr. Opin. Microbiol. **51**: 72-80.
- Hyde, C.C., *et al.* (1988). "Three-dimensional structure of the tryptophan synthase alpha₂-beta₂ multienzyme complex from *Salmonella typhimurium*." J. Biol. Chem. **263**(33): 17857-17871.

- Jan, R., *et al.* (2021). "Over-Expression of chorismate mutase enhances the accumulation of salicylic acid, lignin, and antioxidants in response to the white-backed planthopper in rice plants." Antioxidants (Basel) **10**(11).
- Jensen, R.A., *et al.* (2002). "The correct phylogenetic relationship of KdsA (3-deoxy-D-manno-octulosonate 8-phosphate synthase) with one of two independently evolved classes of AroA (3-deoxy-D-arabino-heptulosonate 7-phosphate synthase)." J. Mol. Evol. **54**(3): 416.
- Johnson, M., *et al.* (2008). "NCBI BLAST: a better web interface." Nucleic Acids Res. **36**(Web Server issue): W5-9.
- Johnson, T.W., *et al.* (2018). "Reviving B-Factors: retrospective normalized B-factor analysis of c-ros oncogene 1 receptor tyrosine kinase and anaplastic lymphoma kinase L1196M with crizotinib and lorlatinib." ACS Med. Chem. Lett. **9**(9): 878-883.
- Jones, J.T., *et al.* (2003). "Characterization of a chorismate mutase from the potato cyst nematode *Globodera pallida*." Mol. Plant Pathol. **4**(1): 43-50.
- Kabir, A., *et al.* (2016). "Effects of ligand binding on the stability of aldo-keto reductases: Implications for stabilizer or destabilizer chaperones." Protein Sci. **25**(12): 2132-2141.
- Kaczmarek, J.A., *et al.* (2020). "Altered conformational sampling along an evolutionary trajectory changes the catalytic activity of an enzyme." Nat. Commun. **11**(1).
- Kapnick, S.M. and Zhang, Y. (2008). "New tuberculosis drug development: targeting the shikimate pathway." Expert. Opin. Drug Discov. **3**(5): 565-577.
- Kasaragod, P., *et al.* (2010). "Crystal structure of liganded rat peroxisomal multifunctional enzyme type 1: a flexible molecule with two interconnected active sites." J. Biol. Chem. **285**(31): 24089-24098.
- Kast, P., *et al.* (1996). "Exploring the active site of chorismate mutase by combinatorial mutagenesis and selection: the importance of electrostatic catalysis." Proc. Natl. Acad. Sci. USA **93**(10): 5043-5048.
- Keith, B., *et al.* (1991). "Differential induction of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase genes in *Arabidopsis thaliana* by wounding and pathogenic attack." Proc. Natl. Acad. Sci. U.S.A. **88**(19): 8821-8825.
- Khan, K., *et al.* (2022). "Integrated bioinformatics-based subtractive genomics approach to decipher the therapeutic drug target and its possible Intervention against *Brucellosis*." Bioengineering (Basel) **9**(11).
- Kieffer, J., *et al.* (2022). "New data analysis for BioSAXS at the ESRF." J. Synchrotron Radiat. **29**: 1318-1328.
- Kim, S.K., *et al.* (2008). "A comparative biochemical and structural analysis of the intracellular chorismate mutase (Rv0948c) from *Mycobacterium tuberculosis* H(37)R(v) and the secreted chorismate mutase (y2828) from *Yersinia pestis*." FEBS J. **275**(19): 4824-4835.
- Kirsch, P., *et al.* (2019). "Concepts and core principles of fragment-based drug design." Molecules **24**(23).
- Klebe, G. (2006). "Virtual ligand screening: strategies, perspectives and limitations." Drug discovery today **11**(13-14): 580-594.
- Kohanski, M.A., *et al.* (2010). "How antibiotics kill bacteria: from targets to networks." Nat. Review. Microbiol. **8**(6): 423-435.

- Kohnhorst, C.L., *et al.* (2017). "Identification of a multienzyme complex for glucose metabolism in living cells." *J. Biol. Chem.* **292**(22): 9191-9203.
- Korasick, D.A. and Tanner, J.J. (2018). "Determination of protein oligomeric structure from small-angle X-ray scattering." *Protein Sci.* **27**(4): 814-824.
- Kovalevskiy, O., *et al.* (2018). "Overview of refinement procedures within REFMAC5: utilizing data from different sources." *Acta Crystallogr. D Struct. Biol.* **74**: 215-227.
- Krug, M., *et al.* (2012). "XDSAPP: a graphical user interface for the convenient processing of diffraction data using XDS." *J. Appl. Crystallogr.* **45**(3): 568-572.
- Künzler, D., *et al.* (2005). "Mechanistic insights into the isochorismate pyruvate lyase activity of the catalytically promiscuous PchB from combinatorial mutagenesis and selection." *J. Biol. Chem.* **280**(38): 32827-32834.
- Kuzmak, A., *et al.* (2019). "Can enzyme proximity accelerate cascade reactions?" *Sci. rep.* **9**(1): 1-7.
- Laganenka, L., *et al.* (2016). "Chemotaxis towards autoinducer 2 mediates autoaggregation in *Escherichia coli*." *Nature communications* **7**(1): 1-11.
- Lambert, K.N., *et al.* (1999). "Cloning and characterization of an esophageal-gland-specific chorismate mutase from the phytoparasitic nematode *Meloidogyne javanica*." *Mol. Plant Microbe Interact.* **12**(4): 328-336.
- Lee, Y., *et al.* (1995). "Atomic structure of the buried catalytic pocket of *Escherichia coli* chorismate mutase." *J. Am. Chem. Soc.* **117**(12): 3627-3628.
- Li, Q. (2020). "Application of fragment-based drug discovery to versatile targets." *Front. Mol. Biosci.* **7**: 180.
- Light, S.H. and Anderson, W.F. (2013). "The diversity of allosteric controls at the gateway to aromatic amino acid biosynthesis." *Protein Sci.* **22**(4): 395-404.
- Light, S.H., *et al.* (2012). "Structural analysis of a 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase with an N-terminal chorismate mutase-like regulatory domain." *Protein Sci.* **21**(6): 887-895.
- Lima, G.M.A., *et al.* (2021). "FragMAXapp: crystallographic fragment-screening data-analysis and project-management system." *Acta Crystallogr. D Struct. Biol.* **77**(Pt 6): 799-808.
- Lima, G.M.A., *et al.* (2020). "FragMAX: the fragment-screening platform at the MAX IV Laboratory." *Acta Crystallogr. D Struct. Biol.* **76**(Pt 8): 771-777.
- Lionta, E., *et al.* (2014). "Structure-based virtual screening for drug discovery: principles, applications and recent advances." *Curr. Top. Med. Chem.* **14**(16): 1923-1938.
- Lisi, G.P. and Loria, J.P. (2017). "Allostery in enzyme catalysis." *Curr. Opin. Struct. Biol.* **47**: 123-130.
- Lutke-Eversloh, T. and Stephanopoulos, G. (2005). "Feedback inhibition of chorismate mutase/prephenate dehydrogenase (TyrA) of *Escherichia coli*: generation and characterization of tyrosine-insensitive mutants." *Appl Environ Microbiol* **71**(11): 7224-7228.
- Lutke-Eversloh, T. and Stephanopoulos, G. (2005). "Feedback inhibition of chorismate mutase/prephenate dehydrogenase (TyrA) of *Escherichia coli*: generation and

characterization of tyrosine-insensitive mutants." Appl. Environ. Microbiol. **71**(11): 7224-7228.

MacBeath, G., *et al.* (1998). "A small, thermostable, and monofunctional chorismate mutase from the archaeon *Methanococcus jannaschii*." Biochemistry **37**(28): 10062-10073.

Maeda, H. and Dudareva, N. (2012). "The shikimate pathway and aromatic amino acid biosynthesis in plants." Annu. Rev. Plant. Biol. **63**: 73-105.

Manalastas-Cantos, K., *et al.* (2021). "ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis." J. Appl. Crystallogr. **54**(Pt 1): 343-355.

Marvin, J.S. and Hellinga, H.W. (2001). "Manipulation of ligand binding affinity by exploitation of conformational coupling." Nat. Struct. Biol. **8**(9): 795-798.

Matthews, D.E. (2007). "An overview of phenylalanine and tyrosine kinetics in humans." J. Nutr. **137**(6 Suppl 1): 1549S-1555S; discussion 1573S-1575S.

Mayr, L.M. and Bojanic, D. (2009). "Novel trends in high-throughput screening." Curr. Opin. Pharmacol. **9**(5): 580-588.

McCandliss, R.J., *et al.* (1978). "3-Deoxy-D-arabino-heptulosonate 7-phosphate synthase. Purification and molecular characterization of the phenylalanine-sensitive isoenzyme from *Escherichia coli*." J. Biol. Chem. **253**(12): 4259-4265.

McConkey, G.A. (1999). "Targeting the shikimate pathway in the malaria parasite *Plasmodium falciparum*." Antimicrob. Agents Chemother. **43**(1): 175-177.

McCoy, A.J., *et al.* (2007). "Phaser crystallographic software." J. Appl. Crystallogr. **40**(Pt 4): 658-674.

McPherson, A. and Gavira, J.A. (2004). "Introduction to protein crystallization." Methods **34**(3): 254-265.

Meng, X.Y., *et al.* (2011). "Molecular docking: a powerful approach for structure-based drug discovery." Curr. Comput. Aided Drug Des. **7**(2): 146-157.

Mertens, H.D. and Svergun, D.I. (2010). "Structural characterization of proteins and complexes using small-angle X-ray solution scattering." J. Struct. Biol. **172**(1): 128-141.

Michael, C.A., *et al.* (2014). "The antimicrobial resistance crisis: causes, consequences, and management." Front. Public Health **2**: 145.

Miles, E.W., *et al.* (1999). "The molecular basis of substrate channeling." J. Biol. Chem. **274**(18): 12193-12196.

Mirdita, M., *et al.* (2022). "ColabFold: making protein folding accessible to all." Nat. Methods: 1-4.

Mohs, R.C. and Greig, N.H. (2017). "Drug discovery and development: Role of basic biological research." Alzheimers. Dement. **3**(4): 651-657.

Monaco, S., *et al.* (2013). "Automatic processing of macromolecular crystallography X-ray diffraction data at the ESRF." J. Appl. Crystallogr. **46**: 804-810.

Morrisette, N.S. and Sibley, L.D. (2002). "Cytoskeleton of apicomplexan parasites." Microbiol. Mol. Biol. Rev. **66**(1): 21-38.

Motlagh, H.N., *et al.* (2014). "The ensemble nature of allostery." Nature **508**(7496): 331-339.

- Munack, S., *et al.* (2016). "Remote control by inter-enzyme allostery: A novel paradigm for regulation of the shikimate pathway." *J. Mol. Biol.* **428**(6): 1237-1255.
- Ngo, H., *et al.* (2007). "Allosteric regulation of substrate channeling in tryptophan synthase: modulation of the L-serine reaction in stage I of the β -reaction by α -site ligands." *Biochemistry* **46**(26): 7740-7753.
- Nussinov, R. (2016). "Introduction to protein ensembles and allostery." *Chem. Rev.* **116**(11): 6263-6266.
- Okvist, M., *et al.* (2006). "1.6 Å crystal structure of the secreted chorismate mutase from *Mycobacterium tuberculosis*: novel fold topology revealed." *J. Mol. Biol.* **357**(5): 1483-1499.
- Palomino, J.C. and Martin, A. (2014). "Drug Resistance Mechanisms in *Mycobacterium tuberculosis*." *Antibiotics (Basel)* **3**(3): 317-340.
- Papaleo, E., *et al.* (2016). "The role of protein loops and linkers in conformational dynamics and allostery." *Chem. Rev.* **116**(11): 6391-6423.
- Parker, M.W. (2003). "Protein structure from X-ray diffraction." *J. Biol. Phys.* **29**(4): 341-362.
- Patel, D., *et al.* (2014). "Advantages of crystallographic fragment screening: functional and mechanistic insights from a powerful platform for efficient drug discovery." *Prog. Biophys. Mol. Biol.* **116**(2-3): 92-100.
- Pearce, N.M., *et al.* (2017). "A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density." *Nat. Commun.* **8**: 15123.
- Perkins, J.R., *et al.* (2010). "Transient protein-protein interactions: structural, functional, and network properties." *Structure* **18**(10): 1233-1243.
- Pinto, J.E., *et al.* (1988). "Glyphosate induces 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase in potato (*Solanum tuberosum* L.) cells grown in suspension culture." *Plant Physiol.* **87**(4): 891-893.
- Pittard, J. and Yang, J. (2008). "Biosynthesis of the Aromatic Amino Acids." *EcoSal Plus* **3**(1).
- Pratap, S., *et al.* (2017). "Structure of chorismate mutase-like domain of DAHPS from *Bacillus subtilis* complexed with novel inhibitor reveals conformational plasticity of active site." *Sci. Rep.* **7**(1): 1-15.
- Qian, Y., *et al.* (2019). "Completion of the cytosolic post-chorismate phenylalanine biosynthetic pathway in plants." *Nat. Commun.* **10**(1): 15.
- Ralston, A. (2008). "Operons and prokaryotic gene regulation." *Nature Education* **1**(1):216.
- Rambo, R.P. and Tainer, J.A. (2013). "Accurate assessment of mass, models and resolution by small-angle scattering." *Nature* **496**(7446): 477-481.
- Ray, J.M. and Bauerle, R. (1991). "Purification and properties of tryptophan-sensitive 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Escherichia coli*." *J. Bacteriol.* **173**(6): 1894-1901.
- Receveur-Bréchet, V. and Durand, D. (2012). "How random are intrinsically disordered proteins? A small-angle scattering perspective." *Current Protein and Peptide Science* **13**(1): 55-75.

- Rees, D.C., *et al.* (2004). "Fragment-based lead discovery." Nat. Rev. Drug. Discov. **3**(8): 660-672.
- Rizzi, C., *et al.* (2005). "DAHPSynthase from *Mycobacterium tuberculosis* H37Rv: cloning, expression, and purification of functional enzyme." Protein Expr. Purif. **40**(1): 23-30.
- Roberts, F., *et al.* (1998). "Evidence for the shikimate pathway in apicomplexan parasites." Nature **393**(6687): 801-805.
- Roderer, K., *et al.* (2014). "Functional mapping of protein-protein interactions in an enzyme complex by directed evolution." PLoS One **9**(12): e116234.
- Rodrigues-Vendramini, F.A.V., *et al.* (2019). "Promising New Antifungal Treatment Targeting Chorismate Synthase from *Paracoccidioides brasiliensis*." Antimicrob. Agents Chemother. **63**(1).
- Rupp, B. (2013). Macromolecular Crystallography: Overview. Encyclopedia of Biophysics. G. C. K. Roberts. Berlin, Heidelberg, Springer Berlin Heidelberg: 1346-1353.
- Sasso, S., *et al.* (2009). "Structure and function of a complex between chorismate mutase and DAHP synthase: efficiency boost for the junior partner." EMBO J. **28**(14): 2128-2142.
- Sasso, S., *et al.* (2005). "Characterization of the secreted chorismate mutase from the pathogen *Mycobacterium tuberculosis*." FEBS J **272**(2): 375-389.
- Schiebel, J., *et al.* (2016). "High-throughput crystallography: Reliable and efficient identification of fragment hits." Structure **24**(8): 1398-1409.
- Schilder, J. and Ubbink, M. (2013). "Formation of transient protein complexes." Curr. Opin. Struct. Biol. **23**(6): 911-918.
- Schilling, L., *et al.* (2014). "Virulence of the maize smut *Ustilago maydis* is shaped by organ-specific effectors." Mol. Plant. Pathol. **15**(8): 780-789.
- Schmid, J., *et al.* (1992). "The in-vitro synthesized tomato shikimate kinase precursor is enzymatically active and is imported and processed to the mature enzyme by chloroplasts." Plant J. **2**(3): 375-383.
- Schoner, R. and Herrmann, K.M. (1976). "3-Deoxy-D-arabino-heptulosonate 7-phosphate synthase. Purification, properties, and kinetics of the tyrosine-sensitive isoenzyme from *Escherichia coli*." J. Biol. Chem. **251**(18): 5440-5447.
- Schultes, S., *et al.* (2010). "Ligand efficiency as a guide in fragment hit selection and optimization." Drug Discov. Today: Technol **7**(3): e157-e162.
- Schupfner, M., *et al.* (2020). "Analysis of allosteric communication in a multienzyme complex by ancestral sequence reconstruction." Proc. Natl. Acad. Sci. U.S.A. **117**(1): 346-354.
- Senisterra, G., *et al.* (2012). "Thermal denaturation assays in chemical biology." Assay and drug development technologies **10**(2): 128-136.
- Seo, M., *et al.* (2014). "Protein conformational dynamics dictate the binding affinity for a ligand." Nat. Commun. **5**(1): 1-7.
- Serral, F., *et al.* (2021). "From genome to drugs: new approaches in antimicrobial discovery." Front. Pharmacol. **12**: 647060.

