

Doctoral Thesis in Biotechnology

Symbiotic and pathogenic factors in plant-microbe interaction: Structural basis of C-glycoside metabolism and lipoprotein transport in bacteria

VALENTINA FURLANETTO



Symbiotic and pathogenic factors in plant-microbe interaction: Structural basis of C-glycoside metabolism and lipoprotein transport in bacteria

VALENTINA FURLANETTO

Academic Dissertation which, with due permission of the KTH Royal Institute of Technology, is submitted for public defence for the Degree of Doctor of Philosophy on Friday the 8th of December 2023, at 1:00 p.m. in Kollegiesalen, Brinellvägen 8, Stockholm, Sweden.

Doctoral Thesis in Biotechnology
KTH Royal Institute of Technology
Stockholm, Sweden 2023

© Valentina Furlanetto

Cover page photo: Valentina Furlanetto

ISBN 978-91-8040-755-7

TRITA-CBH-FOU-2023:53

Printed by: Universitetsservice US-AB, Sweden 2023

Abstract

The communication between plants and bacteria involves a complex chemical signaling network that mediates responses to various biotic and abiotic stresses, as well as establishing symbiotic relationships between different organisms.

The first part of the thesis focuses on a mechanism for symbiotic communication between plants and bacteria and more specifically on how C-glycosylated aromatic polyketide compounds produced by plants can be used as a mechanism for plants to communicate with beneficial bacteria. In their glycosylated form, these compounds are substrates for symbiotic bacteria, which in return deglycosylate them and release the sugar-free part, the active aglycone. The aglycone can then mediate several functions beneficial to the plant, for example facilitating nitrogen fixation or acting as an antibacterial agent against plant pathogens.

Results from studies covered in the thesis show that the soil bacteria *Deinococcus aerius*, *Streptomyces canus* and *Microbacterium testaceum* produce enzymes that can cleave the carbon-carbon bond between the sugar and the aglycone in C-glycosyl compounds. Deglycosylation first requires oxidation of the sugar by an oxidoreductase, after which the carbon-carbon bond can be cleaved by a C-glycosyl deglycosidase (CGD). Biochemical and structural characterization as well as results from phylogenetic analyzes of the amino-acid sequences of CGD enzymes provided new knowledge about these relatively unexplored enzymatic processes, as well as increased insight into how C-glycosylated aromatic polyketides participate in the interaction between plant and bacteria.

The second part of the thesis explores pathogenic interactions between plants and bacteria. The virulence of pathogenic bacteria is dependent on lipoproteins that are attached to the bacteria's outer membrane and that have a decisive role for the bacteria's survival and pathogenicity. The localization of lipoproteins takes place through a process abbreviated Lol. The Lol system of the notorious plant pathogen *Xanthomonas campestris* was studied to better understand the underlying molecular mechanisms of the localization system, which could eventually open new ways to combat the bacterium. Biochemical, structural, and phylogenetic techniques were used also in this project.

Taken together, the results contributed several new discoveries. For the first time, a physical complex between the two proteins responsible for

transporting lipoproteins could be determined and their mutual interactions studied. Furthermore, sequence analyses challenge the generally accepted model of how lipoproteins are released from the bacterial inner membrane before being transported to the outer membrane.

According to the standard model based on *Escherichia coli*, lipoproteins are extracted from the inner membrane by a membrane protein that belongs to the ABC-transporter family and whose structure forms an asymmetric heterodimer (LolCDE). However, our bioinformatic analysis show that most of these ABC transporters, including *X. campestris*, are likely to be homodimers and that *Escherichia coli* is the exception rather than the rule. The difference between an asymmetric and symmetric ABC transporter also has implications for several hypotheses about how these proteins function. Heterologous production of the *X. campestris* ABC transporter confirmed that the protein is a homodimer.

Sammanfattning

Kommunikationen mellan växter och bakterier omfattar ett komplext kemiskt signaleringsnätverk som förmedlar svar på olika biotiska och abiotiska påfrestningar, samt etablerar symbiotiska relationer mellan olika organismer.

Den första delen av avhandlingen fokuserar på en mekanism för symbiotisk kommunikation mellan växter och bakterier och mer specifikt på hur C-glykosylerade aromatiska polyketidföreningar producerade av växter kan utnyttjas som en mekanism för växter att kommunicera med goda bakterier. I sin glykosylerade form utgör dessa föreningar substrat för symbiotiska bakterier som i gengäld deglykosylerar dem och frigör den sockerfria delen, det aktiva aglykonet. Aglykonet kan sedan förmedla ett flertal för växten fördelaktiga funktioner, till exempel underlättande av kvävefixering eller fungera som ett antibakteriellt medel mot växtpatogener.

I avhandlingen visas att jordbakterierna *Deinococcus aerius*, *Streptomyces canus* och *Microbacterium testaceum* producerar enzymer som ansvarar för klyning av kol-kol-bindingen mellan sockret och aglykonet i C-glykosylföreningar. Deglykosylering kräver först oxidation av sockret med hjälp av ett oxidoreduktas, varefter kol-kol-bindningen kan klyvas av ett C-glykosyldeglykosidas (CGD). Biokemisk och strukturell karakterisering samt resultat från fylogenetiska analyser av CGD-enzymers aminosyra-sekvenser ger ny kunskap om dessa relativt outforskade enzymatiska processer, samt ökad insikt om hur C-glykosylerade aromatiska polyketider deltar i samspelet mellan växt och bakterie.

Den andra delen av avhandlingen utforskar patogena interaktioner mellan växter och bakterier. Patogena bakteriers virulens är beroende av lipoproteiner som är fästa i bakteriens ytermembran och som har en avgörande roll för bakteriens överlevnad och patogenicitet. Lokaliseringen av lipoproteiner sker genom en process som förkortas Lol. Lol-systemet hos den problematiska växtpatogenen *Xanthomonas campestris* studerades för att bättre förstå de underliggande molekylära mekanismerna hos lokaliseringssystemet, vilket på sikt kan öppna upp för nya sätt att bekämpa bakterien. Även i detta projekt användes biokemiska, strukturella och fylogenetiska tekniker.

Sammantaget bidrog resultaten med flera nya upptäckter. För första gången kunde ett fysiskt komplex mellan de två proteinerna som ansvarar för transport av lipoproteiner bestämmas och deras inbördes interaktioner studeras. Vidare bidrog sekevensanalyser till ifrågasättande av den allmänt vedertagna modellen för hur lipoproteiner frigörs från bakteriens innermembran innan den transporteras till yttermembranet.

Enligt standardmodellen baserad på *Escherichia coli* extraheras lipoproteiner från innermembranet av ett membranprotein som tillhör ABC-transportörfamiljen och vars struktur bildar en asymmetrisk heterodimer. Våra bioinformatikstudie visade dock att merparten av dessa ABC-transportörer, inklusive den i *X. campestris*, mest troligt är homodimerer och att *Escherichia coli* är ett undantag snarare än en regel. Skillnaden mellan en asymmetrisk och symmetrisk ABC-transportör får även konsekvenser för flera hypoteser om hur dessa proteiner fungerar. Heterolog produktion av ABC-transportören från *X. campestris* bekräftade att proteinet är en homodimer.

Riassunto

La comunicazione tra piante e batteri comprende una complessa rete di segnali chimici che mediano le risposte a vari stress biotici ed abiotici, nonché l'instaurazione di relazioni simbiotiche tra diversi organismi.