- Shumilin, I.A., *et al.* (2002). "Allosteric inhibition of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase alters the coordination of both substrates." J. Mol. Biol. **320**(5): 1147-1156.
- Silvestre, H.L., *et al.* (2013). "Integrated biophysical approach to fragment screening and validation for fragment-based lead discovery." Proc. Natl. Acad. Sci. USA **110**(32): 12984-12989.
- Sliwoski, G., *et al.* (2014). "Computational methods in drug discovery." Pharmacol. Rev. **66**(1): 334-395.
- Smith, D.A., *et al.* (2001). "Characterization of auxotrophic mutants of *Mycobacterium tuberculosis* and their potential as vaccine candidates." Infect. Immun. **69**(2): 1142-1150.
- Sprenger, G.A. (2007). Aromatic amino acids. Amino acid biosynthesis ~ pathways, regulation and metabolic engineering. V. F. Wendisch. Berlin, Heidelberg, Springer Berlin Heidelberg: 93-127.
- Srivastava, D.K. and Bernhard, S.A. (1986). "Enzyme-enzyme interactions and the regulation of metabolic reaction pathways." Curr. Top. Cell. Regul. **28**: 1-68.
- Stephens, C.M. and Bauerle, R. (1991). "Analysis of the metal requirement of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Escherichia coli*." J. Biol. Chem **266**(31): 20810-20817.
- Stephens, C.M. and Bauerle, R. (1992). "Essential cysteines in 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Escherichia coli*. Analysis by chemical modification and site-directed mutagenesis of the phenylalanine-sensitive isozyme." J. Biol. Chem. **267**(9): 5762-5767.
- Sträter, N., *et al.* (1997). "Mechanisms of catalysis and allosteric regulation of yeast chorismate mutase from crystal structures." Structure **5**(11): 1437-1452.
- Sudre, P., *et al.* (1992). "Tuberculosis: a global overview of the situation today." Bull. World Health Organ **70**(2): 149.
- Sundberg, S.A. (2000). "High-throughput and ultra-high-throughput screening: solution- and cell-based approaches." Curr. Opin. Biotechnol. **11**(1): 47-53.
- Suzich, J.A., *et al.* (1985). "3-Deoxy-D-arabino-heptulosonate 7-phosphate synthase from carrot root (*Daucus carota*) is a hysteretic enzyme." Plant physiology **79**(3): 765-770.
- Svedružić, Ž.M., *et al.* (2020). "Substrate channeling via a transient protein-protein complex: the case of D-glyceraldehyde-3-phosphate dehydrogenase and L-lactate dehydrogenase." Scientific reports **10**(1): 1-12.
- Sweetlove, L.J. and Fernie, A.R. (2018). "The role of dynamic enzyme assemblies and substrate channelling in metabolic regulation." Nat. Commun. **9**(1): 1-12.
- Taga, M.E. and Xavier, K.B. (2011). "Methods for analysis of bacterial Autoinducer-2 production." Curr. Protoc. Microbiol. **23**(1): 1C. 1.1-1C. 1.15.
- Tam, R. and Saier, M.H., Jr. (1993). "A bacterial periplasmic receptor homologue with catalytic activity: cyclohexadienyl dehydratase of *Pseudomonas aeruginosa* is homologous to receptors specific for polar amino acids." Res Microbiol. **144**(3): 165-169.
- Taylor, G. (2003). "The phase problem." Acta Crystallogr. D Biol. Crystallogr. **59**(11): 1881-1890.

- Tenover, F.C. and McGowan, J.E. (1996). "Reasons for the emergence of antibiotic resistance." Am. J. Med. Sci. **311**(1): 9-16.
- Tiberi, S., *et al.* (2021). "Accelerating development of new shorter TB treatment regimens in anticipation of a resurgence of multi-drug resistant TB due to the COVID-19 pandemic." Int. J. Infect. Dis. **113 Suppl 1**: S96-S99.
- Timmins, G.S. and Deretic, V. (2006). "Mechanisms of action of isoniazid." Mol. Microbiol. **62**(5): 1220-1227.
- Tzin, V., *et al.* (2012). "Shikimate pathway and aromatic amino acid biosynthesis." eLS.
- Vanholme, B., *et al.* (2009). "Structural and functional investigation of a secreted chorismate mutase from the plant-parasitic nematode *Heterodera schachtii* in the context of related enzymes from diverse origins." Mol. Plant Pathol. **10**(2): 189-200.
- Vieille, C. and Zeikus, G.J. (2001). "Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability." Microbiol. Mol. Biol. Rev. **65**(1): 1-43.
- Vonrhein, C., *et al.* (2018). "Advances in automated data analysis and processing within autoPROC, combined with improved characterisation, mitigation and visualisation of the anisotropy of diffraction limits using STARANISO." Acta Crystallogr. A **74**: A360-A360.
- Walker, G.E., *et al.* (1996). "Evidence for a novel class of microbial 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase in *Streptomyces coelicolor* A3 (2), *Streptomyces rimosus* and *Neurospora crassa*." Microbiology **142**(8): 1973-1982.
- Walsh, C. and Wenczewicz, T. (2020). Antibiotics: challenges, mechanisms, opportunities, John Wiley & Sons.
- Weaver, L.M. and Herrmann, K.M. (1997). "Dynamics of the shikimate pathway in plants." Trends Plant Sci. **2**(9): 346-351.
- Webby, C.J., *et al.* (2005). "The structure of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase from *Mycobacterium tuberculosis* reveals a common catalytic scaffold and ancestry for type I and type II enzymes." J. Mol. Biol. **354**(4): 927-939.
- Webby, C.J., *et al.* (2010). "Synergistic allostery, a sophisticated regulatory network for the control of aromatic amino acid biosynthesis in *Mycobacterium tuberculosis*." J. Biol. Chem. **285**(40): 30567-30576.
- Weinstein, J.N., *et al.* (1997). "An information-intensive approach to the molecular pharmacology of cancer." Science **275**(5298): 343-349.
- Wilson, J.W., *et al.* (2002). "Mechanisms of bacterial pathogenicity." Postgrad. Med. J. **78**(918): 216-224.
- Wilson, K.S. (2018). "An Introduction to X-Ray Crystallography." eLS: 1-16.
- Winston, D.S., *et al.* (2022). "Conformational transitions in yeast chorismate mutase Important for allosteric regulation as identified by nuclear magnetic resonance spectroscopy." J. Mol. Biol.: 167531.
- Xia, T.H., *et al.* (1993). "The AroQ-encoded monofunctional chorismate mutase (CM-F) protein is a periplasmic enzyme in *Erwinia herbicola*." J. Bacteriol. **175**(15): 4729-4737.
- Yates, R.A. and Pardee, A.B. (1956). "Control of pyrimidine biosynthesis in *Escherichia coli* by a feed-back mechanism." J. Biol. Chem. **221**(2): 757-770.

- Zhang, R. and Monsma, F. (2010). "Fluorescence-based thermal shift assays." Curr. Opin. Drug. Discov. Devel. **13**(4): 389-402.
- Zhang, Y., *et al.* (2016). "Proximity does not contribute to activity enhancement in the glucose oxidase–horseradish peroxidase cascade." Nat. Commun. **7**(1): 1-9.
- Zhang, Y.H. (2011). "Substrate channeling and enzyme complexes for biotechnological applications." Biotechnol. Adv. **29**(6): 715-725.
- Zhao, G.S., *et al.* (1993). "Cyclohexadienyl dehydratase from *Pseudomonas aeruginosa* is a periplasmic protein." J. Gen. Microbiol. **139**(4): 807-813.
- Zhao, G.S., *et al.* (1992). "Cyclohexadienyl dehydratase from *Pseudomonas aeruginosa*. Molecular cloning of the gene and characterization of the gene product." J. Biol. Chem. **267**(4): 2487-2493.
- Zheng, X.S., *et al.* (2004). "Genetic and genomic approaches to identify and study the targets of bioactive small molecules." Chem. Biol **11**(5): 609-618.
- Zhou, H., *et al.* (2021). "FRAGSITE: A fragment-based approach for virtual ligand screening." J. Chem. Inf. Model. **61**(4): 2074-2089.
- Zhu, C., *et al.* (2018). "Molecular mechanism of the synergistic activity of ethambutol and isoniazid against *Mycobacterium tuberculosis*." J. Biol. Chem. **293**(43): 16741-16750.
- Zhu, N., *et al.* (2018). "IMB-T130 targets 3-dehydroquinate synthase and inhibits *Mycobacterium tuberculosis*." Sci Rep **8**(1): 17439.

8 Publications

Novel exported bifunctional fusion enzymes with chorismate mutase and cyclohexadienyl dehydratase activity: shikimate pathway enzymes teamed up in no man's land

Christian Stocker^a, Tamjidmaa Khatanbaatar^{b,c}, Kathrin Würth-Roderer^a, Gabriele Cordara^{b,c}, Ute Krengel^{b,c*}, Peter Kast^{a*}

^a Laboratory of Organic Chemistry, ETH Zurich, CH-8093 Zurich, Switzerland

^b Department of Chemistry, University of Oslo, NO-0315 Oslo, Norway

^c Hylleraas Centre for Quantum Molecular Sciences, University of Oslo, NO-0315 Oslo, Norway

*corresponding authors: ute.krengel@kjemi.uio.no, kast@org.chem.ethz.ch

What Drives Chorismate Mutase to Top Performance? Insights from a Combined *In Silico* and *In Vitro* Study

Helen V. Thorbjørnsrud,^{||} Luca Bressan,^{||} Tamjidmaa Khatanbaatar,^{||} Manuel Carrer, Kathrin Würth-Roderer, Gabriele Cordara, Peter Kast,^{*} Michele Cascella,^{*} and Ute Krengel^{*}



Cite This: <https://doi.org/10.1021/acs.biochem.2c00635>



Read Online

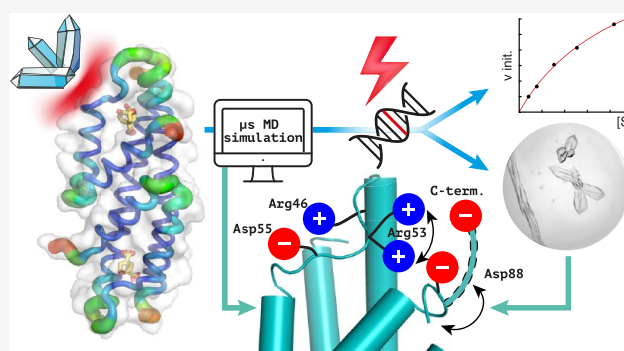
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Unlike typical chorismate mutases, the enzyme from *Mycobacterium tuberculosis* (MtCM) has only low activity on its own. Remarkably, its catalytic efficiency $k_{\text{cat}}/K_{\text{m}}$ can be boosted more than 100-fold by complex formation with a partner enzyme. Recently, an autonomously fully active MtCM variant was generated using directed evolution, and its structure was solved by X-ray crystallography. However, key residues were involved in crystal contacts, challenging the functional interpretation of the structural changes. Here, we address these challenges by microsecond molecular dynamics simulations, followed up by additional kinetic and structural analyses of selected sets of specifically engineered enzyme variants. A comparison of wild-type MtCM with naturally and artificially activated MtCMs revealed the overall dynamic profiles of these enzymes as well as key interactions between the C-terminus and the active site loop. In the artificially evolved variant of this model enzyme, this loop is preorganized and stabilized by Pro52 and Asp55, two highly conserved residues in typical, highly active chorismate mutases. Asp55 stretches across the active site and helps to appropriately position active site residues Arg18 and Arg46 for catalysis. The role of Asp55 can be taken over by another acidic residue, if introduced at position 88 close to the C-terminus of MtCM, as suggested by molecular dynamics simulations and confirmed by kinetic investigations of engineered variants.

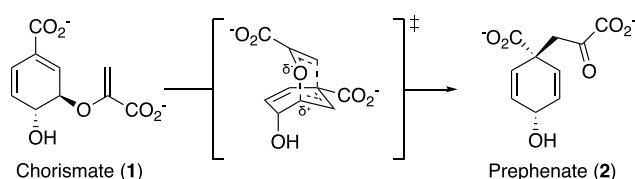


INTRODUCTION

Pericyclic reactions are common in industrial processes, but very rare in biology.^{1–4} Chorismate mutase (CM) catalyzes the only known pericyclic process in primary metabolism, the Claisen rearrangement of chorismate (1) to prephenate (2), via a chair-like transition state (Scheme 1).⁵ This catalytic step at the branch point of the shikimate pathway funnels the key metabolite chorismate toward the synthesis of tyrosine and phenylalanine, as opposed to tryptophan and several aromatic vitamins.^{6,7} The CM reaction is a concerted unimolecular transformation that is well studied by both experimental and computational means.⁸ It proceeds ostensibly *via* the same transition state in both solution and enzyme catalysis.^{9,10} Due to these factors, CM has long been a model enzyme for computational chemists.¹¹

Natural CMs belong to two main classes with two distinct folds AroH and AroQ, which are equally efficient, with typical $k_{\text{cat}}/K_{\text{m}}$ values in the range of $(1–5) \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$.¹² The AroH fold, exemplified by the *Bacillus subtilis* CM, has a trimeric pseudo- α/β -barrel structure,^{13,14} whereas the structures of AroQ enzymes have all- α -helical folds.^{15–21} The AroQ family is further divided into four subfamilies, α – δ .^{20,21} The AroQ₅ subfamily shows abnormally low catalytic activity

Scheme 1. Chorismate Mutase Reaction^a



^aThe Claisen rearrangement catalyzed by chorismate mutase converts chorismate (1) to prephenate (2) and proceeds *via* a highly polarized chair-like transition state carrying partial charges at the C–O bond that is broken during the reaction.

compared to prototypical CM enzymes. In fact, the first discovered AroQ₅ enzyme, the intracellular CM from *Mycobacterium tuberculosis* (MtCM),^{20,21} is on its own only a

Received: November 8, 2022

Revised: December 18, 2022

poor catalyst ($k_{\text{cat}}/K_{\text{m}} = 1.8 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$),²¹ despite its crucial role for producing the aromatic amino acids Tyr and Phe. However, this low activity can be boosted more than 100-fold to a $k_{\text{cat}}/K_{\text{m}}$ of $2.4 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ through formation of a noncovalent complex with the first enzyme of the shikimate pathway, 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase (MtDS) (Figure 1A).²¹

The active site of AroQ CMs is dominated by positive charges, contributed by four arginine residues (Figure 1F). In MtCM, these are Arg18', Arg35, Arg46, and Arg58 (with the prime denoting a different MtCM protomer). Of particular importance for catalysis is Arg46,²¹ or its corresponding cationic residues in other CMs (of both AroH and AroQ families).²² However, high catalytic prowess is only achieved when this cationic residue is optimally positioned such that it can stabilize the developing negative charge at the ether oxygen in the transition state (Scheme 1).^{11,14,21,23–25} In MtCM, this is not the case unless MtCM is activated by MtDS.²¹ The MtDS partner repositions residues of the C-terminus of MtCM for interaction with the H1–H2 loop of MtCM that covers its active site, thereby inducing a characteristic kink in this loop (orange circle in Figure 1C). This interaction leads to a rearrangement of active site residues to catalytically more favorable conformations (Figure 1B)²¹ and is likely a key contributing factor for the increase in CM activity, as shown by randomizing mutagenesis of the C-terminal region followed by selection for functional variants.²⁶ Complex formation also endows MtCM with feedback regulation by Tyr and Phe through binding of these effectors to the MtDS partner.^{21,27,28} Such inter-enzyme allosteric regulation²⁸ allows for dynamic adjustment of the CM activity to meet the changing needs of the cell.

The naturally low activity of MtCM in the absence of its MtDS partner enzyme also provided a unique opportunity for laboratory evolution studies aimed at increasing MtCM efficiency. After four major rounds of directed evolution, the top-performing MtCM variant N-s4.15 emerged,¹² which is abbreviated as MtCM^V in this manuscript. This variant showed autonomous CM activity ($k_{\text{cat}}/K_{\text{m}} = 4.7 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$) twice exceeding that of wild-type MtCM in the MtCM–MtDS complex, and can no longer be activated further through the addition of MtDS.¹² The biggest gains in catalytic activity were due to replacements T52P and V55D in the H1–H2 loop and R87P, L88D, G89A, and H90M at the C-terminus (Figure 1C–E). Of these residues, Pro52 and Asp55 are conserved in the H1–H2 loop of naturally highly active CMs, such as the prototypic CMs from the α - and γ -AroQ subclasses, *i.e.*, EcCM from *Escherichia coli*¹⁶ and *MtCM, the secreted CM from *M. tuberculosis*,²⁰ respectively.¹² The single amino acid exchange that had the largest beneficial effect on activity was V55D (12-fold enhancement of $k_{\text{cat}}/K_{\text{m}}$), followed by T52P (6-fold gain).¹² Combined, these two changes, discussed in detail in a previous publication, gave a $k_{\text{cat}}/K_{\text{m}}$ that was 22 times higher compared to wild-type MtCM.¹² The four C-terminal amino acid replacements together increased the activity more modestly (by a factor of 4), and the five exchanges introduced in the two final evolutionary rounds yielded an additional factor of 5. The resulting combination of large-impact and more subtle residue substitutions in MtCM^V (Figure 1D,E) gave a $k_{\text{cat}}/K_{\text{m}}$ about 500 times greater than that of the parental starting point, thereby reaching the values of the most efficient CMs known to date.¹²

The crystal structure of MtCM^V revealed a strongly kinked conformation of the H1–H2 loop. This is reminiscent of the conformation adopted by MtCM when in the complex with MtDS (the crystal structure of MtDS-bound MtCM is in the following referred to as MtCM^{DS}) and differs considerably from that observed in free MtCM (Figure 1C,D).¹² However, in the crystal structures of free wild-type and top-evolved MtCM^V, both the H1–H2 loop and the C-terminus are involved in extensive crystal contacts, making an unbiased structural evaluation of the sequence alterations in these parts of the enzyme impossible. In solution, these regions are assumed to be more flexible compared to the α -helical segments of MtCM.