La prima parte di questa tesi si centra su un meccanismo di comunicazione simbiotica tra piante e batteri, in particolare su come i composti polichetidici aromatici C-glicosilati prodotti dalle piante possano essere utilizzati nel meccanismo di comunicazione tra piante e batteri benefici. In forma glicosilata, questi composti sono substrati per i batteri simbiotici, i quali, a loro volta, li deglicosilano e rilasciano la parte priva di zucchero, l'aglicone attivo. L'aglicone può quindi svolgere diverse funzioni vantaggiose per la pianta, come ad esempio facilitare la fissazione dell'azoto o agire come agente antibatterico contro i patogeni delle piante.

I risultati degli studi presentati in questa tesi dimostrano che i batteri del suolo *Deinococcus aerius*, *Streptomyces canus* e *Microbacterium testaceum* producono enzimi capaci di scindere il legame carbonio-carbonio tra lo zucchero e l'aglicone nei composti C-glicosilati. La deglicosilazione richiede innanzitutto l'ossidazione dello zucchero da parte di un'ossidoreduttasi, dopodiché il legame carbonio-carbonio può essere scisso da una C-glicosildeglicosidasi (CGD). La caratterizzazione biochimica e strutturale, insieme ai risultati delle analisi filogenetiche delle sequenze di amminoacidi degli enzimi CGD, hanno fornito nuove conoscenze su questi processi enzimatici relativamente poco esplorati, espandendo così la comprensione di come i composti polichetidici aromatici C-glicosilati partecipino all'interazione tra piante e batteri.

La seconda parte della tesi esplora le interazioni patogene tra piante e batteri. La virulenza dei batteri patogeni dipende dalle lipoproteine che sono ancorate alla membrana esterna degli stessi che hanno un ruolo determinante per la loro sopravvivenza e la patogenicità. La localizzazione delle lipoproteine avviene attraverso un processo abbreviato come Lol. Il sistema Lol del noto patogeno delle piante *Xanthomonas campestris* è stato studiato in questa tesi al fine di comprendere meglio i meccanismi molecolari sottostanti a questo sistema di localizzazione. Questi studi forniscono nuove informazioni necessarie a combattere il batterio. Anche in questo progetto sono state utilizzate tecniche biochimiche, strutturali e filogenetiche.

Nel complesso, i risultati hanno contribuito a diverse nuove scoperte. Per la prima volta, è stato possibile determinare la struttura del complesso tra le due proteine responsabili del trasporto delle lipoproteine e studiare le loro interazioni reciproche. Inoltre, le analisi delle sequenze proteiche mettono in discussione il modello generalmente accettato su come le lipoproteine siano rilasciate dalla membrana interna del batterio prima di essere trasportati alla membrana esterna.

Secondo il modello standard basato su *Escherichia coli*, le lipoproteine sono estratte dalla membrana interna grazie ad una proteina di membrana appartenente alla famiglia di trasportatori ABC e la cui struttura forma un eterodimero asimmetrico (LolCDE). Tuttavia, le nostre analisi bioinformatiche dimostrano che la maggior parte di questi trasportatori ABC, compreso quello in *X. campestris*, sono probabilmente omodimeri e che *Escherichia coli* rappresenta l'eccezione piuttosto che la regola. La differenza tra un trasportatore ABC asimmetrico e uno simmetrico ha anche implicazioni per diverse ipotesi su come queste proteine operino. La produzione eterologa del trasportatore ABC del batterio *X. campestris* ha confermato che la proteina è un omodimero.

List of appended papers

Paper I

Furlanetto V*, Kalyani DC*, Kostelac A, Haltrich D, Hällberg BM, Divne C (2023). Structural and functional characterization of the gene cluster responsible for deglycosylation of C-glucosyl flavonoids and xanthonoids by *Deinococcus aerius*. Manuscript, to be submitted.

*Shared first authorship.

Paper II

Kostelac A, Sützl L, Puc J, **Furlanetto V**, Divne C, Haltrich D (2022). Biochemical characterization of pyranose oxidase from *Streptomyces canus*-towards a better understanding of pyranose oxidase homologues in bacteria. *Int. J. Mol. Sci.* 23:13595.

<https://doi.org/10.3390/ijms232113595>

Co-authored conference proceedings for OxiZymes meeting 2022.

Paper III

Furlanetto V, Kalyani DC, Srivastava V, Hällberg BM, Divne C. Deglycosylation of C-glycosylflavonoids by the endophytic bacterium *Microbacterium testaceum*. Manuscript, to be submitted.

Paper IV

Furlanetto V, Divne C (2023). LolA and LolB from the plant-pathogen *Xanthomonas campestris* forms a stable heterodimeric complex in the absence of lipoprotein. *Front. Microbiol.* 14:1216799.

<https://doi.org/10.3389/fmicb.2023.1216799>

Paper V

Furlanetto V, Divne C. Experimental evidence for the lipoprotein extractor complex in *Xanthomonas campestris* pv. *campestris* forming a LolFD homodimer and comparative structural modeling of the LolFD complex. Manuscript, to be submitted.

Respondent's contribution to the included papers

Paper I

Conceptualization, validation, cloning experiments, gene expression, protein purification and crystallization, synchrotron data collection, structure determination, model refinement, cryo-EM grid preparation, cryo-EM screening and data collection, model building, structural analysis, model validation, visualization, manuscript writing.

Paper II

Contribution to validation, structural analysis.

Paper III

Conceptualization, validation, crystallization, synchrotron data collection, structure determination, model building, model refinement, cryo-EM grid preparation, cryo-EM screening and data collection, cryo-EM data processing, structural analysis, model validation, visualization, manuscript writing.

Paper IV

Conceptualization, validation, cloning experiments, gene expression, protein purification and crystallization, synchrotron data collection, structure determination, model building, model refinement, structural analysis, model validation, design and execution of interaction experiments, visualization, manuscript writing.

Paper V

Conceptualization, validation, cloning experiments, gene expression, protein purification, comparative template-based structure modeling and validation, structural analysis, visualization, manuscript writing.

Published papers not included in the thesis

Chang SC, Kao MR, Karmakar Saldivar R, Diaz-Moreno S, Zing X, Furlanetto V, Tayo J, Divne C, Vilaplana F, Abbott DW, Hsieh YSY (2023). The Gram-positive bacterium *Romboutsia ilealis* harbors a polysaccharide synthase that can produce (1,3;1,4)- β -D-glucans. *Nat. Commun.* 14:4526. <https://doi.org/10.1038/s41467-023-40214-z>

Humer D, Furlanetto V, Schruet AK, Wlodarczyk A, Kuttke M, Divne C, Spadiut O (2021). Potential of unglycosylated horseradish peroxidase variants for enzyme prodrug cancer therapy. *Biomed. Pharmacother.* 142:112037. <https://doi.org/10.1016/j.biopha.2021.112037>

List of abbreviations

AA3	Auxiliary Activity Family 3
ABC	ATP-binding cassette
ADP	adenosine diphosphate
AMP-PNP	adenylyl-imidodiphosphate
ATP	adenosine triphosphate
CGD	CGD, C-glycosyl/glycoside deglycosidase
CGF	C-glycosylflavonoids
CGO	C-glycosyl/glycoside oxidoreductase
cryo-EM	cryo-electron microscopy
CSS	complexation significance score
CTF	contrast-transfer-function
ESI	electrospray Ionization
ETI	effector-triggered immunity
ETS	effector-triggered susceptibility
FAD	flavin adenine dinucleotide
FSC	Fourier Shell Correlation
FT	Fourier transform
G+	Gram-positive
G-	Gram-negative
GFO	glucose-fructose oxidoreductase
GIM	GFO/IDH/MocA
GMC	Glucose/Methanol/Choline
HPLC	High-performance liquid chromatography
HR	hypersensitive response
IDH	inositol 2-dehydrogenase
IM	inner membrane
IMAC	Immobilized metal-ion affinity chromatography
IPTG	isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
LIC	ligation-independent cloning
LP	lipoprotein
MALDI	Matrix Assisted Laser Desorption/Ionization
MR	molecular replacement
MS	Mass spectrometry
MSA	Multiple sequence alignment
NB-LRR	nucleotide-binding and a leucine-rich repeat domain