Here, we used molecular dynamics (MD) simulations to investigate the behavior of MtCM in the absence or presence of ligands and to analyze whether the protein is able to interconvert between activated and nonactivated conformations in the absence of the MtDS partner enzyme. We also compared the wild-type MtCM with the evolved MtCM^V, to see if the acquired amino acid substitutions introduced any new interactions or if they altered the probabilities of existing ones, with potential impact on catalytic activity. From an assessment of the dynamic properties of MtCM and MtCM^V, we proposed a set of single, double, and triple C-terminal variants of the enzyme and subsequently tested these experimentally.

MATERIALS AND METHODS

Construction of Untagged MtCM Variants. General cloning was carried out in *E. coli* DH5 α or XL1-Blue (both Stratagene, La Jolla, California). All cloning techniques and bacterial culturing were performed according to standard procedures.²⁹ Oligonucleotide synthesis and DNA sequencing were performed by Microsynth AG (Balgach, Switzerland).

For the construction of expression plasmids pKTCMM-H-V55D and pKTCMM-H-T52P for the native MtCM single variants, the individual site-directed mutants were first constructed in the pKNTET background (providing an N-terminal His₆ tag, first 5 residues missing). Parts of the MtCM gene (Gene Rv0948c) were amplified using oligonucleotides 412-MtCM-N-V55D (5'-GTTTCGCTAGCGGAGGTA-CACGTTTGGATCATAGTCGGGAGATGAAGGTCATC-GAAC) or 413-MtCM-N-T52P (5'-GTTTCGCTAGCG-GAGGTCCGCGTTTGGTCCATAGTCGGGAGATGAAGGTCATCGAAC) together with oligonucleotides 386-LpLib-N2 (5'-GGTTAAAGCTTCCGCAGCCACTAGT-TATTAGTGACCGAGGCGGCCACGGCCCAAT) on template pMG248¹² to create a 163 bp PCR product. The PCR products were restriction digested with *Nhe*I and *Hind*III and the resulting 148 bp fragments were individually ligated to the accordingly cut 2873 bp fragment from acceptor vector pKNTET-0.¹² The ligation was performed with T4 DNA ligase (New England Biolabs, Ipswich, Massachusetts) overnight at 16 °C. The ligation products were transformed into chemically competent *E. coli* XL1-Blue cells. The cloned PCR'd DNA fragments were confirmed by Sanger sequencing. Subsequently, the genes for MtCM-T52P and MtCM-V55D were isolated by restriction digestion using enzymes *Xho*I and *Spe*I followed by a preparative agarose gel, yielding corresponding 260 bp fragments. pKTCMM-H²¹ was used as acceptor vector and was accordingly restriction digested with *Xho*I and *Spe*I, yielding a 4547 bp acceptor fragment. The fragments were ligated overnight at 16 °C, using T4 DNA

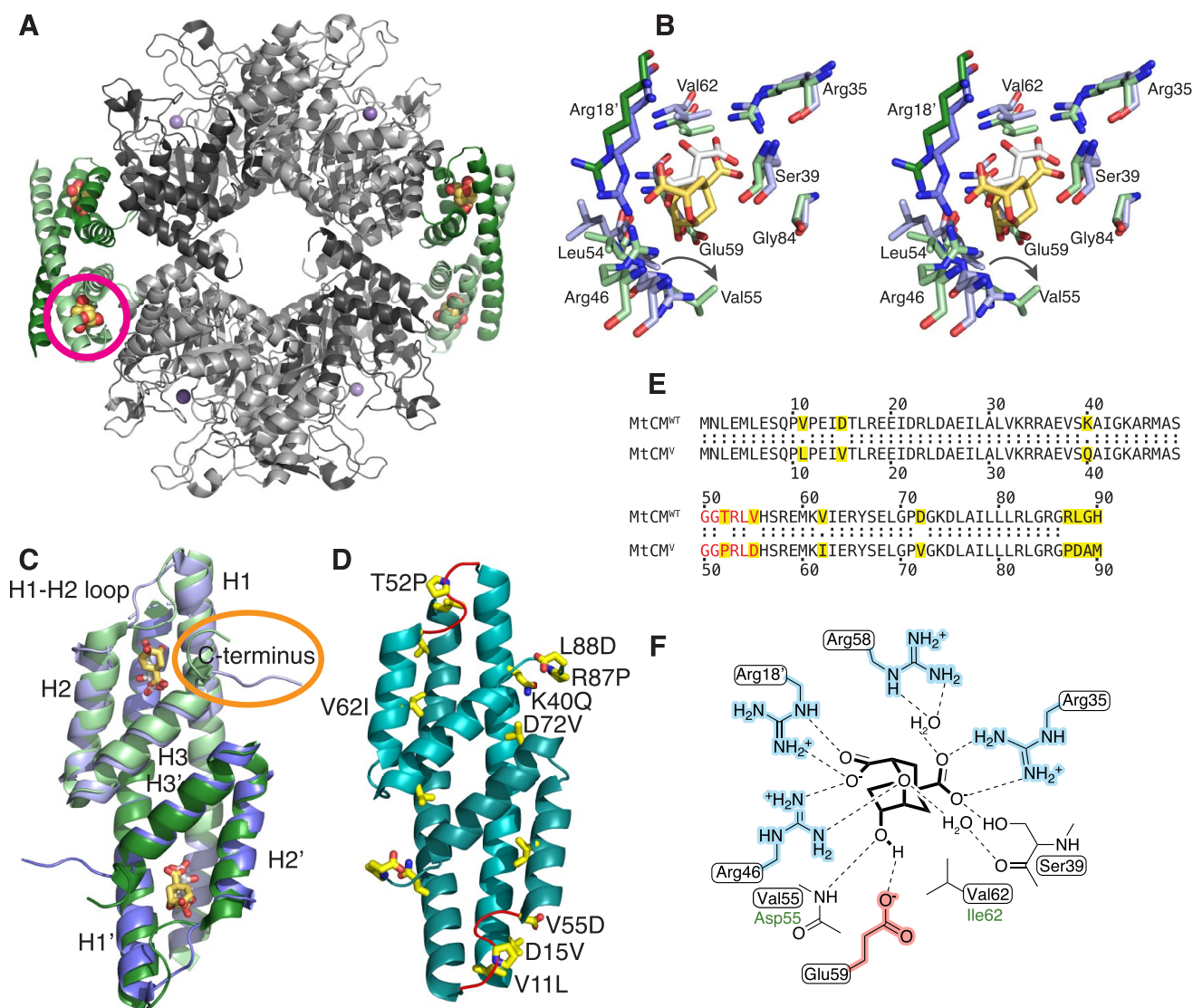


Figure 1. Structural information on *M. tuberculosis* chorismate mutase. (A) Cartoon illustration of the heterooctameric complex of MtCM with DAHP synthase (MtDS) (Protein Data Bank (PDB) ID: 2W1A).²¹ MtCM is colored in shades of green and MtDS in shades of gray to emphasize individual subunits; Bartlett’s transition state analogue (TSA)⁵¹ is shown with golden spheres. The location of one of the four active sites of MtCM is marked with a circle. (B) Stereo superimposition of CM active sites of MtCM (shades of violet, with malate bound, white sticks) (PDB ID: 2VKL)²¹ and of the MtCM–MtDS complex (with TSA bound, golden sticks) (PDB ID: 2W1A),²¹ showing several active site residues as sticks. Shades of violet/green (and prime notation for Arg18’) illustrate separate protomers of MtCM/MtCM–MtDS structures, respectively. An arrow shows the shift in the position of Val55 upon MtDS binding, allowing H-bond formation of its backbone to TSA. (C) Cartoon superimposition of MtCM (PDB ID: 2VKL,²¹ violet, with white sticks for malate ligand) and activated MtCM (MtCM^{DS}) from the MtCM–MtDS complex (PDB ID: 2W1A,²¹ green, with golden sticks for TSA). The biggest structural changes upon activation are a kink observed in the H1–H2 loop and interaction of the C-terminus (circled in orange) with the active site of MtCM. (D) Cartoon representation of the artificially evolved MtDS-independent super-active MtCM variant N-s4.15 (PDB ID: SMPV;¹² cyan), dubbed MtCM^V in this work, having a k_{cat}/K_m typical for the most efficient CMs known to date.¹² Amino acid replacements accumulated after four cycles of directed evolution are emphasized as yellow side-chain sticks (A89 and M90 are not resolved) and labeled for one of the protomers. The H1–H2 loop (shown in red) adopts a kinked conformation similar to that observed for the MtDS-activated MtCM^{DS} shown in (C). (E) Sequence alignment of wild-type MtCM (MtCM^{WT}) and the highly active variant N-s4.15 (MtCM^V).¹² Substituted residues are highlighted in yellow, and the H1–H2 loop is colored red. (F) Schematic representation of the active site of MtCM with bound TSA. Boxed residues refer to the wild-type enzyme, and green font color (Asp55, Ile62) refers to those substituted in MtCM^V. Charged residues are highlighted in red and blue.

ligase. The ligation products were transformed into chemically competent *E. coli* KA12 cells²³ and the inserts were analyzed by Sanger sequencing. The gene for variant PHS10-3p3,¹² carrying an N-terminal His₆-tag and missing the first five residues, was recloned into the native format provided by plasmid pKTCMM-H. Acceptor vector pKTCMM-H and

pKTNTET-PHS10-3p3 were restriction digested with *Xho*I and *Spe*I, and the fragments were isolated from preparative agarose gels. The 4547 bp and 260 bp fragments were ligated overnight at 16 °C with T4 DNA ligase and transformed into chemically competent XL1-Blue cells. The relevant gene sequence was confirmed by Sanger sequencing.

Different C-terminal variants of the MtCM gene were generated by PCR mutagenesis. DNA fragments were amplified with the same forward primer (containing an *NdeI* site, underlined) and different reverse primers (containing an *SpeI* site, underlined) on different DNA templates. The gene encoding MtCM L88D was produced by PCR with primers LB5 (5'-TCCGCACATATGAACCTGGAAATG) and LB4 (5'-TAAGCAACTAGTTATTAGTGACCGTCGCG) on the template plasmid pKTCMM-H carrying the wild-type gene.²¹ The gene for the triple variant MtCM (T52P V55D L88D) was assembled with primers LB5 and LB4 on a pKTCMM-H derivative containing MtCM variant 3p3 (T52P V55D).¹² The gene for MtCM variant PNAM (D88N) was generated with primers LB5 and LB6 (5'-TAAGCAACTAGTTATTACATAGCATTCGGA), and for the MtCM variant PLAM (D88L) with primers LB5 and LB7 (5'-TAAGCAACTAGTTATTAGTGACCAAGCGGA), in both cases using a version of the template plasmid pKTCMM-H, into which the gene for the top-evolved s4.15 variant had been inserted.¹² The resulting 296 bp PCR fragments containing *NdeI* and *SpeI* restriction sites at the 5' and 3' ends of the MtCM gene, respectively, were digested with the corresponding enzymes to yield 278 bp fragments. These fragments were ligated to the 4529 bp *NdeI*–*SpeI* fragment of pKTCMM-H yielding the final 4807 bp plasmids.

Protein Production and Purification. *E. coli* strain KA13^{18,30} carrying an endogenous UV5 P_{lac} -expressed T7 RNA polymerase gene was used to overproduce the (untagged) MtCM variants. KA13 cells were transformed by electroporation with the appropriate pKTCMM-H plasmid derivative that carries the desired MtCM gene variant.

For the two crystallized MtCM variants T52P (MtCM^{T52P}) and V55D (MtCM^{V55D}), the transformed cells were grown in baffled flasks at 30 °C in LB medium containing 100 μ g/mL sodium ampicillin until the OD₆₀₀ reached 0.5. Gene expression was induced through the addition of isopropyl- β -D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM, and incubation was continued overnight. The cells were harvested by centrifugation (6500g for 20 min at 4 °C) and frozen at –80 °C before being resuspended in a buffer suitable for ion exchange chromatography, supplemented with DNase I (Sigma), 150 μ M phenylmethanesulfonyl fluoride (PMSF) and cOmplete protease inhibitor cocktail (Roche). The cells were lysed using BeadBeater (BioSpec BSP 74340, Techtum Lab AB), with four times 30 s pulses with a 60 s wait between each pulse. Insoluble debris was removed by centrifugation (48,000g for 30 min at 4 °C).

The resuspension buffer was selected based on the theoretical isoelectric point (pI) of the protein. MtCM T52P has a pI of 8.14, so the pellet was resuspended in 50 mM 2-(*N*-morpholino)ethanesulfonic acid (MES), pH 6.5. MtCM V55D has a pI of 6.74; therefore, the pellet was resuspended in 50 mM acetic acid, pH 5.25. After lysis and centrifugation, the soluble lysate was loaded onto a HiTrap XL SP column (GE Healthcare) for cation exchange chromatography and eluted with a 0–0.5 M NaCl gradient. The purity of the eluted fractions was gauged by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis and sufficiently pure fractions were pooled and concentrated using concentrator tubes with a 5 kDa molecular mass cutoff (Vivaspin MWCO 5K). The proteins were then further purified by size-exclusion chromatography using a Superdex 75 300/10 column (GE Healthcare) with running buffer 20 mM 1,3-bis[tris-

(hydroxymethyl)methylamino]propane (BTP), pH 7.5, 150 mM NaCl. Finally, the proteins were concentrated (Vivaspin MWCO 5K), frozen, and stored at –80 °C.

For the sets of MtCM variants probed for the catalytic impact of particular C-terminal amino acid exchanges, 500 mL LB medium cultures containing 150 μ g mL^{–1} sodium ampicillin were inoculated with 5 mL overnight culture of the desired transformant and grown at 37 °C and 220 rpm shaking to an OD_{600nm} of 0.3–0.5. Protein production was induced by the addition of IPTG to 0.5 mM, and culture growth was continued overnight at 30 °C.

The cells were harvested by centrifugation (17,000g for 10 min at 4 °C) and washed once with 100 mM tris(hydroxymethyl)aminomethane (Tris)–HCl, pH 7.5. The cells were pelleted again, and the cell pellet was either frozen for storage at –20 °C or directly resuspended in 80 mL of sonication buffer (50 mM sodium phosphate, 0.3 M NaCl, pH 7.0). The cells were disrupted by sonication on ice (15 min total pulse time with 45 s pulse/30 s pause cycles at 50% amplitude; Q700 sonicator, QSonica). The crude lysate was cleared by centrifugation (20,000g for 20 min at 4 °C). The supernatant was supplemented with sonication buffer to 100 mL, 42 g of ammonium sulfate was added, and the solution was stirred at 4 °C for 1.5 h. The precipitate was pelleted by centrifugation (10,000g for 30 min at 4 °C), dissolved in 8 mL of low-salt buffer (20 mM piperazine, pH 9.0), and dialyzed against 1 L of low-salt buffer overnight. Dialysis was repeated against another 1 L of low-salt buffer for 3 h before application to a MonoQ (MonoQ HR 10/10, Pharmacia) FPLC column (Biologic Duoflow system, Bio-Rad). The sample was eluted over 80 mL in 20 mM piperazine by applying a gradient from 0 to 30% of a high-salt buffer (20 mM piperazine, 1 M NaCl, pH 9.0).

The MonoQ fractions containing the protein of interest were pooled and concentrated to less than 1 mL. The concentrated sample was directly applied to a gel-filtration column (Superdex Increase 75 10/300 GL, GE Healthcare) and eluted in 20 mM BTP, 150 mM NaCl, pH 7.5. Protein identity was confirmed by liquid chromatography–mass spectrometry (LC-MS) (MoBIAS facility, Laboratory of Organic Chemistry, ETH Zurich), with the observed mass being within 1 Da of the calculated mass. Protein purity was assessed by SDS-PAGE (PhastGel Homogenous 20 precast gels, GE Healthcare) and the enzyme concentration ($[E]$) was determined using the Bradford assay.³¹

X-ray Crystallography. MtCM variants T52P (MtCM^{T52P}) and V55D (MtCM^{V55D}) were crystallized in 96-well two-drop MRC crystallization plates (SWISSCI) by the sitting drop vapor diffusion technique. Diffraction-quality crystals of MtCM^{T52P} grew at 20 °C from a 1:1 (375 nL + 375 nL) mixture of protein (28 mg mL^{–1} in 20 mM BTP, pH 7.5) and reservoir solution containing 0.2 M sodium malonate, 20% PEG 3350 (w/v), and 0.1 M Bis Tris propane buffer, pH 8.5 (PACT premier crystallization screen, condition H12; Molecular Dimensions Ltd.). Crystals of MtCM^{V55D} were obtained from a 1:1 (375 nL + 375 nL) mixture of protein (44 mg mL^{–1} in 20 mM Bis Tris propane, pH 7.5, 150 mM NaCl) and reservoir solution containing 0.2 M zinc acetate dihydrate, 10% w/v PEG 3000, and 0.1 M sodium acetate, pH 4.5 (JCSG-plus crystallization screen, condition C7; Molecular Dimensions Ltd.) at 20 °C.

Diffraction data of MtCM^{T52P} and MtCM^{V55D} crystals were collected at the European Synchrotron Radiation Facility

(ESRF, Grenoble, France) at the ID30A-3/MASSIF-3 (Dectris Eiger X 4M detector) and ID29 (Pilatus detector) beamlines, respectively, covering 120° with 0.1° oscillation. Diffraction images were integrated and scaled using the *XDS* software package;³² merging and truncation were performed with *AIMLESS*³³ from the *CCP4* program suite.³⁴ Since data collection statistics of both crystals suggested the presence of anisotropy, the *XDS* output was reprocessed for anisotropy correction and truncation using the *STARANISO* server.³⁵ The “aniso-merged” output files (merged MTZ file with an anisotropic diffraction cutoff) were subsequently used for structure solution and refinement (Table S1).

The crystal structures of MtCM^{T52P} and MtCM^{V55D} were solved by molecular replacement with the program *Phaser*.³⁶ The structure of the top-evolved MtCM variant MtCM^V (PDB ID: SMPV)¹² was used as a search model for solving the structure of MtCM^{T52P} since it was expected to be a better match at the Pro52-containing H1–H2 loop compared to wild-type MtCM. For MtCM^{V55D}, we used the MtCM structure from the MtCM–MtDS complex (PDB ID: 2W1A)²¹ as a search model, after truncation of the termini and the H1–H2 loop, and removal of the ligand.

The two structures were subsequently refined, alternating between real-space refinement cycles using *Coot*³⁷ and maximum-likelihood refinement with *REFMAC5*.³⁸ The models were improved stepwise by first removing ill-defined side chains, and subsequently adding missing structural elements as the quality of the electron density map improved. Water molecules and alternative side-chain conformations were added to the MtCM^{T52P} model toward the end of the refinement process, where positive peaks in the σ_A -weighted $F_o - F_c$ difference map and the chemical surroundings allowed for their unambiguous identification. As a last step, occupancy refinement was carried out with *phenix.refine*, a tool of the *PHENIX* software suite.³⁹ The final structure of MtCM^{T52P} was deposited in the Protein Data Bank (PDB)⁴⁰ with deposition code 6YGT. Data collection and refinement statistics are summarized in Supporting Table S1.

Determination of Enzyme Kinetic Parameters.

Michaelis–Menten kinetics of the untagged purified MtCM variants were determined by a continuous spectroscopic chorismate depletion assay (Lambda 20 UV/VIS spectrophotometer, PerkinElmer). The purified enzymes were diluted into 20 mM potassium phosphate, pH 7.5, containing 0.01 mg mL⁻¹ bovine serum albumin to obtain suitable working concentrations for starting the reactions, depending on the activity of individual variants. The assays were performed at 30 °C in either 50 mM potassium phosphate, pH 7.5, or 50 mM BTP, pH 7.5. Different chorismate concentrations ($[S]$) ranging from 10 to 1500 μ M were used at 274 nm ($\epsilon_{274} = 2630$ M⁻¹ cm⁻¹) or 310 nm ($\epsilon_{310} = 370$ M⁻¹ cm⁻¹). Chorismate disappearance upon enzyme addition was monitored to determine the initial reaction velocity (v_0). The obtained data were fitted to the Michaelis–Menten equation with the program *KaleidaGraph* (Synergy Software, Reading, Pennsylvania) to obtain the catalytic parameters k_{cat} and K_m .

Molecular Dynamics Simulations. Molecular dynamics (MD) simulations were carried out on a number of representative structures for CM. They included two independent sets of simulations for apo MtCM, starting either from the X-ray crystal structure of MtCM in complex with malate (after removing malate) (PDB ID: 2VKL)²¹ or from the structure of the CM polypeptide in the apo MtCM–MtDS

complex (PDB ID: 2W19,²¹ chain D). The malate complex was chosen over ligand-free MtCM (PDB ID: 2QBV)⁴¹ due to its higher resolution and better refinement statistics. Both simulations gave essentially the same result; therefore, we will not refer to the second data set any further. For the highly active evolved MtCM variant (MtCM^V), we used the recent crystal structure (PDB ID: SMPV).¹² The MtCM–ligand complex (MtCM^{LC}) was taken from PDB ID: 2W1A,²¹ excluding the MtDS partner protein, where MtCM was co-crystallized with a transition state analog (TSA) in its active site (Figure 1). Finally, the V55D variant was modeled based on a partially refined experimental structure (Table S1). Residues that were not fully defined were added to the models using (often weak) electron density maps as reference in *Coot*.³⁷ When no interpretable density was visible, geometric restraints (and α -helical restraints for residues in helix H1) were applied during model building, to ensure stable starting geometries. The N-termini of all of the models were set at Glu13, corresponding to the first defined residue in almost all of the resolved structures available. Glu13 was capped with an acetyl group to imply the continuation of the H1 helix. CM dimers were generated by 2-fold crystallographic symmetry.

Missing H-atoms were added to the model and the systems were solvated in a periodic box filled with explicit water molecules, retaining neighboring crystallographic waters, and keeping the protein at least 12 Å from the box boundaries. The systems were neutralized through the addition of Cl⁻ ions at a minimum distance of 7 Å from the protein and each other. Additional buffering moieties like glycerol or sulfate ions found in the crystals were not considered. MD simulations were run using the Gromacs 5.1.4 package^{42,43} using the AMBER 12 force fields for the protein moieties^{44,45} and the TIP3P model for water.⁴⁶ The ligand was modeled using the GAFF force field.⁴⁷ The smooth particle mesh Ewald method was used to compute long-range electrostatic interactions,⁴⁸ while a cutoff of 11 Å was used to treat the Lennard–Jones potential.

The systems were minimized using the steepest descent/conjugate gradients algorithms for 500/1500 steps until the maximum force was less than 1000 kJ mol⁻¹ nm⁻¹. To equilibrate and heat the systems, first we ran 100 ps MD in the NVT ensemble starting from a temperature of 10 K, using the canonical velocity rescaling thermostat⁴⁹ followed by 100 ps in the NpT ensemble with a Parrinello–Rahman barostat⁵⁰ targeting a final temperature of 310 K and a pressure of 1 atm. After initial equilibration, 1 μ s of MD simulation was performed for each system. In all MD simulations, the time step size was set to 2 fs.

RESULTS

The fact that MtCM exhibits only low natural catalytic activity provided us with a perfect opportunity to probe features that optimize CM catalysis by directed evolution.¹² Since the biggest gains in catalytic activity were contributed by exchanging the H1–H2 loop residues 52 (T52P) and 55 (V55D), we set out to determine the crystal structures of these two enzyme variants. Together, these two substitutions led to an increase in k_{cat}/K_m by 22-fold compared to the parent enzyme.¹²

Crystal Structures of MtCM^{T52P} and MtCM^{V55D}. Whereas MtCM^{T52P} crystals had the same space group ($P4_32_12$) and similar cell parameters as the wild-type enzyme (PDB IDs: 2VKL²¹ and 2QBV⁴¹), with one protomer in the asymmetric unit, MtCM^{V55D} crystallized in a different space group

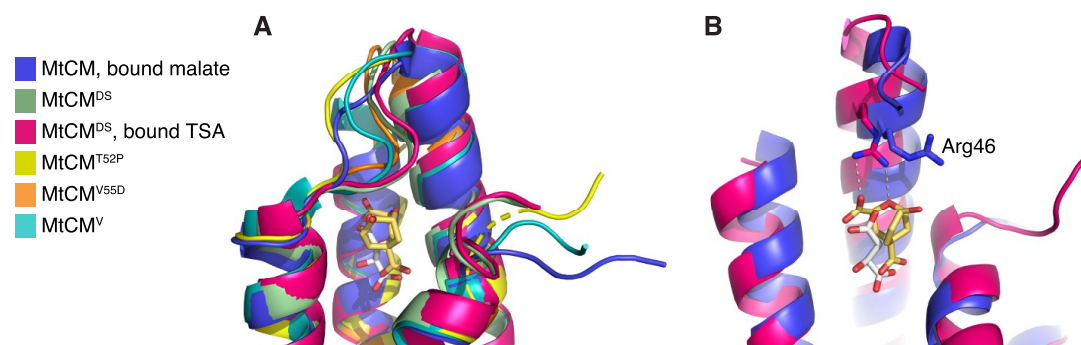


Figure 2. Comparison of MtCM crystal structures, with focus on Arg46 and H1–H2 loop. (A) Superimposition of the active site of MtCM (PDB ID: 2VKL;²¹ violet), MtCM^{DS} (PDB ID: 2W19;²¹ green), MtCM^{DS}–TSA complex (PDB ID: 2W1A;²¹ pink), MtCM^{TS2P} (PDB ID: 6YGT, this work; yellow), MtCM^{V55D} (this work; orange), and top-evolved MtCM^V (PDB ID: SMPV;¹² cyan); cartoon representation featuring the H1–H2 loop, with the ligands depicted as sticks. (B) Superimposition of MtCM (PDB ID: 2VKL;²¹ violet, with bound malate in gray sticks) and MtCM in the MtCM–MtDS complex (PDB ID: 2W1A;²¹ pink, with TSA in golden sticks, corresponding to MtCM^{LC}) in cartoon representation, with the catalytically important Arg46 depicted as sticks. MtDS binding promotes the catalytically competent conformation of Arg46. Helix H2 was removed for clarity.

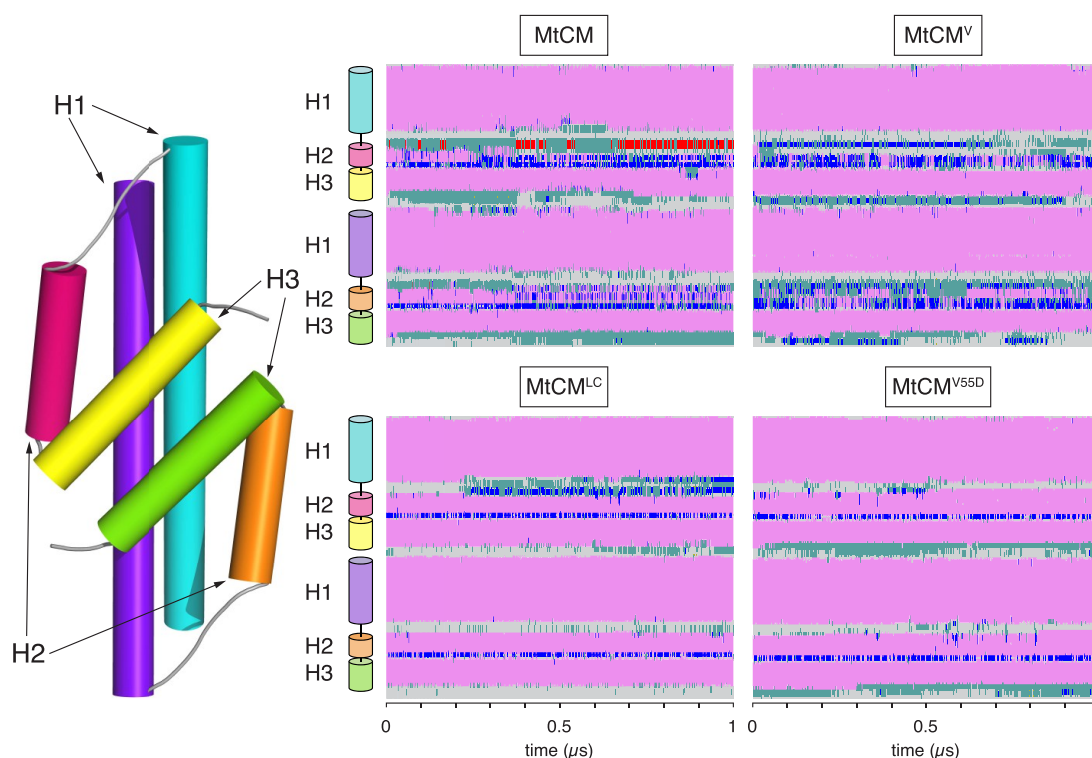


Figure 3. Secondary structure changes of MtCM during MD simulations. Estimated secondary structure of MtCM over 1 μ s of MD simulation. Color code: α -helical structure (magenta), 3_{10} helix (blue), π -helix (red), turn (green), coil (gray). The top panels report data for apo MtCM and MtCM^V, showing clear instability of H2. The bottom panels present the data for the holo-MtCM^{LC} system and for the single V55D variant (apo structure), which in contrast retained all secondary structure elements within the simulation time.

(P22₁), where the asymmetric unit contained the biological dimer. The MtCM^{TS2P} structure was refined to 1.6 Å and $R_{\text{work}}/R_{\text{free}}$ values of 24.0/26.5% (Table S1 and Figure S1B), whereas MtCM^{V55D} diffraction data yielded lower-quality electron density, particularly for the H1–H2 loop (Figure S1C,D). Consistent with this, the Wilson B -factor of MtCM^{V55D} is high (57.8 Å²), indicating structural disorder. Refinement of the 2.1 Å MtCM^{V55D} model stalled at $R_{\text{work}}/R_{\text{free}}$ values of 27.6/34.9%, with very high B -factors for H1–H2 loop residues, especially for protomer B. For both structures, residues preceding residue Glu13 and C-terminal to Leu88

showed poorly defined electron density. Therefore, the terminal residues were not included in the final model.

Overall, the crystal structures of both MtCM^{TS2P} (PDB ID: 6YGT) and MtCM^{V55D} are very similar to the structure of substrate-free wild-type MtCM (PDB ID: 2QBV),⁴¹ with RMSD = 0.3 and 0.4 Å, respectively. However, the H1–H2 loops (⁴⁷MASGGPRLDHS⁵⁷) of both protomers of MtCM^{V55D} adopt a different conformation (RMSD = 2.3 Å compared to PDB ID: 2QBV), which most closely resembles the kinked conformation in the MtCM–MtDS complex (PDB ID: 2W19;²¹ RMSD = 0.8 Å) (Figure 2A). In the crystal structure

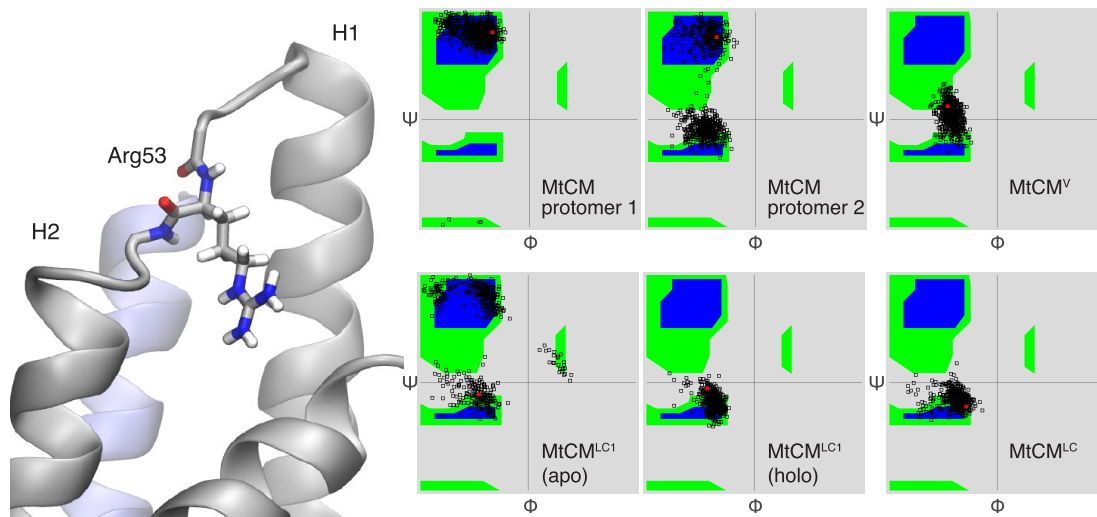


Figure 4. Conformation of Arg53 in the H1–H2 loop. Ramachandran plot showing backbone dihedral angles Φ and ψ for Arg53. Red dots mark starting conformations. For MtCM, the H1–H2 loops from the two protomers (left and middle plots in the top row) assume different ensembles of conformations; overall, the catalytically favored conformation ($\Psi \sim 0$) is observed less frequently than the nonproductive one. In contrast, in MtCM^V, both loops retained the active conformation during the entire course of the simulation, similarly to that observed for MtCM^{LC}. When only one ligand was bound (MtCM^{LC1}), the TSA-loaded site (holo) retained the active conformation, while the loop in the other protomer (apo) remained flexible.

of MtCM^{VSSD}, Asp55 in the H1–H2 loop forms a salt bridge with Arg46, similar to the one in MtCM^V (compare Figure S1E,G,H). This interaction preorganizes the active sites of both MtCM variants for catalytic activity, mimicking MtCM in the complex with MtDS (Figure 2B). However, the overall conformation of the active site loop, which is involved in extensive crystal contacts that are highly distinct for the different crystal forms (Figure S2), differs significantly between the structures (Figures S1 and 2A).