OM	outer membrane
PAMP	pathogen-associated molecular pattern
PD	periplasmic domain
PF	Patterson function
PG	peptidoglycan
PGPR	plant growth-promoting rhizobacteria
PHBH	<i>p</i> -hydroxybenzoate hydroxylase
POx	pyranose oxidase
PRR	pattern recognition receptor
PTI	PAMP-triggered immunity
ROS	reactive oxygen species
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	size-exclusion chromatography
TIR	Toll-interleukin 1 receptor
TM	transmembrane
TMD	transmembrane domain
TOF	Time-of-Flight
TLR	Toll-like receptor

Table of contents

1	Introduction	1
1.1	Plant-bacteria interactions	1
1.2	Glycosylated aromatic polyketides in plants	5
1.3	C-glycosyl deglycosylation by symbiotic bacteria	7
1.3.1	C-glycosyl-oxidizing enzymes	8
1.3.2	C-glycosyl-cleaving enzymes	15
1.4	Pathogenic plant-bacteria interactions	18
1.5	Localization of lipoprotein (Lol) pathway	21
2	Methods	26
2.1	Cloning and recombinant protein production	26
2.1.1	Cloning	26
2.1.2	Recombinant gene expression	27
2.1.3	Protein purification	27
2.2	Experimental determination of protein structures	29
2.2.1	Single-crystal protein X-ray crystallography	29
2.2.2	Single-particle cryo-electron microscopy	33
2.3	Other biophysical methods	36
2.3.1	Isothermal titration calorimetry	36
2.3.2	Mass spectrometry	37
2.3.3	High-performance liquid chromatography	38
2.4	Phylogenetic analysis	39
3	Present investigation	40
3.1	Aim of the thesis	40
3.2	Paper I. C-glycosyl deglycosylation in <i>D. aerius</i>	41
3.2.1	<i>Deinococcus aerius</i>	41
3.2.2	Biochemical characterization of the CGD gene cluster	42
3.2.3	Structure determination of DaCGO1 and DaCGD	44
3.2.4	C-glycosyl deglycosylation mechanism	46
3.3	Paper II. C-glycosyl oxidoreductase in <i>S. canus</i>	51
3.3.1	<i>Streptomyces canus</i>	51

3.3.2 Biochemical characterization of ScPOx	52
3.3.3 Catalytic activity of ScPOx	52
3.4 Paper III. C-glycosyl deglycosylation in <i>M. testaceum</i>	55
3.4.1 <i>Microbacterium testaceum</i>	55
3.4.2 Biochemical characterization of <i>MteCGD</i>	55
3.4.3 Structure determination of <i>MteCGD</i>	56
3.4.4 Catalytic activity of <i>MteCGD</i>	59
3.4.5 Phylogenetic analysis of CGDs in Prokaryota	61
3.5 Paper IV-V. Lipoprotein localization in <i>Xcc</i>	63
3.5.1 <i>Xanthomonas campestris</i> pv. <i>campestris</i>	63
3.5.2 LolA-LolB complex formation and interactions	64
3.5.3 LolFD, a novel ABC transporter for lipoprotein extraction	68
4 Concluding remarks and outlook	75
5 Acknowledgement	77
6 References	80

1 Introduction

1.1 Plant-bacteria interactions

Approximately 470 million years ago, the emergence of terrestrial (land-living) plants marked a monumental milestone in the developing narrative of Earth's evolution (Morris et al., 2018). Early terrestrial plants, known as Bryophytes, lacked vascular systems, and included liverworts, hornworts, and mosses.

Around 420 million years ago, vascular plants emerged, initially without seeds (lycophytes and ferns), and then, 360 million years ago, with seeds (gymnosperms and angiosperms) (Kenrick et al., 1997; Ruzsala et al., 2011). Terrestrial plants exhibit remarkable diversity and exhibit structural adaptations influenced by environmental factors (Morris et al., 2018).

Important environmental factors that helped shape the development of terrestrial plants include the interaction of plants with endophytic microorganisms, such as bacteria and fungi (Lodewyckx et al., 2002), and rhizobiome, namely microorganisms living in the region of the soil influenced by root secretions or rhizosphere (Lugtenberg et al., 2009; Hartmann et al., 2008).

Endophytic microorganisms that colonize plant tissues can trigger interactions that are detrimental (parasitic), neutral (symptomless) or beneficial (mutualistic), and the latter may become pathogenic under certain conditions and within different host genotypes. Rhizobacteria can be pathogenic or plant growth-promoting rhizobacteria (PGPR), displaying beneficial effects on the plant directly or indirectly (**Figure 1**).

Mutualistic relationships between organisms throughout the tree of life play a crucial role in shaping the diversity and functioning of ecosystems. The communication between plant, bacterial and fungal kingdoms, and the causes of synergy or antagonism of different microorganisms is a complex, fascinating network to disentangle.

The complex interplay between plants and symbiotic bacteria lies in the necessity of both organisms to survive, which is promoted by establishing to an intricate network of biochemical communication. Bacterial symbionts of plants, endophytic or in the rhizosphere, attempt to avoid triggering the plants' defense systems.

2 INTRODUCTION

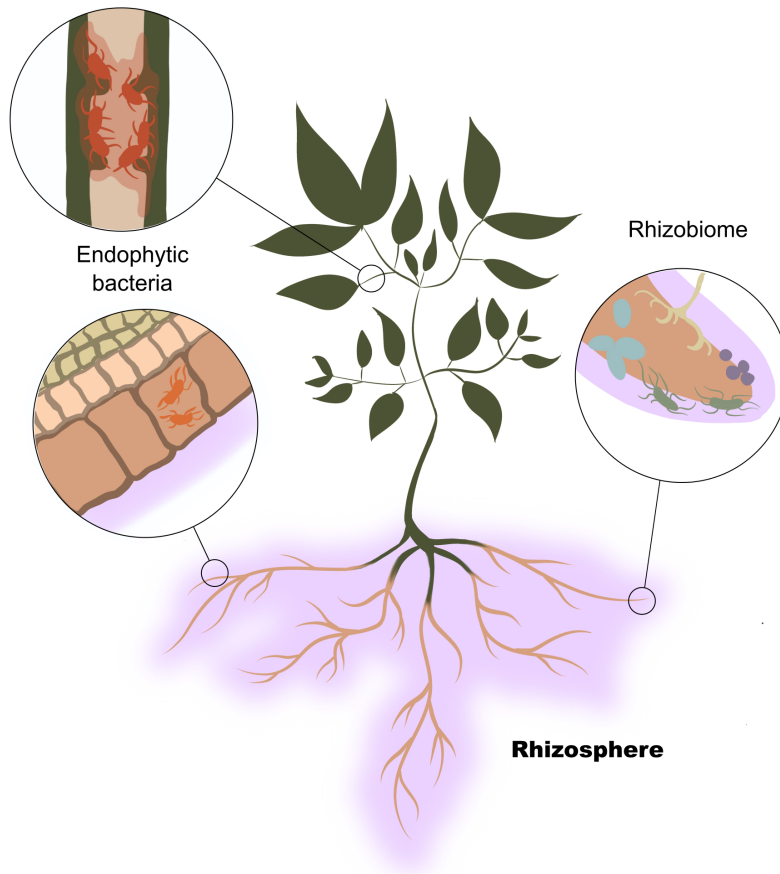


Figure 1: The rhizosphere is the region in close contact with the roots of the plant, colonized by microorganisms influenced by the root secretions. These micro-organisms constitute the rhizobiome. Microorganisms can also colonize plant tissues, for example endophytic bacteria that can spread through the xylem vessels of the plant. Picture copyright © Valentina Furlanetto 2023.