MD Simulations. To evaluate the behavior of MtCM in the absence of crystal contacts, we probed the MtCM structures by MD simulations. We used four model systems: low-activity apo wild-type MtCM, MtCM^{LC} (“ligand complex”: wild-type MtCM from the MtCM–MtDS structure in complex with TSA, the transition state analog of the CM reaction;⁵¹ Scheme 1 and Figure 1), MtCM^V, corresponding to the highly active evolved variant N-s4.15,¹² and MtCM^{VSSD}, which shows the highest catalytic activity among the single-substitution MtCM variants.¹² We compared the overall dynamic profiles of these models and inspected the interactions formed between the C-termini and the H1–H2 loops covering the active sites, to find general features that could be associated with increased catalytic competence.

Apo Structures of MtCM Are Characterized by Significant Flexibility. We anticipated that the model systems would more or less retain the same fold as observed in the crystal structures, but that regions associated with crystal contacts, like the C-termini and the H1–H2 loop, would rapidly move away from their starting positions. Instead, the MD simulations revealed large changes from the initial crystal geometries in the apo protein structures, causing a rather high root-mean-square deviation (RMSD) from the original crystal structure geometry for the CM core regions (RMSD = 2.8 ± 1.2 Å (MtCM) or 3.4 ± 1.5 Å (MtCM^V)). In particular, helix H2 showed a tendency to unravel (Figure 3). Due to the large flexibility observed, the two protomers making up the biological dimer instantaneously broke their symmetry,

independently exploring different conformations in two chains. In contrast, the ligand-bound structure MtCM^{LC} retained the secondary structure throughout the 1 μ s simulation (Figure 3), with a lower RMSD (1.7 ± 0.6 Å) than the two apo structures. Intriguingly, a similar stabilization was observed for the unliganded variant MtCM^{VSSD} (Figure 3).

Kinked Conformation of the H1–H2 Loop. One of the biggest conformational changes in the crystal structure upon formation of the MtCM–MtDS complex occurs in the H1–H2 loop (Figures 1C and 2A).²¹ Whereas in the X-ray structure of the MtDS-activated MtCM, the H1–H2 loop is strongly kinked, this is not the case in nonactivated MtCM. We investigated the conformational landscape of this loop by simulations, using Arg53 from the loop as reporter residue. As shown in Figure 4, in one of the two protomers of MtCM, Arg53 remained in an extended conformation for the entire 1 μ s MD simulation. In contrast, the same amino acid in the other protomer oscillated between the extended and the helical region of the Ramachandran plot (Figure 4), the latter being characteristic of the catalytically active conformation of the loop. Statistically averaging the two distributions, it appears that the apo form of MtCM is preferentially found in its inactive conformation, whereas in MtCM^V both protomers assumed the kinked active loop conformation, and retained it for the whole length of the simulation. However, TSA binding promoted the active conformation also in wild-type MtCM (represented by MtCM^{LC}). The fact that the fluctuations of the MtCM^V H1–H2 loop are contained within the conformational basin of the catalytically competent geometry (Figure 4 and Table 1) is an indication that MtCM^V has an intrinsically preorganized loop, a condition that helps to minimize the entropy loss during substrate binding and consequently favors catalysis.

To test the effect of ligand binding, we repeated simulations of MtCM loaded with only one TSA ligand (MtCM^{LC1}). Interestingly, ligand presence in one of the two binding pockets was sufficient to stabilize the structure of the whole

Table 1. Root-Mean-Square Fluctuation (RMSF) Values of Selected Active Site Residues^a

	RMSF (Å)			
	MtCM ^{WT}	MtCM ^V	MtCM ^{LC}	MtCM ^{VSSD}
Arg18'	2.2 ± 0.5	1.5 ± 1.0	0.7 ± 0.2	1.1 ± 0.5
Arg35	0.8 ± 0.4	0.7 ± 0.4	0.4 ± 0.1	0.6 ± 0.2
Arg46	2.0 ± 0.6	1.7 ± 0.7	0.7 ± 0.5	0.9 ± 0.5
Arg58	2.3 ± 1.0	1.9 ± 2.0	2.0 ± 0.9	1.5 ± 0.6

^aRMSF values were calculated as an average over all nonhydrogen atoms for each residue compared to the average structure of the simulation. The reported σ values reflect the different relative fluctuations of the individual atoms composing the residues in the two symmetric protomers.

dimer. Nonetheless, the H1–H2 active site loop of the apo protomer retained its intrinsic flexibility (Figure 4). The fact that the active site loop of the unloaded protomer behaved like the apo MtCM system suggests that the two active sites in MtCM retain considerable independence.

Contrary to MtCM and MtCM^V (Figure 1C,D), the presence of the additional carboxylate group in MtCM^{VSSD} promoted the elongation of helix H2, resulting in a significant shortening of the active H1–H2 loop (Figure 5A). This structural rearrangement is associated with the formation of persistent salt bridges between Asp55, now localized in the first turn of H2, and active site residues Arg18' and Arg46 that were retained for the entire length of the simulation. Noticeably, in MtCM^{WT}, where such stabilizing electrostatic interactions are absent, no such contacts were observed, with the side chain of Val55 keeping a distance of more than 10 Å from the side chains of both Arg18' and Arg46 for the whole duration of the simulation.

Overall, MtCM^{WT} shows a noisier RMSF profile over the whole amino acid sequence compared to MtCM^V and to the ligand complex MtCM^{LC} (Figure S3). This result reflects the expected rigidification occurring upon substrate binding due to additional protein–ligand interactions in MtCM^{LC}. MtCM^V accomplishes rigidification as a direct consequence of its evolved sequence. Interestingly, also MtCM^{VSSD} shows generally dampened fluctuations, possibly due to the extended helical motif observed in that structure.

Positioning of Active Site Residues. The MtCM active site contains four arginine residues (Figure 1F), among them the key catalytic residue Arg46. In contrast to the observation in the two MtCM–MtDS crystal structures, the conformation of Arg46 was not strictly maintained during MD simulations. In the absence of a ligand, Arg18', Arg46, and Arg58 repelled each other, and at least one of the residues was pushed out of the active site in the majority of the simulations. Only one of the four arginine residues (Arg35) maintained its position (Table 1), appropriately placed for substrate binding by wild-type MtCM, with an RMSF below 1 Å, while RMSF values >2 Å for the other Arg residues signal a substantial increase in the conformational freedom. This changes upon complex formation with MtDS, guiding also the important Arg46 into a catalytically competent conformation.

In contrast to MtCM^{WT}, the two variants MtCM^V and MtCM^{VSSD} exhibited lower RMSF values for all active site Arg residues (Table 1 and Figure 1F) and maintained their catalytically competent conformation during the MD simulations even in the apo forms (Figure 5). The more stable positioning of Arg18' and Arg46 appears to be a direct

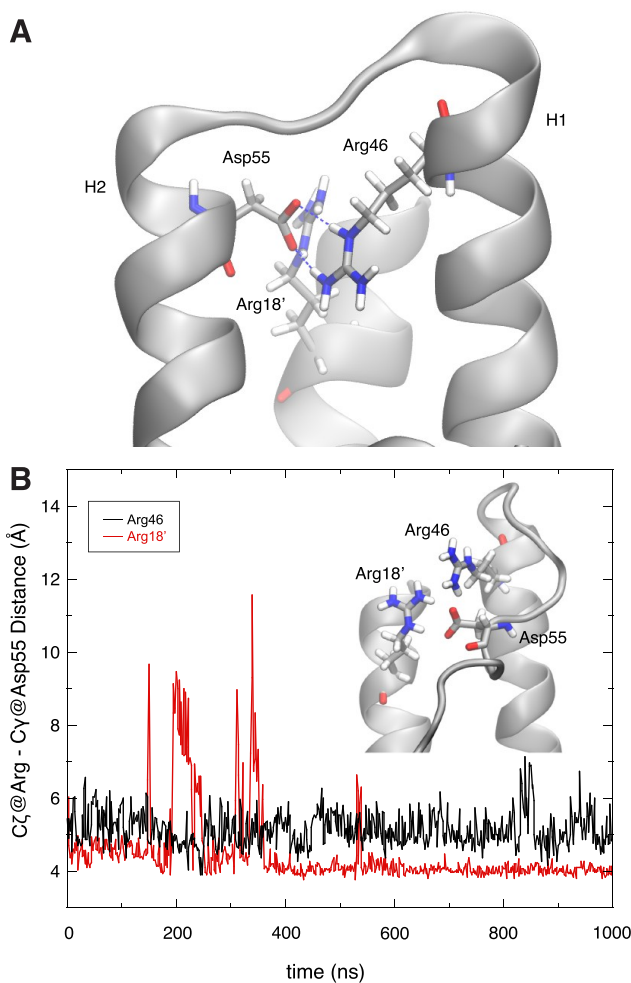


Figure 5. Role of MtCM^V residue Asp55 in positioning active site residues. (A) Extension of H2 and stabilization of the H1–H2 loop by residue Asp55. Substitution of Val55 by Asp stabilizes helix H2 through interactions with Arg18' and Arg46 across the active site (the image shows the structure of V55D after 1 μ s of MD simulations). Note that Arg46 is a catalytically essential residue for MtCM and its correct orientation is critical for catalytic proficiency. (B) Distance plotted between MtCM^V Arg46 (black, chain A) or Arg18' (red, from chain B), and Asp55 (chain A) observed during the simulation. In both cases, the distance measured is between Asp C γ and Arg C γ using PDB nomenclature.

consequence of the replacement of Val55 with Asp, which introduces a negative charge, mitigating the surplus positive charges in the active site.

Interactions between C-Terminal Residues and H1–H2 Loop. A crucial factor for the enhanced activity of MtCM in the MtCM–MtDS complex is an MtDS-induced interaction between MtCM's H1–H2 loop and its C-terminus.²¹ The interaction can be divided into two contributions: a salt bridge between the C-terminal carboxylate and the side chain of Arg53, and a hydrophobic contact between Leu54 and Leu88 (Figure S4A).

Our 1 μ s-long simulations detected persistent, multiple interactions involving the C-terminal carboxylate. In contrast, the hydrophobic contacts between Leu54 and Leu88 were disrupted in the first nanoseconds, and almost never observed again during the rest of the simulation time (Figure S4B,C).

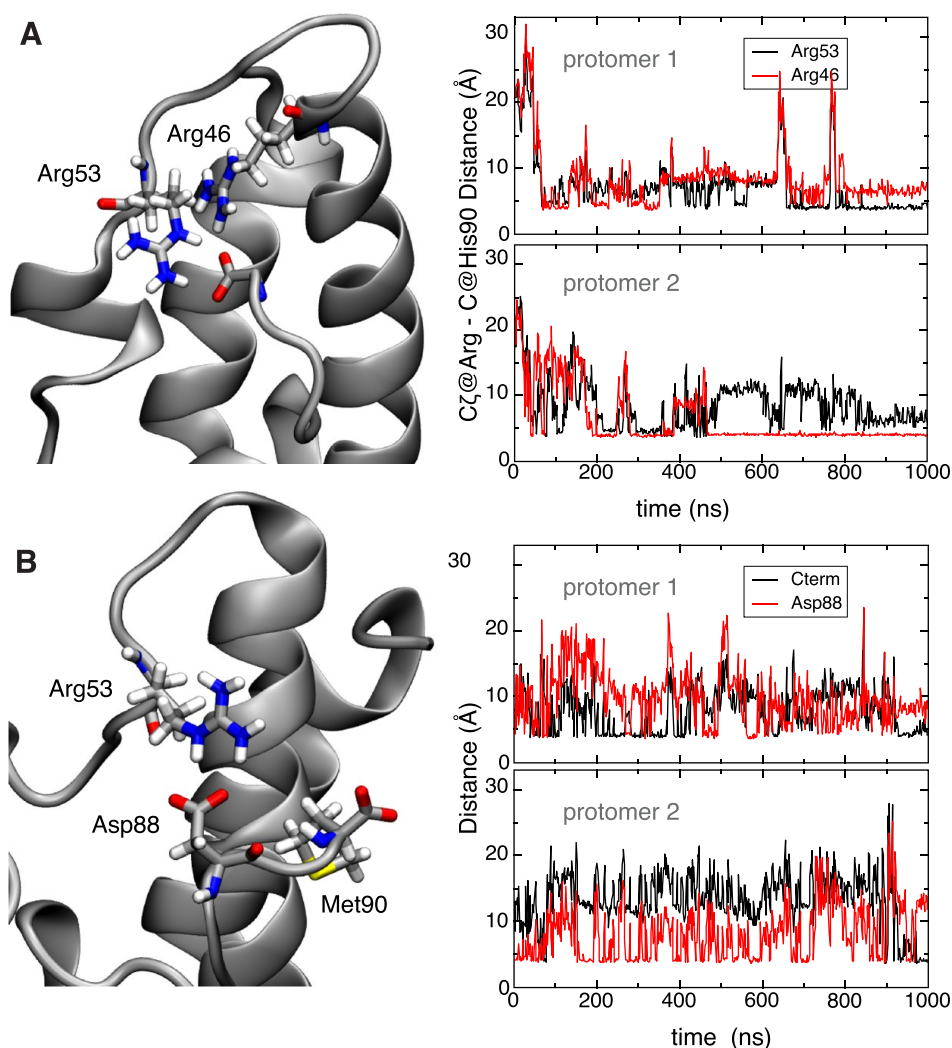


Figure 6. Interaction between C-terminal carboxylate of MtCM and H1–H2 loop. (A) Right panels show the distance between the C_{ζ} carbon of arginine residues 53 or 46 and the carboxylate carbon of the C-terminus of MtCM in the two protomers (top and bottom panels). Formation of a steady contact (<5 Å) with Arg46 (bottom panel) corresponds to stabilization of the catalytically productive H1–H2 loop conformation, which allows for stabilization of the transition state of the chorismate to prephenate rearrangement (Figure 1B,F). (B) Interaction between C-terminal residues and the H1–H2 loop in MtCM^V. Salt bridge contacts between Arg53 and the carboxyl groups of Asp88 (red line) and Met90 (C-terminus; black line) in MtCM^V in the two protomers. The top and bottom panels on the right show the evolution of the distances over time between Arg53's C_{ζ} and the corresponding carboxylate carbons for each of the two protomers of MtCM^V.

Salt Bridges with C-Terminus. In our MD simulations, the C-terminal carboxylate formed interchangeable contacts with Arg53 and the catalytically important Arg46,²¹ which is located in the last turn of helix H1 (Figure 6A). Notably, the presence of a salt bridge between Arg46 and the C-terminus correlated with the apparently active conformation of the H1–H2 loop (Figure 6A).

The observed fluctuations suggest that the catalytically competent conformation of the binding site is malleable in wild-type MtCM and that additional interactions, *i.e.*, with the substrate, are required to stabilize it. This is in line with studies of a topologically redesigned monomeric CM from *Methanococcus jannaschii*. This artificial enzyme was found to be catalytically active in the presence of the substrate despite showing extensive structural disorder without a ligand, reminiscent of a molten globule.⁵²

MtCM^V Exhibits Strengthened Interactions between C-Terminus and H1–H2 Loop. In MtCM^V, the four C-terminal

residues Arg–Leu–Gly–His (RLGH) are substituted with Pro–Asp–Ala–Met (PDAM) at positions 87–90, which include another carboxylate, introduced through Asp88. Our MD simulations show that the Asp88 carboxylate in the evolved variant MtCM^V offers an alternative mode of interaction with Arg53 of the H1–H2 loop (Figures 6B and S5), which is not possible for wild-type MtCM. This allows for a persistent interaction of C-terminal residues with the H1–H2 loop throughout the simulation, while maintaining a highly flexible C-terminus. Moreover, in MtCM^V, Arg46 is topologically displaced from its original position with respect to the loop and no longer able to engage in a catalytically unproductive salt bridge with the C-terminus.

Another interesting substitution, which emerged within the four C-terminal residues during the laboratory evolution toward variant MtCM^V, is a proline residue (RLGH to PDAM).¹² However, in contrast to Pro52, Pro87 did not appear to have a major influence on the simulations. While

Table 2. Catalytic CM Activities of Purified MtCM Variants to Experimentally Address the Importance of the Leu88 to Asp88 Substitution that Emerged during the Directed Evolution of MtCM^V

Variant	Residue changes	k_{cat} (s ⁻¹) ^a	K_{m} (μM) ^a	$k_{\text{cat}}/K_{\text{m}}$ (M ⁻¹ s ⁻¹) ^a
MtCM ^V (PDAM)	PD/PDAM ^b , V62I, D72V, V11L, D15V, K40Q	9.4 ± 1.3	22 ± 2	430,000 ± 30,000
MtCM PNAM	MtCM ^V , D88N	7.6 ± 0.2	45 ± 7	170,000 ± 20,000
MtCM PLAM	MtCM ^V , D88L	6.0 ± 0.4	38 ± 1	160,000 ± 10,000
MtCM ^{WTc}		1.7 ± 0.2	980 ± 80	1700 ± 300
MtCM L88D ^c	L88D	3.0 ± 0.1	1110 ± 70	2700 ± 300
MtCM 3p3	T52P V55D	8.5 ± 0.0	450 ± 70	19,000 ± 3000
MtCM Triple	T52P V55D L88D	11 ± 2	510 ± 130	22,000 ± 1000

^aAll values are experimental means from assays performed with at least two independently produced and purified protein batches with their calculated standard deviations (σ_{n-1}). The $k_{\text{cat}}/K_{\text{m}}$ parameters were obtained as the mean from averaging $k_{\text{cat}}/K_{\text{m}}$ values derived directly from individually fitted independent Michaelis–Menten plots with the calculated error of the corresponding average. ^bPD/PDAM indicates amino acid substitutions T52P, V55D, R87P, L88D, G89A, and H90M. ^cThe default for measuring kinetics involved assays performed in 50 mM K-phosphate, pH 7.5, at 274 nm, whereas the kinetic parameters of these low-performing variants were determined in 50 mM BTP, pH 7.5, at 310 nm. Measuring these variants in 50 mM K-phosphate, pH 7.5, resulted in ~40% reduction in k_{cat} , as was already observed previously for wild-type MtCM.²¹

Pro52 is likely contributing to H1–H2 loop rigidity, with an average RMSF of 1.6 Å in MtCM^V compared to 2.5 Å (MtCM) for this region, the C-termini showed similarly high RMSF values in the two models (>3 Å). Although Pro87 induced a kink at the C-terminus, this did not appear to affect the flexibility of the three terminal residues Asp88–Ala89–Met90.

Kinetic Analysis to Probe Predicted Key Interactions of Engineered MtCM Variants. In the course of the directed evolution of MtCM^V, the L88D replacement was only acquired after the H1–H2 loop-stabilizing substitutions T52P and V55D were already introduced. Guided by the outcome of the MD simulations, we therefore probed the kinetic impact of the innocuous single L88D exchange in the context of three different sets of MtCM variants to experimentally assess the benefit of the introduced negative charge for fine-tuning and optimizing catalytic efficiency. We looked at (i) changing Asp88 in the MtCM^V sequence ⁸⁷PDAM⁹⁰ into Asn88 or Leu88, (ii) directly introducing Asp88 into the MtCM wild-type sequence, and (iii) the triple variant T52P V55D L88D (MtCM Triple). All variants were obtained in their native format, *i.e.*, with their native N-terminus and without a His-tag, to allow for optimal comparison with the structural and computational results. The variants were purified by ion-exchange and size-exclusion chromatography from the *E. coli* host strain KA13, which is devoid of CM genes to rule out contamination by endogenous CMs.^{18,30} Subsequently, the enzymes' kinetic parameters were characterized by a spectrophotometric chorismate depletion assay.

As shown in Table 2, removing the negative charge at residue 88 by replacing Asp with Asn in the top-evolved variant MtCM^V leads to a 2.5-fold drop in the catalytic efficiency $k_{\text{cat}}/K_{\text{m}}$ to $1.7 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$. This decrease is due both to a slightly lower catalytic rate constant (k_{cat}) as well as a reduced substrate affinity (doubled K_{m}). When residue 88 is further changed to the similarly sized but nonpolar wild-type residue Leu88 in variant MtCM PLAM, the catalytic parameters essentially remain the same as for the Asn88 variant (Table 2), independently confirming the catalytic advantage of the negative charge introduced through Asp88.

For the second set of variants that directly started out from the sluggish MtCM wild-type enzyme (MtCM^{WT}), a trend for an increase in catalytic activity upon replacing Leu88 by Asp88 was observed (1.6-fold higher $k_{\text{cat}}/K_{\text{m}}$, reaching $2.7 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$; Table 2). This is mainly caused by an increase in k_{cat} rather

than an altered substrate affinity. Interestingly, the L88D exchange together with T52P and V55D in the MtCM triple variant does not lead to a significant increase in $k_{\text{cat}}/K_{\text{m}}$ compared to MtCM 3p3,¹² which just carries the two loop substitutions T52P and V55D.

Thus, the substitution of Leu88 with Asp88 indeed results in a beneficial effect on the performance of MtCM. However, this effect is only prominent in combination with other selected exchanges, such as those present in MtCM^V. As a single amino acid replacement in the wild-type enzyme or on top of the two substitutions in the H1–H2 loop, the effect of L88D is less noticeable, if present at all.

In summary, a comparison of the dynamic behavior of wild-type MtCM in its apo and ligand-bound states with MtCM^V and MtCM^{V55D} revealed that the catalytically favorable conformation of the active site is achieved by the interplay of several interactions, which balance charges and entropic disorder of the H1–H2 loop. Structuring is promoted, in particular, by increasing the number of the negatively charged carboxylate groups that can both shield the electrostatic charge of the various arginine side chains within or next to the active site and orient catalytically important residues by hydrogen bonding and salt bridge formation. Simulations of MtCM^V revealed the special importance of Asp55 in the V55D variant for coordinating Arg18' and Arg46, thus promoting the preorganization of the active site region. These results echo the conclusions from directed evolution, which also identified the V55D substitution as the most important contributor for catalytic enhancement, causing a 12-fold increase in $k_{\text{cat}}/K_{\text{m}}$.¹² At the same time, we determined and rationalized the more subtle and context-dependent effect of the L88D replacement that introduced an additional negative charge for electrostatic preorganization of the active site. Overall, the high catalytic activity of MtCM^V clearly results from many individual larger and smaller contributions mediated by substitutions at diverse locations within the enzyme structure.

DISCUSSION

Important Activating Factors in MtCM^{DS} and MtCM^V. MtCM has intrinsically low activity but can be activated to rival the performance of the best CMs known to date¹² through the formation of a heterooctameric complex with MtDS,²¹ which aligns crucial active site residues to catalytically competent conformations. Most importantly, binding to MtDS induces preorganization of Arg46 into a catalytically favorable

conformation (Figure 2B), via H-bonding to the carbonyl oxygens of Thr52 and Arg53.²¹ Arg46 is the crucial catalytic residue interacting with the ether oxygen of Bartlett's transition state analogue (TSA)⁵¹ in the complex with MtDS (PDB ID: 2W1A)²¹ (Figures 1B,F and 2B); upon replacing Arg with Lys, the enzyme's efficiency drops 50-fold.²¹

Both MtCM^{DS} and MtCM^V exhibit a kinked H1–H2 loop conformation (Figures 1C,D and 2A), which was hypothesized to be important for increased catalytic efficiency.¹² However, in MtCM^V and MtCM^{V5SD}, the kink is exacerbated by crystal contacts, which are different in the two crystal forms (Figure S2). This kink is much less prominent in wild-type MtCM, or even MtCM^{T52P} (Figures 2A and S1B), and completely lost during the simulations of MtCM^{WT} (we did not carry out simulations on the single variant MtCM^{T52P}). Thus, this conformation may well be a crystallization artifact rather than a prerequisite for an active MtCM.

Nevertheless, preorganization and prestabilization appear to be of crucial importance for the catalytic prowess of MtCM. The largest boost in catalytic efficiency (12-fold enhancement) by a single substitution was observed for the V5SD replacement found in the evolved MtCM^V.¹² This residue is located on the C-terminal side of the H1–H2 loop (Figure 1D,E) and forms a salt bridge to the catalytically important Arg46 at the top of helix H1 (Figure 5A), an interaction that is also observed in the crystal structure of MtCM^{V5SD} (Figure S1G,H). During the MD simulations of MtCM^V and the single variant MtCM^{V5SD}, the presence of Asp55 reduced the mobility of active site residues. By interacting with Arg18' and Arg46, this residue helps to preorganize the active site for catalysis and reduce unfavorable conformational fluctuations caused by electrostatic repulsion in the absence of a substrate. This is supported by the lower RMSF values of MtCM^V compared to uncomplexed wild-type MtCM (Table 1) and by a slightly higher melting temperature of MtCM^{V5SD} ($\Delta T = 3$ °C from differential scanning fluorimetry (DSF) measurements; preliminary data). By decreasing thermal fluctuations in the active site, Asp55 likely also reduces the entropic penalty associated with substrate binding. Pro52 appears to exert a similar stabilizing effect on the protein, despite the rather small structural changes, as suggested by a 2 °C increase in melting temperature of MtCM^{T52P} in DSF experiments compared to MtCM (preliminary data). This single substitution alone raises the $k_{\text{cat}}/K_{\text{m}}$ value of the enzyme by a factor of six.¹² It is worth noting that the simultaneous substitution of T52P and V5SD increased the melting temperature by 6 °C (monitored by circular dichroism spectroscopy) and boosted $k_{\text{cat}}/K_{\text{m}}$ by 22-fold.¹² The top-evolved MtCM^V even showed a melting temperature of 83 °C compared to 74 °C for the parent MtCM.¹²

Importance of the C-Terminus. MtCM activation by MtDS involves a change in conformation of the C-terminus of MtCM and its active site H1–H2 loop.²¹ Specifically, a salt bridge is formed between the C-terminal carboxylate of MtCM (which is repositioned upon MtDS binding) and loop residue Arg53, possibly bolstered by a newly formed hydrophobic interaction between Leu88 and Leu54 (Figure S4A). The 1 μs simulations suggest that salt bridge formation with Arg53 occurs in solution in all tested cases, whereas the hydrophobic contact is less important.

Directed evolution experiments carried out by randomizing the final four C-terminal positions 87–90 of MtCM had previously revealed that a great variety of residues with quite

distinct physico-chemical properties are compatible with a functional catalytic machinery.²⁶ Conserved positions emerged only when probing for an intact activation mechanism by MtDS.²⁶ Still, when residues 87–90 of MtCM^V were evolved from Arg–Leu–Gly–His to Pro–Asp–Ala–Met (Figure 1E), an increase in $k_{\text{cat}}/K_{\text{m}}$ by roughly a factor of four was achieved.¹² Here, we resolved this apparent paradox by investigating C-terminal factors important for the fine-tuned optimization of CM function. Even though the replacement R87P induced a kink in the structure, the presence of the proline did not appear to have a major influence in the simulations. Notably, the C-terminal substitutions together result in a change in net charge from +1 to –2, including the terminal carboxylate, providing the basis for more extensive electrostatic interactions with the positively charged Arg53 than is possible for wild-type MtCM. Indeed, our kinetic analysis of Asp88-containing MtCM variants demonstrates that this residue increases CM's catalytic efficiency (Table 2). The fact that Asp88 did not significantly augment $k_{\text{cat}}/K_{\text{m}}$ in the context of the MtCM double variant T52P V5SD (*i.e.*, MtCM Triple; Table 2) suggests that the extent of catalytic improvement by L88D depends on the particular structural context.

Our simulations indicate that in free wild-type MtCM, an interaction of the C-terminal carboxylate with the key active site residue Arg46 is possible but infrequent due to fluctuations (Figure 6A and Table 1). In contrast, in MtCM^V and MtCM^{DS} the side chain of Arg46 points toward the catalytic pocket (Figures 5 and 2B), and any unproductive reorientation of Arg46 toward the C-terminus would easily result in a clash with the H1–H2 loop. Thus, an additional feature of this loop may be to act as a conditional shield (illustrated for MtCM^V in Figure 7). In the conformation assumed in MtCM^V and MtCM^{DS}, this loop blocks the reorientation of Arg46 toward the C-terminus and hence prevents an unproductive conformation accessible for free wild-type MtCM. MtCM^{DS} and MtCM^V use different means to correctly position active site residues, which correlates with a bent H1–H2 loop in both cases. This is either achieved through conformational changes imposed upon MtCM^{DS} by MtDS binding, or by establishing a salt bridge across the active site, between Arg46 and Asp55, as seen for MtCM^V and also for the single variant MtCM^{V5SD} (Figures 5 and S1E,G,H).

General Implications for CM Catalysis. It is obviously impossible to directly transfer our findings of critical detailed molecular contacts from the AroQ _{δ} subclass CM of *M. tuberculosis* to the evolutionary distinct AroH class CMs, or even to the structurally and functionally divergent AroQ _{α} , AroQ _{β} , and AroQ _{γ} subclasses.⁵³ Neither of those groups of CMs have evolved to be deliberately poor catalysts that become proficient upon regulatory interaction with a partner protein such as MtDS.²¹ To be amenable to 'inter-enzyme allosteric' regulation,²⁸ the H1–H2 loop in MtCM must be malleable and allow for conformational switching between a poorly and a highly active form. In contrast, this region is rigidified in a catalytically competent conformation in the overwhelming majority of CMs from other subclasses. This is exemplified by the prototypic EcCM (AroQ _{α} subclass) and the secreted *MtCM (AroQ _{γ}), which possess the sequence ⁴⁵PVRD⁴⁸ and ⁶⁶PIED⁶⁹, respectively, at the position corresponding to the malleable H1–H2 loop sequence ⁵²TRLV⁵⁵ of wild-type MtCM.¹² Remarkably, the two most impactful substitutions T52P and V5SD occurring during the

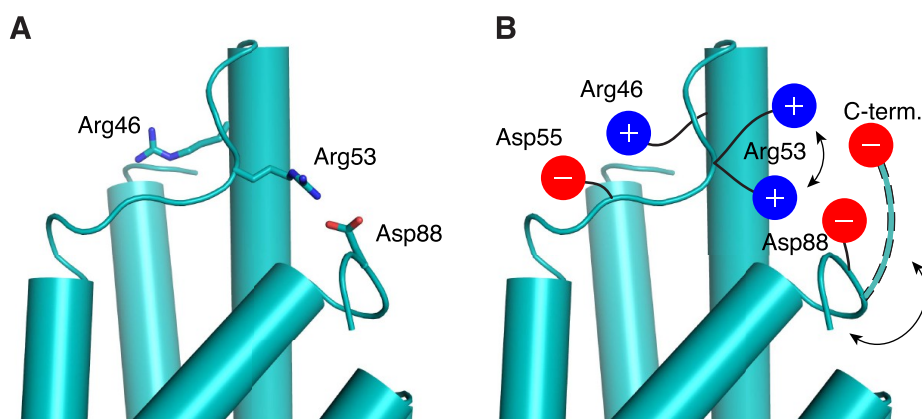


Figure 7. Shielding interaction mediated by the H1–H2 loop. (A) Conformation of important Arg residues in chain A of MtCM^V (cyan) after 31.7 ns of MD simulations. The key active site residue Arg46 is positioned on the opposite side of the H1–H2 loop, which in turn is bolted to the C-terminus by a salt bridge between Arg53 and Asp88 (cartoon representation, with side chains shown as sticks). (B) Cartoon summarizing the important stabilizing interactions in the top-evolved variant MtCM^V depicted in (A) that properly position Arg46 for catalysis. Asp55 stabilizes the stretched-out conformation of Arg46, whereas alternating salt bridges accessible for Arg53 with the negatively charged groups present in the C-terminal region hinder Arg46 from adopting an unfavorable interaction with the C-terminal carboxylate. One example of an alternative backbone conformation that allows for interactions between the C-terminal carboxylate and Arg53 is depicted with dashed outlines.

evolution of MtCM^V have led to the tetrapeptide sequence ⁵²PRLD⁵⁵, with both Pro and Asp being conserved in naturally highly active CMs.¹²

The AroQ₅ subclass CM from *Corynebacterium glutamicum* is another structurally well-characterized poorly active CM ($k_{\text{cat}}/K_{\text{m}} = 110 \text{ M}^{-1} \text{ s}^{-1}$) that requires complex formation with its cognate DAHP synthase for an impressive 180-fold boost in catalytic efficiency.⁵⁴ In that case, inter-enzyme allosteric regulation involves a conformational change of a different malleable segment between helices H1 and H2. Thus, while the molecular details important for the activation of a particular AroQ₅ CM cannot be transferred directly from one system to another, our findings suggest as a general regulatory principle the deliberate and reversible destabilization of a catalytically critical loop conformation.

In both the *M. tuberculosis*¹² and the *C. glutamicum* systems,⁵⁴ crystal contacts in the H1–H2 loop region impede the structural interpretation of the activity switching. The MD simulations shown here represent an interesting alternative approach to dynamic high-resolution structure determination methods for sampling the conformational space adopted by malleable peptide segments with and without ligands.

CONCLUSIONS

MD greatly aided the analysis of crystal structures that were compromised or biased by extensive crystal contacts at the most interesting structural sites. Our aim was to obtain insight into the crucial factors underlying CM activity by comparing the structure and dynamics of the poorly active wild-type MtCM ($k_{\text{cat}}/K_{\text{m}} = 1.7 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$) with the top-performing MtCM variant MtCM^V ($k_{\text{cat}}/K_{\text{m}} = 4.3 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$), which emerged from directed evolution experiments. Both in MtDS-activated wild-type MtCM and in MtCM^V, high activity correlated with a kinked H1–H2 loop conformation and an interaction of this region with the C-terminus of MtCM. The autonomously fully active variant MtCM^V had amino acid changes in both of these regions that augment these structural features. In this report, we focussed on substitutions T52P, V55D, and L88D.

The active site of all natural CMs contains a high density of positive charges. In MtCM, four arginine residues (Arg18', Arg35, Arg46, and Arg58, of which Arg18' is contributed by a different MtCM protomer) are responsible for binding and rearranging the doubly negatively charged substrate chorismate. Only one of these residues (Arg35) is firmly in position before the substrate enters the active site. Of critical importance for catalysis is Arg46. During the MD simulations, Arg46 competes with another arginine residue (Arg53) for binding to the C-terminal carboxylate (Figure 6A) and adopts a catalytically unproductive conformation unless an aspartate residue (Asp55 or Asp88) comes to its rescue. As shown here, Asp55 not only properly orients Arg46 for catalysis but additionally stabilizes the active site. Together with T52P, which preorders the H1–H2 loop, the V55D exchange results in reduced mobility of residues in the active site through stabilizing interactions, thereby preorganizing it for efficient catalysis and lowering the entropic cost of substrate binding. Another aspartate residue (Asp88), also acquired in the top-evolved MtCM^V,¹² helps to balance charges, and—by interacting with Arg53—imposes a steric block that prevents nonoptimal positioning of Arg46 (Figure 7), explaining why the L88D exchange can increase $k_{\text{cat}}/K_{\text{m}}$ by about 2- to 3-fold.