The successful symbiosis between plant and bacteria relies on the ability of certain rhizobacteria and endophilic bacteria not only to survive without inducing harm to the plant, but also to provide useful products to the host (Ma et al., 2022; Lugtenberg et al., 2009). The growth-promoting effect of bacteria on plants can affect the plant through different mechanisms, ranging from stress control to biofertilization, and stimulation of root growth.

Beneficial bacteria can produce phytohormones, enhance plant tolerance to abiotic stress, such as frost or heat stress, drought, heavy metal accumulation (Dai et al., 2021), and increase mineral availability (Lugtenberg et al., 2009). Phosphorous intake and nitrogen fixation can also be mediated by bacteria, resulting in plant growth promotion (Buchanan et al., 2015a; Mus et al., 2016).

Plants and microorganisms can synthesize compounds such as aromatic polyketides (Thomas, 2001; Sharma et al., 2022; Shen, 2003). These compounds are not only important for plant and animal health but are metabolites responsible directly or indirectly for plant defense and plant-host communication. Aromatic polyketides in the rhizosphere play a crucial role in mediating plant-microbe interactions acting as chemical signals, and mediating plant responses to various stresses, including both abiotic and biotic stressors. Biotic stress, for instance, can arise from the attack of plant tissues by pathogenic microorganisms (Dixon et al., 1995), and among the many biological functions of these active compounds are protection against plant diseases and communication with beneficial soil microbes (Yu et al., 2008).

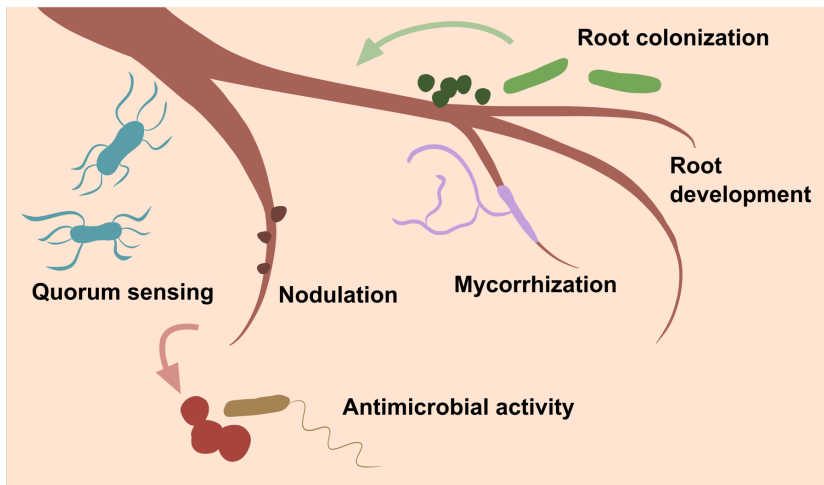
For example, xanthenes can have antimicrobial effects that benefit the plant (Liu et al., 2022; Han et al., 2018). Flavonoids have, in addition to their key role in nitrogen fixation processes (Yu et al., 2021), an important role in quorum sensing in the rhizosphere, inducing symbiotic microbial interactions, for example initiating mycorrhizal spore germination and hyphal branching and colonization of the roots (Siqueira et al., 1991; Scervino et al., 2005; Steinkellner et al., 2007) (**Figure 2**).

Aromatic polyketides such as flavonoids are involved in the beneficial microbial enrichment of the rhizosphere with inducing soil suppressive effects (Yu et al., 2021). The soil-suppressive effects are linked to the ability of certain beneficial bacteria to perform nitrogen fixation for the plant. The main input to the nitrogen pool in natural ecosystems is from reduction of atmospheric and dissolved molecular nitrogen (N_2) to ammonia (NH_3).

Nitrogen fixation is carried out only by prokaryotes and archaea, and in natural terrestrial ecosystems about 80-90% of the nitrogen available to plants originate from nitrogen fixation (Buchanan et al., 2015a; Cocking, 2003). The nitrogen fixated by symbiotic microorganisms is assimilated *in planta* into amino acids and other nitrogen-containing compounds, like nucleic acids, cofactors, and other common metabolites like chlorophyll.

4 INTRODUCTION

A



B

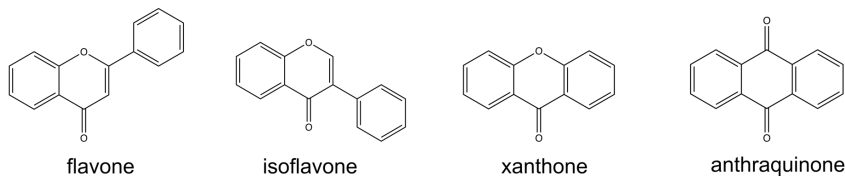


Figure 2: The roots and surrounding rhizosphere are rich in various compounds, including diverse aromatic polyketides in both glycosylated and deglycosylated forms. (A) Illustration of various roles of active compounds in the rhizosphere. (B) Some examples of active aromatic polyketides of plant origin that can serve as secondary metabolites. Flavones and isoflavones, belonging to the flavonoid class, participate in plant-bacteria quorum sensing. Xanthones exhibit antimicrobial properties, while anthraquinones influence both biotic and abiotic factors in plants. Picture copyright © Valentina Furlanetto 2023.

Several plant hormones and secondary metabolites contain nitrogen, but even the ones that do not, like flavonoids and other plant phenolic compounds, are derived from phenylalanine, and thereby linked to amino-acid metabolism (Remali et al., 2022; Hassan et al., 2012; Yu et al., 2008; Yonekura-Sakakibara et al., 2019).

Nitrogen fixation is carried out in the plant nodules, a process that is induced by the production of flavonoid compounds (**Figure 3**). These compounds induce the transcription of *nod* genes and production of nodulation (Nod) factors in symbiotic bacteria, typically *N*-acetylated chito-oligosaccharides (Long et al., 1996), that signal to the plant to

establish nodules for nitrogen fixation and transport of the resulting ammonia throughout the plant (Hassan et al., 2012; Long et al., 1996; Mus et al., 2016).

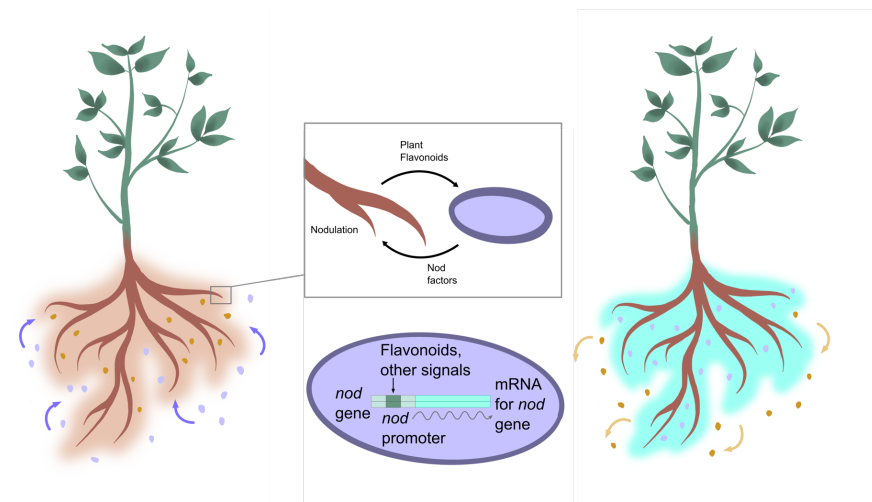


Figure 3: One of the many symbiotic interactions between plants and soil bacteria involves flavonoid-mediated nodulation for nitrogen fixation. Certain beneficial bacteria carry *nod* genes, which are activated in response to plant flavonoids. The resulting Nod factors play a crucial role in initiating nodule formation in the roots, where nitrogen fixation occurs. This process, coupled with other mechanisms, enriches the soil with beneficial bacteria, leading to soil suppressiveness. Picture copyright © Valentina Furlanetto 2023.