In summary, we tested our hypotheses on the specific importance of critical substitutions acquired during the directed evolution of MtCM^V, namely, T52P, V55D, and L88D by investigating single variants as well as combinations with other residue replacements that were found to augment catalysis. The variants were characterized by crystallography, MD simulations, and enzyme kinetics. The two residues Pro52 and Asp55 exert a major impact by prestabilization and preorganization of catalytically competent conformations of active site residues, while Asp88 contributes to fine-tuning and optimizing the catalytic process. By expanding on the previous directed evolution studies, we have shown here how the accumulated set of amino acid substitutions found in MtCM^V has resulted in an activity level matching that of the most active CMs known to date.¹²

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.biochem.2c00635>.

Crystal structures and electron density maps shown for catalytically important regions (Figure S1); crystal contacts of the H1–H2 loop (Figure S2); root-mean-square fluctuations of MtCM during MD simulations (Figure S3); interactions in MtCM between its C-terminus and H1–H2 loop (Figure S4); MD snapshots of interactions between C-terminus and H1–H2 loop of MtCM^V (Figure S5); and data collection and refinement statistics (Table S1) (PDF)

Accession Codes

UniProt ID: P9WIC1 (Gene Rv0948c). PDB ID: 6YGT.

■ AUTHOR INFORMATION

Corresponding Authors

Peter Kast – Laboratory of Organic Chemistry, ETH Zürich, CH-8093 Zürich, Switzerland; orcid.org/0000-0002-0209-8975; Email: kast@org.chem.ethz.ch

Michele Cascella – Department of Chemistry and Hylleraas Centre for Quantum Molecular Sciences, University of Oslo, Oslo 0315 NO, Norway; orcid.org/0000-0003-2266-5399; Email: michele.cascella@kjemi.uio.no

Ute Krengel – Department of Chemistry and Hylleraas Centre for Quantum Molecular Sciences, University of Oslo, Oslo 0315 NO, Norway; orcid.org/0000-0001-6688-8151; Email: ute.krengel@kjemi.uio.no

Authors

Helen V. Thorbjørnsrud – Department of Chemistry and Hylleraas Centre for Quantum Molecular Sciences, University of Oslo, Oslo 0315 NO, Norway

Luca Bressan – Laboratory of Organic Chemistry, ETH Zürich, CH-8093 Zürich, Switzerland

Tamjidmaa Khatanbaatar – Department of Chemistry and Hylleraas Centre for Quantum Molecular Sciences, University of Oslo, Oslo 0315 NO, Norway

Manuel Carrer – Department of Chemistry and Hylleraas Centre for Quantum Molecular Sciences, University of Oslo, Oslo 0315 NO, Norway

Kathrin Würth-Roderer – Laboratory of Organic Chemistry, ETH Zürich, CH-8093 Zürich, Switzerland; orcid.org/0000-0002-0233-9863

Gabriele Cordara – Department of Chemistry and Hylleraas Centre for Quantum Molecular Sciences, University of Oslo, Oslo 0315 NO, Norway; orcid.org/0000-0001-8029-8043

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.biochem.2c00635>

Author Contributions

^{||}H.V.T., L.B., and T.K. contributed equally to this work.

Author Contributions

U.K. conceived the study. H.V.T., P.K., and Mi.C. were additionally involved in the planning of the experiments. H.V.T. performed most of the calculations, transformed, produced, purified, and crystallized the two single MtCM variants, and solved the crystal structure of MtCM^{V55D}, supervised by Mi.C. and U.K., respectively. Ma.C. contributed with additional simulations, supervised by Mi.C. T.K. solved

the crystal structure of MtCM^{T52P} and refined the crystal structures of both MtCM variants, supervised by G.C. and U.K., who also validated the structures. L.B. constructed, produced, and purified additional sets of MtCM variants and characterized their kinetic parameters to validate computational results, and K.W.-R. designed and constructed the MtCM variants T52P and V55D and prepared the final figures; both were supervised by P.K. The initial version of the manuscript was written by H.V.T. and U.K., which was complemented with contributions from all authors and revised by P.K., Mi.C., and U.K.

Funding

This work was funded by grants from the Swiss National Science Foundation to P.K. (grant 310030M_182648), the Norwegian Research Council (grants 247730 and 245828), and CoE Hylleraas Centre for Quantum Molecular Sciences (grant 262695), and through the Norwegian Supercomputing Program (NOTUR) (grant NN4654K). The work was additionally supported through funds from the University of Oslo (position of H.V.T.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Regula Grüninger-Stössel for help with the construction of the T52P and V55D variants of MtCM, and Joel B. Heim for collecting X-ray data for MtCM^{T52P}. The experiments were performed on beamlines ID30A-3/MASSIF-3 and ID29 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. We are grateful to Montserrat Soler Lopez and Daniele De Sanctis at the ESRF for providing assistance in using beamlines ID30A-3/MASSIF-3 and ID29, respectively. Finally, they acknowledge services provided by the MoBIAS facility, Laboratory of Organic Chemistry, ETH Zurich, and by the UiO Structural Biology Core Facilities.

■ ABBREVIATIONS

BTP, 1,3-bis[tris(hydroxymethyl)methylamino]propane; CM, chorismate mutase; DAHP, 3-deoxy-D-arabino-heptulosonate 7-phosphate; IPTG, isopropyl-β-D-thiogalactopyranoside; LC, ligand complex (*i.e.*, complex with TSA); MD, molecular dynamics; MES, 2-(N-morpholino)ethanesulfonic acid; MtCM, chorismate mutase from *M. tuberculosis*; MtCM^{DS}, MtCM from MtCM–MtDS complex; MtCM^{LC}, TSA-bound MtCM from MtCM–MtDS complex (LC1 refers to only one of the protomers containing TSA); MtCM^{T52P}, MtCM variant T52P; MtCM^V, top-performing MtCM variant N-s4.15 from directed evolution study; MtCM Triple, MtCM variant with the three substitutions T52P, V55D, and L88D; MtCM^{V55D}, MtCM variant V55D; MtCM^{WT}, wild-type MtCM; MtDS, DAHP synthase from *M. tuberculosis*; MWCO, molecular weight cutoff; pI, isoelectric point; PMSF, phenylmethanesulfonyl fluoride; RMSD, root-mean-square deviation (or root-mean-square difference, if concerning structural comparisons); RMSF, root-mean-square fluctuation; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; Tris, tris(hydroxymethyl)aminomethane; TSA, transition state analogue

■ REFERENCES

(1) Ohashi, M.; Liu, F.; Hai, Y.; Chen, M.; Tang, M.-C.; Yang, Z.; Sato, M.; Watanabe, K.; Houk, K. N.; Tang, Y. SAM-dependent

enzyme-catalysed pericyclic reactions in natural product biosynthesis. *Nature* **2017**, *549*, 502–506.

(2) Minami, A.; Oikawa, H. Recent advances of Diels-Alderases involved in natural product biosynthesis. *J. Antibiot.* **2016**, *69*, 500–506.

(3) Tang, M.-C.; Zou, Y.; Watanabe, K.; Walsh, C. T.; Tang, Y. Oxidative cyclization in natural product biosynthesis. *Chem. Rev.* **2017**, *117*, 5226–5333.

(4) Lin, C.-I.; McCarty, R. M.; Liu, H.-W. The enzymology of organic transformations: A survey of name reactions in biological systems. *Angew. Chem., Int. Ed.* **2017**, *56*, 3446–3489.

(5) Sogo, S. G.; Widlanski, T. S.; Hoare, J. H.; Grimshaw, C. E.; Berchtold, G. A.; Knowles, J. R. Stereochemistry of the rearrangement of chorismate to prephenate: Chorismate mutase involves a chair transition state. *J. Am. Chem. Soc.* **1984**, *106*, 2701–2703.

(6) Bentley, R. The shikimate pathway - A metabolic tree with many branches. *Crit. Rev. Biochem. Mol. Biol.* **1990**, *25*, 307–384.

(7) Herrmann, K. M.; Weaver, L. M. The shikimate pathway. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1999**, *50*, 473–503.

(8) Lamb, A. L. Pericyclic reactions catalyzed by chorismate-utilizing enzymes. *Biochemistry* **2011**, *50*, 7476–7483.

(9) Copley, S. D.; Knowles, J. R. The uncatalyzed Claisen rearrangement of chorismate to prephenate prefers a transition state of chairlike geometry. *J. Am. Chem. Soc.* **1985**, *107*, 5306–5308.

(10) Wiest, O.; Houk, K. N. Stabilization of the transition state of the chorismate-prephenate rearrangement: An *ab initio* study of enzyme and antibody catalysis. *J. Am. Chem. Soc.* **1995**, *117*, 11628–11639.

(11) Burschowsky, D.; Kreggel, U.; Uggerud, E.; Balcells, D. Quantum chemical modeling of the reaction path of chorismate mutase based on the experimental substrate/product complex. *FEBS Open Bio* **2017**, *7*, 789–797.

(12) Fahrig-Kamarauskaitė, J.; Würth-Roderer, K.; Thorbjørnsrud, H. V.; Mailand, S.; Kreggel, U.; Kast, P. Evolving the naturally compromised chorismate mutase from *Mycobacterium tuberculosis* to top performance. *J. Biol. Chem.* **2020**, *295*, 17514–17534.

(13) Chook, Y. M.; Ke, H.; Lipscomb, W. N. Crystal structures of the monofunctional chorismate mutase from *Bacillus subtilis* and its complex with a transition state analog. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8600–8603.

(14) Chook, Y. M.; Gray, J. V.; Ke, H.; Lipscomb, W. N. The monofunctional chorismate mutase from *Bacillus subtilis*. Structure determination of chorismate mutase and its complexes with a transition state analog and prephenate, and implications for the mechanism of the enzymatic reaction. *J. Mol. Biol.* **1994**, *240*, 476–500.

(15) Xue, Y.; Lipscomb, W. N.; Graf, R.; Schnappauf, G.; Braus, G. The crystal structure of allosteric chorismate mutase at 2.2-Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10814–10818.

(16) Lee, A. Y.; Karplus, P. A.; Ganem, B.; Clardy, J. Atomic structure of the buried catalytic pocket of *Escherichia coli* chorismate mutase. *J. Am. Chem. Soc.* **1995**, *117*, 3627–3628.

(17) Sträter, N.; Schnappauf, G.; Braus, G.; Lipscomb, W. N. Mechanisms of catalysis and allosteric regulation of yeast chorismate mutase from crystal structures. *Structure* **1997**, *5*, 1437–1452.

(18) MacBeath, G.; Kast, P.; Hilvert, D. A small, thermostable, and monofunctional chorismate mutase from the archaeon *Methanococcus jannaschii*. *Biochemistry* **1998**, *37*, 10062–10073.

(19) Calhoun, D. H.; Bonner, C. A.; Gu, W.; Xie, G.; Jensen, R. A. The emerging periplasm-localized subclass of AroQ chorismate mutases, exemplified by those from *Salmonella typhimurium* and *Pseudomonas aeruginosa*. *Genome Biol.* **2001**, *2*, 1–16.

(20) Ökvist, M.; Dey, R.; Sasso, S.; Grahn, E.; Kast, P.; Kreggel, U. 1.6 Å crystal structure of the secreted chorismate mutase from *Mycobacterium tuberculosis*: Novel fold topology revealed. *J. Mol. Biol.* **2006**, *357*, 1483–1499.

(21) Sasso, S.; Ökvist, M.; Roderer, K.; Gamper, M.; Codoni, G.; Kreggel, U.; Kast, P. Structure and function of a complex between

chorismate mutase and DAHP synthase: Efficiency boost for the junior partner. *EMBO J.* **2009**, *28*, 2128–2142.

(22) Lee, A. Y.; Stewart, J. D.; Clardy, J.; Ganem, B. New insight into the catalytic mechanism of chorismate mutases from structural studies. *Chem. Biol.* **1995**, *2*, 195–203.

(23) Kast, P.; Asif-Ullah, M.; Jiang, N.; Hilvert, D. Exploring the active site of chorismate mutase by combinatorial mutagenesis and selection: The importance of electrostatic catalysis. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 5043–5048.

(24) Kast, P.; Grisostomi, C.; Chen, I. A.; Li, S.; Kreggel, U.; Xue, Y.; Hilvert, D. A strategically positioned cation is crucial for efficient catalysis by chorismate mutase. *J. Biol. Chem.* **2000**, *275*, 36832–36838.

(25) Burschowsky, D.; van Eerde, A.; Ökvist, M.; Kienhöfer, A.; Kast, P.; Hilvert, D.; Kreggel, U. Electrostatic transition state stabilization rather than reactant destabilization provides the chemical basis for efficient chorismate mutase catalysis. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 17516–17521.

(26) Roderer, K.; Neuenschwander, M.; Codoni, G.; Sasso, S.; Gamper, M.; Kast, P. Functional mapping of protein-protein interactions in an enzyme complex by directed evolution. *PLoS One* **2014**, *9*, No. e116234.

(27) Blackmore, N. J.; Nazmi, A. R.; Hutton, R. D.; Webby, M. N.; Baker, E. N.; Jameson, G. B.; Parker, E. J. Complex formation between two biosynthetic enzymes modifies the allosteric regulatory properties of both: An example of molecular symbiosis. *J. Biol. Chem.* **2015**, *290*, 18187–18198.

(28) Munack, S.; Roderer, K.; Ökvist, M.; Kamarauskaitė, J.; Sasso, S.; van Eerde, A.; Kast, P.; Kreggel, U. Remote control by inter-enzyme allostery: A novel paradigm for regulation of the shikimate pathway. *J. Mol. Biol.* **2016**, *428*, 1237–1255.

(29) Sambrook, J.; Russel, D. W. *Molecular Cloning: A Laboratory Manual*, 3rd ed.; Cold Spring Harbor Laboratory Press: New York, 2001.

(30) MacBeath, G.; Kast, P. UGA read-through artifacts—When popular gene expression systems need a pATCH. *Biotechniques* **1998**, *24*, 789–794.

(31) Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, *72*, 248–254.

(32) Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 133–144.

(33) Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69*, 1204–1214.

(34) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 235–242.

(35) Tickle, I. J.; Flensburg, C.; Keller, P.; Paciorek, W.; Sharff, A.; Vonrhein, C.; Bricogne, G. STARANISO; Global Phasing Ltd.: Cambridge, United Kingdom, 2018.

(36) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658–674.

(37) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486–501.

(38) Kovalevskiy, O.; Nicholls, R. A.; Long, F.; Carlon, A.; Murshudov, G. N. Overview of refinement procedures within REFMAC5: Utilizing data from different sources. *Acta Crystallogr. D Struct. Biol.* **2018**, *74*, 215–227.

(39) Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkóczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L.-W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev,

O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallogr. D Struct. Biol.* **2019**, *75*, 861–877.

(40) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(41) Kim, S.-K.; Reddy, S. K.; Nelson, B. C.; Robinson, H.; Reddy, P. T.; Ladner, J. E. A comparative biochemical and structural analysis of the intracellular chorismate mutase (Rv0948c) from *Mycobacterium tuberculosis* H₃₇R_v and the secreted chorismate mutase (γ2828) from *Yersinia pestis*. *FEBS J.* **2008**, *275*, 4824–4835.

(42) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

(43) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

(44) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(45) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*; University of California: San Francisco, 2012.

(46) Jorgensen, W. L.; Madura, J. D. Solvation and conformation of methanol in water. *J. Am. Chem. Soc.* **1983**, *105*, 1407–1413.

(47) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(48) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(49) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.

(50) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(51) Bartlett, P. A.; Johnson, C. R. An inhibitor of chorismate mutase resembling the transition-state conformation. *J. Am. Chem. Soc.* **1985**, *107*, 7792–7793.

(52) Vamvaca, K.; Vögeli, B.; Kast, P.; Pervushin, K.; Hilvert, D. An enzymatic molten globule: Efficient coupling of folding and catalysis. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 12860–12864.

(53) Roderer, K.; Kast, P. Evolutionary cycles for pericyclic reactions – Or why we keep mutating mutases. *CHIMIA* **2009**, *63*, 313–317.

(54) Burschowsky, D.; Thorbjørnsrud, H. V.; Heim, J. B.; Fahrig-Kamarauskaitė, J.; Würth-Roderer, K.; Kast, P.; Kregel, U. Inter-enzyme allosteric regulation of chorismate mutase in *Corynebacterium glutamicum*: structural basis of feedback activation by Trp. *Biochemistry* **2018**, *57*, 557–573.