Nitrogen fixation offers benefits to the plant by considerably increasing growth and yield, which becomes evident when studying plants grown in nitrogen-poor soil (Bowsher et al., 2018; Pankiewicz et al., 2022). The nitrogen fixed by symbiotic bacteria is exchanged for carbon fixed by the plant, and the optimized physiological conditions provided by the plant environment increase the efficiency of nitrogen fixation (Hassan et al., 2012; Mus et al., 2016; Cocking, 2003). This interplay can overcome constraints that often limit nitrogen fixation by nonsymbiotic bacteria.

1.2 Glycosylated aromatic polyketides in plants

Flavonoids are common aromatic polyketide compounds produced by plants. These aromatic polyketides can undergo various modifications, including glycosylation where the aromatic polyphenolic aglycone group is

6 INTRODUCTION

modified by a sugar moiety attached *via* an *O*- or *C*-bond (**Figure 4**). Glycosylation of aromatic polyketides enhances their solubility in aqueous solutions and influences their mobility in the soil (Hassan et al., 2012).

Examination of root tissues and root exudates has shown the presence of both deglycosylated and glycosylated aromatic polyketides (Weisskopf et al., 2006). The glycosylated compounds are rapidly deglycosylated by endophytic and rhizosphere microorganisms (Hartwig et al., 1991). Glycosylated plant aromatic polyketides serve as secondary metabolites that are important for plant adaptation to changing environmental conditions and for effectively managing biotic and abiotic stresses. They are also playing a key role in the selective crosstalk between plants and the beneficial soil microbiome (Bag et al., 2022).

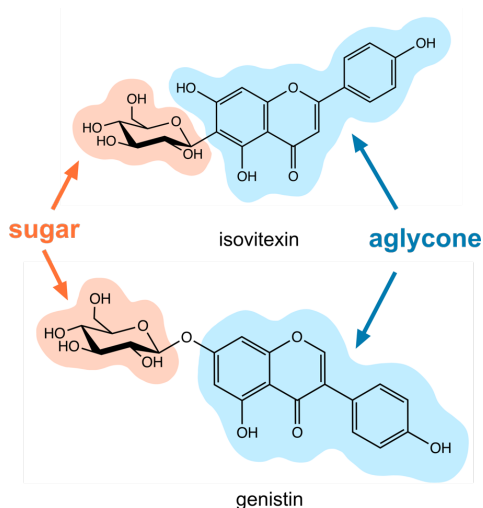


Figure 4: Illustration of glycosylated aromatic polyketides, including the C-glycosylated flavone isovitexin and the O-glycosylated isoflavone genistin. Picture copyright © Valentina Furlanetto 2023.

As will be discussed in the next chapter, bacterial plant symbionts have evolved efficient enzymatic systems for deglycosylation of a wide range of naturally occurring glycosylated aromatic polyketides, as quorum sensing mechanism between the beneficial microorganisms and plants. The model substrates for the studied enzymatic systems discussed in this thesis present glucose linked to the aglycone moiety, hence can be referred to as glucosylated aromatic polyketides.

1.3 C-glycosyl deglycosylation by symbiotic bacteria

In the case of glycosylated aromatic polyketides produced by plants, O-glycosylflavonoids are the most common, however recently, there has been an increased interest in the less common C-glycosylflavonoids (CGFs), and other C-glycosylated aromatic polyketides raised interest in the scientific community (Casas et al., 2016; Bucar et al., 2021)

The carbon-carbon (C-C) bond that attaches the sugar and the aromatic aglycone in C-glycosylated aromatic polyketides is highly stable and recalcitrant to chemical and enzymatic hydrolysis compared with other glycosidic linkages (Ji et al., 2020). Bacteria with the ability to deglycosylate C-glycosyl compounds was first identified in the human intestinal flora in late 80s and included enzymatic activity on barbaloin, isoorientin and mangiferin (Hattori et al., 1988a; 1988b; 1989).

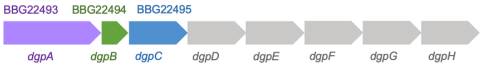
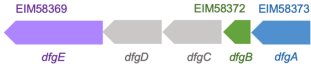


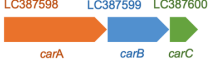
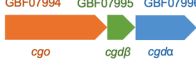
However, the bacterial enzymes responsible for C-deglycosylation have only recently been described: gene clusters coding for C-deglycosylating enzymes were identified and characterized for the intestinal bacteria *Eubacterium cellulosolvens* (Braune et al., 2016; Mori et al., 2021) and PUE (Nakamura et al., 2019; 2020; Mori et al., 2021; He et al., 2023); and for the soil bacteria *Microbacterium trichothecenolyticum* (Mori et al., 2021), *Microbacterium* sp. 5-2b (Kumano et al., 2021; Mori et al., 2021), and *Arthrobacter globiformis* (Mori et al., 2021).

The C-deglycosylation gene clusters typically include genes coding for two types of enzymatic activities: i) one or several C-glycosyl oxidoreductases (CGOs) responsible for oxidation of the sugar moiety to generate an oxo substrate for subsequent enzymatic cleavage of the C-C bond; and ii) one or several C-glycoside/glycosyl deglycosidases (CGDs) that cleave the C-C cleavage of the bond between the oxidized sugar and the aglycone (**Table 1**).

The *cgd* gene cluster in human intestinal bacterium PUE (Nakamura et al., 2019) includes the genes *dgpA-dgpH* where *dgpA* codes for a GFO/IDH/MocA-type oxidoreductase and *dgpB* and *dgpC* code for the β - and α -subunits of a CGD complex, respectively. In *E. cellulosolvens* 6 (Mori et al., 2021), a gene cluster composed of the genes *dfgA-dfgE* where *dfgE* corresponds to a GFO/IDH/MocA-type enzyme (*dfgE*), and *dfgA* and *dfgB* to the CGD α - and β -subunits.

8 INTRODUCTION

Table 1. CGD gene clusters.

Organism	Genome ID	Gene cluster
Human intestinal bacterium PUE	LC422372	
<i>Eubacterium cellulosolvens</i> 6	CM001487	
<i>Arthrobacter globiformis</i> NBRC 12137	BAEG01000075	
<i>Microbacterium trichothecenolyticum</i> DSM 8608	JYJA01000033	
<i>Microbacterium</i> sp. 5-2b		
<i>Deinococcus aerius</i> TR0125 (this work)	BFAG01000020 (contig_20)	

GMC-type oxidoreductase (orange); GIM-type oxidoreductase (purple); CGD α -subunit (blue); CGD β -subunit (green). The GenBank accession is given for the GIM-type oxidoreductase from *A. globiformis* NBRC12137. Table adapted from Paper I.

The corresponding gene cluster in the soil bacterium *A. globiformis* NBRC12137 (Mori et al., 2021) includes one gene coding for a Glucose/Methanol/Choline (GMC)-type oxidoreductase (*carA*), and four genes coding for two CGD complexes (*carB1-carC1* and *carB2-carC2*) where the *carB* genes correspond to the α -subunit and *carC* genes for the β -subunits. Additionally, the cluster includes a gene coding for a GFO/IDH/MocA-type oxidoreductase. The genomes of the soil bacteria *M. trichithecenolyticum* DSM 86o8 and *Microbacterium* sp. 5-2b both have gene clusters coding for one GMC oxidoreductase and the α - and β -subunits of a CGD complex.

1.3.1 C-glycosyl-oxidizing enzymes (CGOs)

The GMC superfamily of oxidoreductases was first defined in 1992 by Cavener (Cavener, 1992) through sequence comparison which identified

close relationships between *Drosophila melanogaster* glucose dehydrogenase, *Escherichia coli* choline dehydrogenase, *Aspergillus niger* glucose oxidase and *Hansenula polymorpha* methanol oxidase. Since then, many enzymes have been added to this family, including representatives from bacteria, yeast, filamentous fungi, and insects. The GMC superfamily now appear as part of the Auxiliary Activity Family 3 (AA3) in the Carbohydrate-Active enZYmes (CAZy) database (Levasseur et al., 2013; www.cazy.org).

GMC oxidoreductases feature a common highly conserved flavin adenine dinucleotide (FAD)-binding domain, where the FAD cofactor can be covalently or non-covalently bound, and a more variable substrate-binding domain. The electron-donor substrates can be sugars, alcohols, cholesterol, or choline, and the sequence variation of the substrate-binding domain reflects the different substrate preferences of the functionally diverse members of the GMC superfamily. Various electron acceptors can be used, from oxygen to quinones, phenol radicals or metal ions. The preference for the electron acceptor determines whether the GMC enzyme is primarily an oxidase using O₂ as electron acceptor, or a dehydrogenase.

GMC oxidoreductases fold into a *p*-hydroxybenzoate hydroxylase (PHBH)-like fold (Wierenga et al, 1979; Mattevi, 1998) where the FAD binding domain has a Rossmann-type fold with a $\beta\alpha\beta$ mononucleotide-binding motif (Rossmann et al., 1975; Wierenga et al., 1983; 1986); and a substrate binding domain composed of a six-stranded β -sheet and three α -helices. A typical feature of the GMC members is the presence of either a loop-and-arm region or of a variable loop-and-lid region (typically involved in the eventual oligomerization of the GMC). For instance, the structure of the fungal GMC oxidoreductase pyranose 2-oxidase from *Trametes ochracea* (ToP2O; former *T. multicolor*) includes a loop-and-arm region (**Figure 5**) that stabilizes the homotetrameric assembly (Hallberg et al, 2004). Other GMC oxidoreductases present instead loop-and-lid regions of varying length (Hallberg et al., 2002). An active-site loop gates the entrance of the active site positioned at the interface between the two subdomains.

10 INTRODUCTION

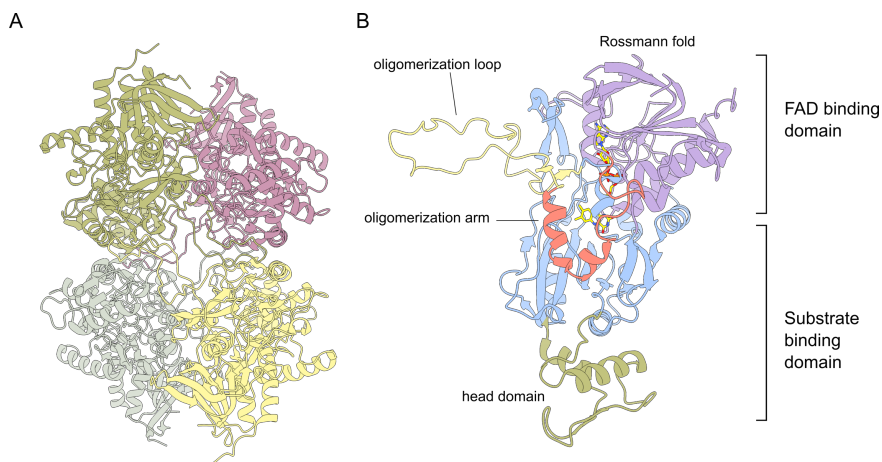


Figure 5: The structure of fungal *ToP2O*. (A) *ToP2O* assembles as a homotetramer that is stabilized by the oligomerization arm in each subunit (PDB 1TT0; Hallberg et al., 2004). (B) The structure of the *ToP2O* monomer with the FAD coactor shown in yellow. Picture copyright © Valentina Furlanetto 2023.

As mentioned above, some GMC oxidoreductases are CGOs that perform the necessary oxidation of the sugar in *C*-glycosylated aromatic polyketides prior to C-C-bond cleavage. Kumano and co-workers identified and characterized the first GMC-type CGO, *CarA*, produced by the soil bacterium *Microbacterium* sp. 5-2b (Kumano et al., 2021). We will hereafter refer to this enzyme as *MiCarA*.

They also identified two homologs of *MiCarA* from the soil bacteria *A. globiformis* and *M. trichothecenolyticum*, *AgCarA* and *MtCarA*, respectively. Activity studies showed that these enzymes could oxidize the glucose moiety in aromatic polyketide compounds: *MiCarA* oxidized isovitexin, isoorientin, mangiferin, and to some extent carminic acid; *AgCarA* oxidized isovitexin, isoorientin, and mangiferin; and *MtCarA* oxidized isoorientin and mangiferin (**Figure 6A**; **Table 2**). None of the GMC-type oxidoreductases could oxidize C-8 glycosylated flavonoids. The enzymes showed no activity on glucose and broad, but low, activity on *O*-glycosylated compounds.

They found that the reaction products from carminic acid using cell-free extracts from *Microbacterium* sp. 5-2b were C3'-keto carminic acid, C2'-keto carminic acid and kermesic acid. It could also be determined that C2'-

keto carminic acid is the isomerization product of C3'-keto carminic acid in aqueous solution.

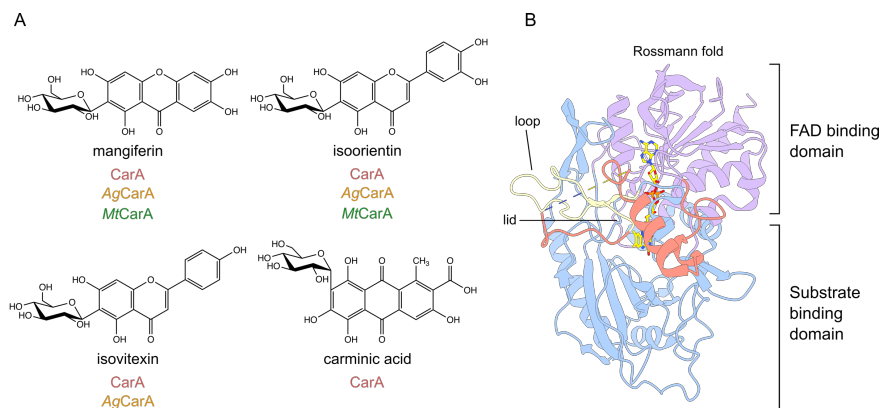


Figure 6: (A) Glycosylated aromatic polyketides as substrates of the GMC enzymes discussed. (B) Overall structure of the monomeric *MtCarA* (PDB 7DVE). *MtCarA* has the same fold as *ToP2O* but is lacking the head domain and the long oligomerization arm. Picture copyright © Valentina Furlanetto 2023.