Recommended by ACS

Transferability of the Electrostatic Parameters of the Polarizable Gaussian Multipole Model

Shiji Zhao, Ray Luo, *et al.*

JANUARY 25, 2023
JOURNAL OF CHEMICAL THEORY AND COMPUTATION

READ 

Phosphorylation Regulation Mechanism of β2 Integrin for the Binding of Filamin Revealed by Markov State Model

Xiaokun Hong, Hai-Feng Chen, *et al.*

JANUARY 06, 2023
JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 

Water Leakage Pathway Leads to Internal Hydration of the p53 Core Domain

Igor D. M. Lima, Elio A. Cino, *et al.*

DECEMBER 19, 2022
BIOCHEMISTRY

READ 

Modeling the Orthosteric Binding Site of the G Protein-Coupled Odorant Receptor OR5K1

Alessandro Nicoli, Antonella Di Pizio, *et al.*

JANUARY 25, 2023
JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 

Get More Suggestions >

SUPPORTING INFORMATION

What drives chorismate mutase to top performance? Insights from a combined *in silico* and *in vitro* study

Helen V. Thorbjørnsrud^{1,2#}, Luca Bressan^{3#}, Tamjidmaa Khatanbaatar^{1,2#}, Manuel Carrer^{1,2}, Kathrin Würth-Roderer³, Gabriele Cordara^{1,2}, Peter Kast^{3*}, Michele Cascella^{1,2*}, Ute Krengel^{1,2*}

¹ Department of Chemistry, University of Oslo, NO-0315 Oslo, Norway

² Hylleraas Centre for Quantum Molecular Sciences, University of Oslo, NO-0315 Oslo, Norway

³ Laboratory of Organic Chemistry, ETH Zurich, CH-8093 Zurich, Switzerland

#H.V.T., L.B. and T.K. contributed equally to this work.

*corresponding authors: kast@org.chem.ethz.ch, michele.cascella@kjemi.uio.no, ute.krengel@kjemi.uio.no

LIST OF MATERIAL INCLUDED:

Figures S1-S5 (S1, Crystal structures and electron density; S2, Crystal contacts; S3 RMSF plot; S4-5, MD snapshots)

Table S1 (Data collection and refinement statistics)

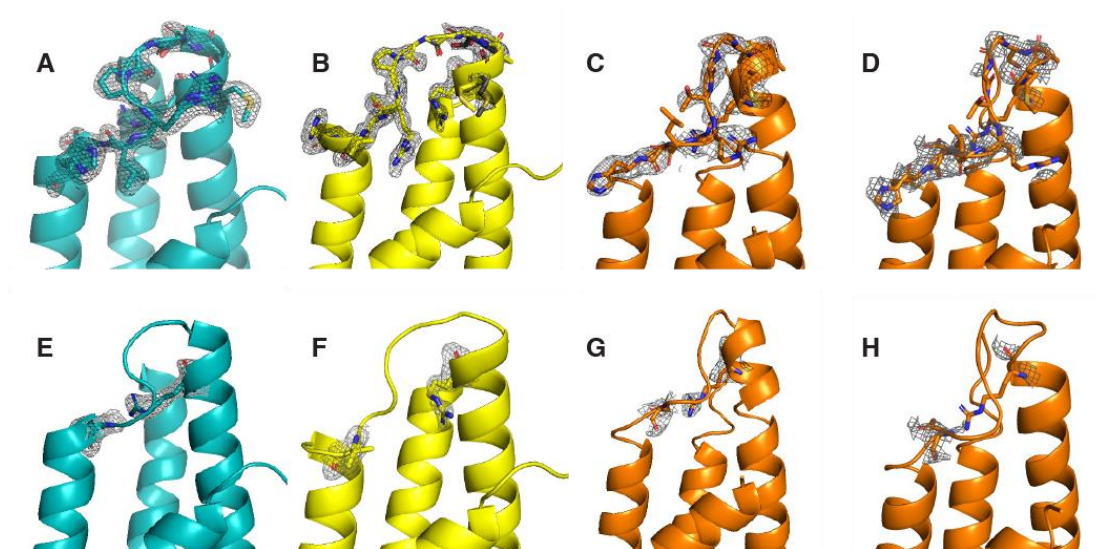


Figure S1. Crystal structures and electron density maps shown for catalytically important regions. Top row, active site H1-H2 loop residues (⁴⁷MASGGPRLVHS⁵⁷). Bottom row, residue 55 and the catalytically crucial Arg46. **(A)** and **(E)**, top-evolved MtCM^V (PDB ID: 5MPV;¹ cyan); **(B)** and **(F)**, MtCM^{T52P} (PDB ID: 6YGT, this work; yellow); **(C)** and **(G)**, MtCM^{V55D}, protomer A (this work; orange); **(D)** and **(H)**, MtCM^{V55D}, protomer B (this work; orange). The MtCM^{V55D} structure is of low quality, precluding final refinement; therefore, the coordinates were not deposited in the PDB. All electron density maps are σ_A -weighted $2mFo-DFc$ maps, depicted at $\sigma = 1.0$.

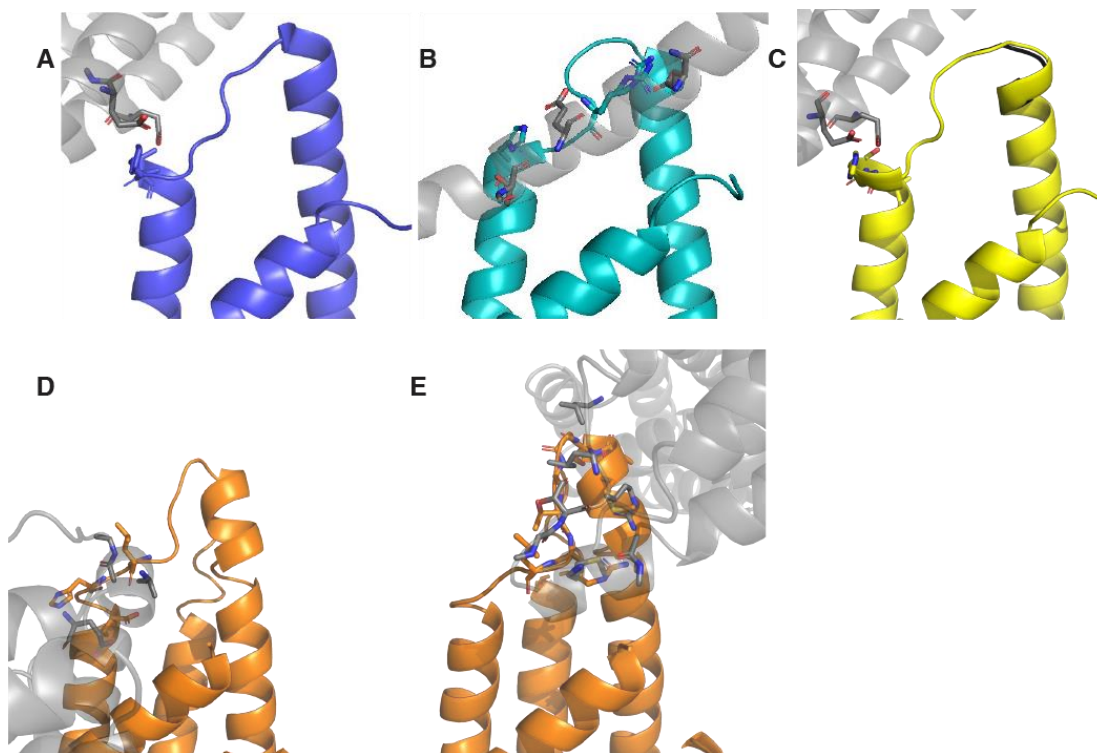


Figure S2. Crystal contacts of the H1-H2 loop with the closest neighboring molecule (shown in grey). **(A)** Wild-type MtCM (PDB ID: 2VKL;² purple), **(B)** Top-evolved MtCM^V (PDB ID: 5MPV;¹ cyan), **(C)** MtCM^{T52P} (PDB ID: 6YGT, this work; yellow), **(D)** MtCM^{V55D}, protomer A (this work; orange), **(E)** MtCM^{V55D}, protomer B (this work; orange). The structures in panels A and C have the same crystal form.

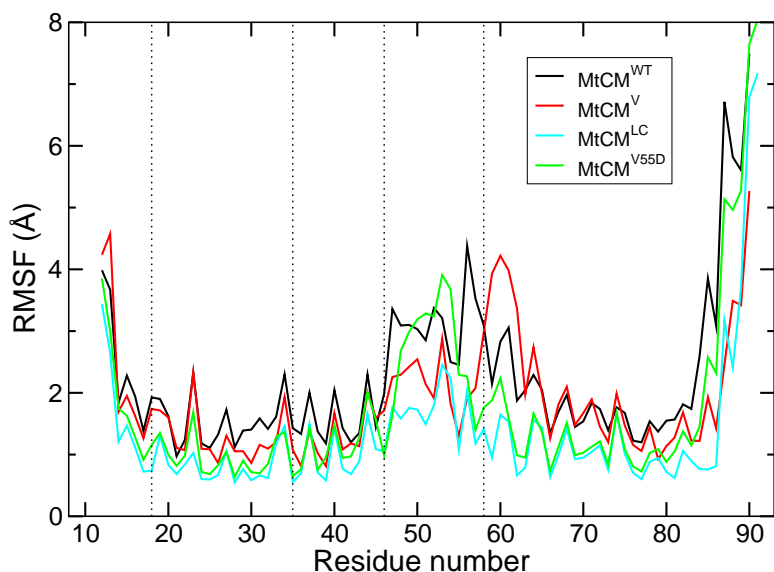


Figure S3. Root mean square fluctuations of MtCM during MD simulations. The graph reports the mean RMSF values averaged per residue for MtCM^{WT} and the super-active variant MtCM^V as well as for the ligand complex MtCM^{LC} and the MtCM^{V55D} single variant. The vertical dotted lines indicate the positions of the four key arginine residues listed in Table 1.

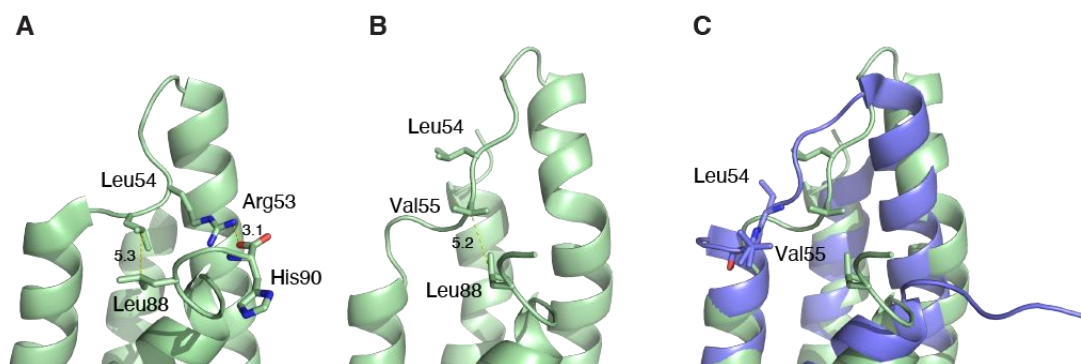


Figure S4. Interactions in MtCM between its C-terminus and H1-H2 loop. (A) Cartoon illustration of MtCM^{DS} (MtCM from the MtCM-MtDS crystal structure; PDB ID: 2W19²). MtCM^{DS} is colored green and residues involved in prominent interactions are shown as sticks. Two interactions are highlighted: hydrophobic contacts between Leu54 of the H1-H2 loop and Leu88 close to the C-terminus of MtCM (distance measured between C_γ of Leu side chains), and a salt bridge between Arg53 of the H1-H2 loop and the C-terminal carboxylate (His90). (B) Cartoon illustration of van der Waals interactions observed for chain A of MtCM^{DS} at time step 250 (25 ns simulation). Note the distinct conformation of the H1-H2 loop, with Val55 temporarily taking over the role of Leu54. However, this conformational change is catalytically unfavorable, as it interferes with substrate binding to the Val55 main chain amide group, as shown in Fig. 1F. (C) Superimposition of MtCM^{DS} (simulated structure from B, green) with MtCM crystal structure (purple, PDB ID: 2VKL²). The corresponding residues Val55 and Leu54 in the two structures, shown as sticks, occupy shifted positions, but in similar (catalytically unfavorable) orientations in the respective H1-H2 loops.

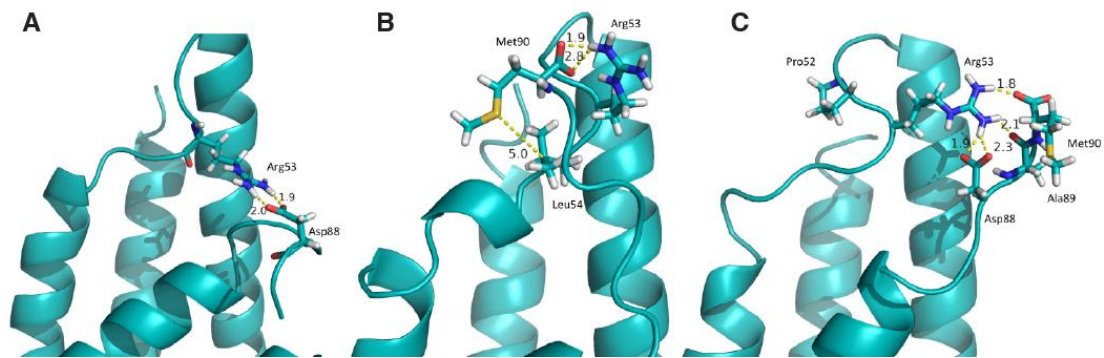


Figure S5. MD snapshots of interactions between C-terminus and H1-H2 loop of MtCMV. The C-terminus is very flexible, but recurring interactions are made by carboxylic acid groups from the C-terminal sequence (C-terminal carboxylate or Asp88 side chain) to Arg53 of the H1-H2 loop. Representative frames are shown in cartoon representation, with relevant residues as sticks. **(A)** Chain A at 15.1 ns, **(B)** Chain B at 17.8 ns, **(C)** Chain A at 4.1 ns (intermediate conformation).

Table S1 – Data collection and refinement statistics

	MtCM ^{T52P}	MtCM ^{V55D}
Data collection		
Beamline	ESRF ID3-A30/MASSIF-3	ESRF ID29
Wavelength (Å)	0.9677	0.9753
Space group	<i>P</i> 4 ₃ 2 ₁ 2	<i>P</i> 2 2 ₁ 2 ₁
Cell parameters - <i>a</i> , <i>b</i> , <i>c</i> (Å)	59.6, 59.6, 46.6	32.2, 59.7, 72.1
Protein chains in a.s.u.	1	2
Matthew's coefficient (Å ³ /Da)	2.0	1.7
Resolution (Å) ^a	36.7-1.64 (1.72-1.64)	45.94-2.06 (2.25-2.06)
CC _{1/2} (%) ^{a, b}	99.5 (45.1)	99.9 (47.6)
Mean I/σ(I) ^a	15.7 (1.1)	14.6 (1.1)
Completeness (%) ^a	88.1 (33.4)	82.4 (22.3)
Number of unique reflections ^a	9570 (479)	7445 (438)
Multiplicity ^a	8.7 (9.8)	6.0 (6.1)
Wilson B-factor (Å ²)	22.2	57.8
Refinement^d		
Resolution range (Å)	29.8-1.64	
<i>R</i> _{work} / <i>R</i> _{free} (%) ^c	24.0/26.5	
Average <i>B</i> -factor (Å ²)	32.6	
Number of atoms		
Protein	674	
Water	18	
r.m.s.d. from ideal geometry		
Bond lengths (Å)	0.01	
Bond angles (deg.)	1.2	
Ramachandran plot ^e		
Favored (%)	93.1	
Allowed (%)	5.2	
Outliers (%)	0.0	
PDB code	6YGT	

^a Values in parentheses refer to highest resolution shell. The data completeness falls below the 95% threshold beyond ~1.85 Å and ~2.40 Å resolution for MtCM^{T52P} and MtCM^{V55D}, respectively, which can be considered the effective resolution of the data sets.

^b Reflections up to the highest resolution limit were included assessing the data using the CC_{1/2} parameter, as suggested by Diederichs and Karplus.^{3,4}

^c *R*_{free} was calculated from 5% of randomly selected reflections for each data set.

^d Refinement of MtCM^{V55D} stalled at *R*_{work}/*R*_{free} values of 27.6/34.9%. The structure was therefore not included in this Table or submitted to the PDB.

^e Calculated with *SFCHECK*.⁵ Ser49 (and Gly51) have torsion angles bordering to outlier regions in the Ramachandran plot, and were identified as outliers by the PDB.

Supporting References

(1) Fahrig-Kamarauskaitė, J.; Würth-Roderer, K.; Thorbjørnsrud, H. V.; Mailand, S.; Krenzel, U.; Kast, P. Evolving the naturally compromised chorismate mutase from *Mycobacterium tuberculosis* to top performance. *J. Biol. Chem.* **2020**, *295*, 17514-17534.

(2) Sasso, S.; Ökvist, M.; Roderer, K.; Gamper, M.; Codoni, G.; Krenzel, U.; Kast, P. Structure and function of a complex between chorismate mutase and DAHP synthase: Efficiency boost for the junior partner. *EMBO J.* **2009**, *28*, 2128-2142.

(3) Diederichs, K.; Karplus, P. A. Better models by discarding data? *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69*, 1215-1222.

(4) Karplus, P. A.; Diederichs, K. Assessing and maximizing data quality in macromolecular crystallography. *Curr. Opin. Struct. Biol.* **2015**, *34*, 60-68.

(5) Vaguine, A. A.; Richelle, J.; Wodak, S. J. *SFCHECK*: A unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. D Biol. Crystallogr.* **1999**, *55*, 191-205.