MtCarA is currently the only GMC oxidoreductase known to oxidize C6-glycosylated aromatic polyketides that has been structurally characterized (PDB 7DVE; **Table 2**) (Kumano et al., 2021). *MtCarA* presents the typical GMC-oxidoreductase fold with an FAD-binding domain with non-covalently bound FAD and a substrate binding domain (**Figure 6B**).

Structural comparison of *MtCarA* with the experimental structure of the fungal pyranose 2-oxidase *PeP2O* from the fungal *Peniophora* in complex with the reaction product (PDB 1YZL), was used to dock a carminate in the *MtCarA* active site. The superimposition of the pyranose ring of carminate on the pyranose ring of the reaction product of *PeP2O* revealed that, while the active-site residues are conserved in *MtCarA* and *PeP2O*, the docking of the carminate aglycone moiety requires a conformational change of the loop gating the substrate entrance.

The gene *carA* is part of a C-glycoside deglycosylation gene cluster in *Microbacterium* sp. 5-2b, where the adjacent genes *carB* and *carC* code the proteins CarB and CarC, displaying sequences similar to the subunits of the CGD formed by DgpC and DgpB, respectively (Nakamura et al., 2020). The genes *dgpC* and *dgpB* belong to the C-deglycosylation cluster in the human intestinal bacteria PUE.

Table 2. Overview of previously characterized CGO and CGD enzymes.

Organism	CGO type	CGO activity (reference)	CGO PDB accession (state, resolution, method, reference)	CGD	Oligomer	CGD activity (reference)	CGD PDB accession (state, resolution, method, reference)
<i>Human intestinal bacterium PUE</i>	DgpA GIM	Puerarin Vilexin (Nakamura et al., 2019; He et al., 2023)	7XRE (X-ray, 2.70 Å, He et al., 2023) 7XR9 (glucose, X-ray, 2.42 Å, He et al., 2023)	PuCGD (DgpB-DgpC)	αβ ₄	3"-oxo-puerarin 3"-oxo-orientin	7BYR (Mn ²⁺ , X-ray, 2.60 Å, Mori et al. 2021) 7DRD (apo, cryo-EM, 2.85 Å, Mori et al. 2021) 7EXZ (Mn ²⁺ , X-ray, 2.50 Å, Mori et al. 2021) 7XRF (Mn ²⁺ , X-ray, 2.14 Å, He et al. 2023)
<i>Eubacterium cellulosolvens</i> 6	DfgE GIM	Isoorientin Isovitexin (Braune et al., 2016; Nakamura et al., 2019)	Not determined	EuCGD (DfgA-DfgB)	αβ ₄	3"-oxo-isovitexin 3"-oxo-isoorientin	7BVS (Mn ²⁺ , X-ray, 2.85 Å, Mori et al., 2021) 7DRE (apo, cryo-EM, 2.54 Å, Mori et al., 2021) 7EXB (Mn ²⁺ , X-ray, 2.40 Å, Mori et al., 2021)
<i>Arthrobacter globiformis</i> NBRC <u>12137</u>	AgCarA GMC	Isovitexin Isoorientin Mangiferin	Not determined	AgCGD1 (AgCarB1-C1)	αβ	3"-oxo-isovitexin 3"-oxo-isoorientin 3"-oxo-mangiferin	Not determined
<i>Arthrobacter globiformis</i> NBRC <u>12137</u>	AgCarA GMC	Isovitexin Isoorientin Mangiferin	Not determined	AgCGD2 (AgCarB2-C2)	αβ	3"-oxo-orientin	7DNM (Mn ²⁺ , X-ray, 2.30 Å, Mori et al., 2021) 7DNN (Mn ²⁺ , isoorientin, X-ray, 2.25 Å, Mori et al., 2021)
<i>Microbacterium trichothecenolyticum</i>	MtCarA GMC	Isoorientin Mangiferin	7DVE (X-ray, 2.40 Å, Kumano et al., 2021)	MtCGD (CarB-CarC)	αβ	3"-oxo-isovitexin 3"-oxo-isoorientin 3"-oxo-mangiferin	Not determined
<i>Microbacterium</i> sp. 5-2b	MtCarA (CarA) GMC	Isovitexin Isoorientin Mangiferin Carminic acid	Not determined	MtCGD	αβ	3"-oxo-puerarin 3"-oxo-isoorientin 3"-oxo-mangiferin 3"-oxo-carminate	Not determined

The *cgo* gene that codes for DgpA is not homologous to *MiCarA* and is instead classified as GFO/IDH/MocA. The GFO/IDH/MocA superfamily includes oxidoreductases that depend on NADP or NAD (Rowland et al., 1994; Kingston et al., 1996). The GFO/IDH/MocA abbreviations refer to glucose-fructose oxidoreductase (GFO), inositol 2-dehydrogenase (IDH), and the rhizopine catabolism protein MocA (MocA). In analogy with the GMC acronym for the Glucose/Methanol/Choline oxidoreductase family, we propose in Paper I to use the acronym GIM for the GFO/IDH/MocA enzymes. The GIM oxidoreductases consist of two domains, an N-terminal NAD(P)-binding Rossmann-like fold (Rossmann et al., 1975; Taberman et al., 2016) and a C-terminal substrate binding α/β domain (Taberman et al., 2016). The C-terminal domain consists of a two-layered α/β -sandwich that also participates in oligomerization.

The proposed oxidation reaction mechanism of GIM oxidoreductases includes hydride abstraction by $\text{NAD}^+/\text{NADP}^+$ from the sugar carbon defined as the site of oxidation to generate a protonated ketone intermediate. An amino acid acting as a base is responsible for abstracting the proton from the charged glycosyl intermediate to form the final oxidized product (Taberman et al., 2016).

As mentioned above, the enzyme DgpA from the human intestinal bacterium PUE catalyzes the oxidation of C-glycoside aromatic polyketides. Specifically, this CGO was originally identified for its ability to oxidize of the C-6 glucosylated substrate puerarin (Nakamura et al., 2020). Nakamura and co-workers found unexpectedly that the NAD(H)-dependent enzyme was able to oxidize puerarin to 3''-oxo puerarin, the latter undergoing non enzymatic isomerization to 2''-oxo puerarin in aqueous conditions. They suggested a “ping-pong” mechanism where the associated CGD was responsible for catalyzing the C-C bond cleavage of oxo-puerarin to yield 1,5-anhydro-D-erythro-hex-1-en-3-ulose and the aglycone daidzein. The 1,5-anhydro-D-erythro-hex-1-en-3-ulose would then be converted to 3-oxo glucose in aqueous solution and act as a hydride acceptor for puerarin oxidation from DfgA. Experimental crystal structures of DgpA were only recently determined (**Figure 7; Table 2**), specifically the unliganded state (PDB 7XRE; He et al., 2023) and the complex with β -D-glucose (PDB 7XR9; He et al., 2023).

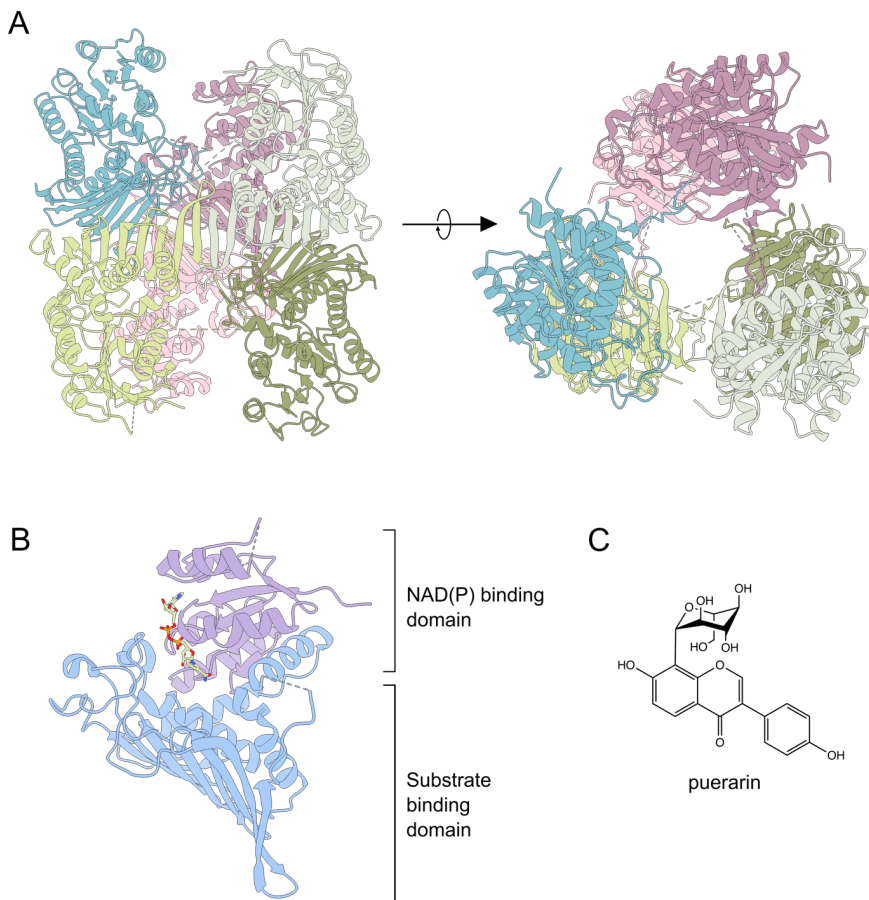


Figure 7: Crystal structures of the GIM-type oxidoreductase DgpA. (A) DgpA forms a hexamer where the interactions between the monomers have auto-inhibitory effects due to the blockage of the substrate binding pocket from a loop of the neighboring DgpA. Deletion of the loop causes a 3-fold increase of the DgpA activity towards glucose, compared to the wild type (PDB 7XRE; He et al., 2023). (B) The monomers display the fold of GIM-type oxidoreductases. (C) The C-8 glucosylated puerarin substrate for PUE DgpA (Nakamura et al., 2020). Picture copyright © Valentina Furlanetto 2023.

The complex of DgpA with β -D-glucose showed that the glucosyl moiety was suitably oriented for oxidation at the glucosyl C3 position (PDB 7XR9). Based on the experimental glucose complex, the authors docked puerarin to the active site of DgpA and proposed that recognition and binding of the glucose moiety is the key specificity determinant.

Activity studies showed that glucose and puerarin bind competitively to DgpA. Analysis of the reaction products from the combined DgpA/DgpB/DgpC reaction showed that the CGD was responsible for determining the specificity of the overall reaction towards C- and O-glycosidic (iso)flavonoids. Interestingly, they also observed an intriguing isomerase activity for O- and C-glycosidic compounds (He et al., 2023).

1.3.2 C-glycosyl-cleaving enzymes (CGDs)

The C-C bond cleavage reaction to release the bioactive aglycone can take place only after the C-glycosyl compound has been oxidized, and as discussed above the enzymes that are currently known to perform the oxidation reactions are either a GMC- or GIM-type oxidoreductases. Thus, the identified substrates for the CGDs can be either a C3"-oxidized sugar attached to the C6 or C8 position of a flavonoid aglycone.

All CGDs identified to date are metalloenzymes. Prior to the work presented in this thesis, all structurally characterized CGDs also contained separate α - and β -subunits that assemble to form a competent enzyme complex (**Table 2**). While CGDs produced by the intestinal bacteria PUE (*PuCGD*) and *E. cellulosolvens* 6 (*EuCGD*) were found to be $\alpha_4\beta_4$ heterooctamers, the CGDs from the soil bacteria *A. globiformis* NBRC 12137 (*AgCGD2*) and *M. trichothecenolyticum* (*MtCGD*) assembled as $\alpha\beta$ heterodimers (Mori et al., 2021).

The α -subunit includes the PFAM motif "AP_endonuc_2" that folds as a xylose isomerase-like TIM-barrel. This subunit also contains a lid-domain inserted in one β/α connection of the TIM-barrel that folds into a 4-helix bundle. The β -subunit presents an immunoglobulin-like β -sandwich fold with the PFAM motif "DUF6379". Existing experimental CGD structures indicate that both the helix bundle and the β -subunit can move in response to substrate binding and departure.

The active site of known CGDs show a conserved "E,D,HxK,E" sequence motif that includes the ligands for the catalytic metal ion. The catalytic residues have been proposed to be a histidine and a glutamic acid that are located close to the metal-binding site.

The CGD from the intestinal bacteria PUE (*PuCGD*; DgpB-DgpC) was identified as the enzyme responsible for C-C bond cleavage of chemically synthesized 3"-oxo puerarin (Nakamura et al., 2019), and enzyme system was later characterized in more detail to also identify the C-C bond-

cleavage product 1,5-anhydro-D-erythro-hex-1-en-3-ulose mentioned above (Nakamura et al., 2020)

Mori and coworkers characterized *PuCGD* biochemically and structurally (**Table 2**) using X-ray crystallography (PDB 7EXZ) and cryo-EM (PDB 7DRD). *PuCGD* could cleave the C-C bond in 3''-oxo forms of the C-8 glucosyl flavonoids puerarin and orientin, but not in the C-6 glycosylated oxo compounds, regardless of backbone structure (i.e., flavonoid, hydroxyanthrapurin or xanthonoid).

The experimental structure of *PuCGD* was found to be an $\alpha_4\beta_4$ heterooctamer composed by four DgpC-DgpB heterodimers (**Figure 8A; Table 2**). DgpC and DgpB were also renamed *PuCGD* $^\alpha$ and *PuCGD* $^\beta$, respectively (Mori et al., 2021). The *PuCGD* $^\alpha$ subunits displayed the predicted TIM-barrel fold and the *PuCGD* $^\beta$ subunit folded a 7-stranded β -sandwich fold.

The active sites are present in the *PuCGD* $^\alpha$ subunits and located at the interface between *PuCGD* $^\alpha$ and *PuCGD* $^\beta$. The active sites are exposed to the surrounding solvent (**Figure 8A**) and their entrances gated by a flexible 4-helix bundle in *PuCGD* $^\alpha$ (lid domain) and a long loop in the *PuCGD* $^\beta$ subunit. The lid domain is absent in the electron potential map of the cryo-EM structure (PDB 7DRD), and deletion of the lid abolished *PuCGD* activity but did not disrupt the oligomeric assembly (Mori et al., 2021). These observations emphasize the dynamic character of the lid domain, and its importance for catalysis.

EuCGD and its associated GIM-type CGO (DfgE) produced by the intestinal bacterium *E. cellulosolvens* 6 were produced and characterized by Mori and coworkers. The structure of *EuCGD* (**Figure 8B; Table 2**) was determined by x-ray crystallography (PDB 7EXB) and cryo-EM (PDB 7DRE).

Like *PuCGD*, *EuCGD* is an $\alpha_4\beta_4$ heterooctamer formed by four heterodimers composed of an α - (DfgA) and a β -subunit (DfgB) that are homologous to *PuCGD* $^\alpha$ (DgpC) and *PuCGD* $^\beta$ (DgpB). An important difference is that the 4-helix bundle that forms a flexible lid in *PuCGD* is replaced by an extended flexible loop in *EuCGD*. *EuCGD* could use the C-6 flavonoids 3''-oxo-isorientin and 3''-oxo-isovitexin as substrates (Mori et al., 2021).

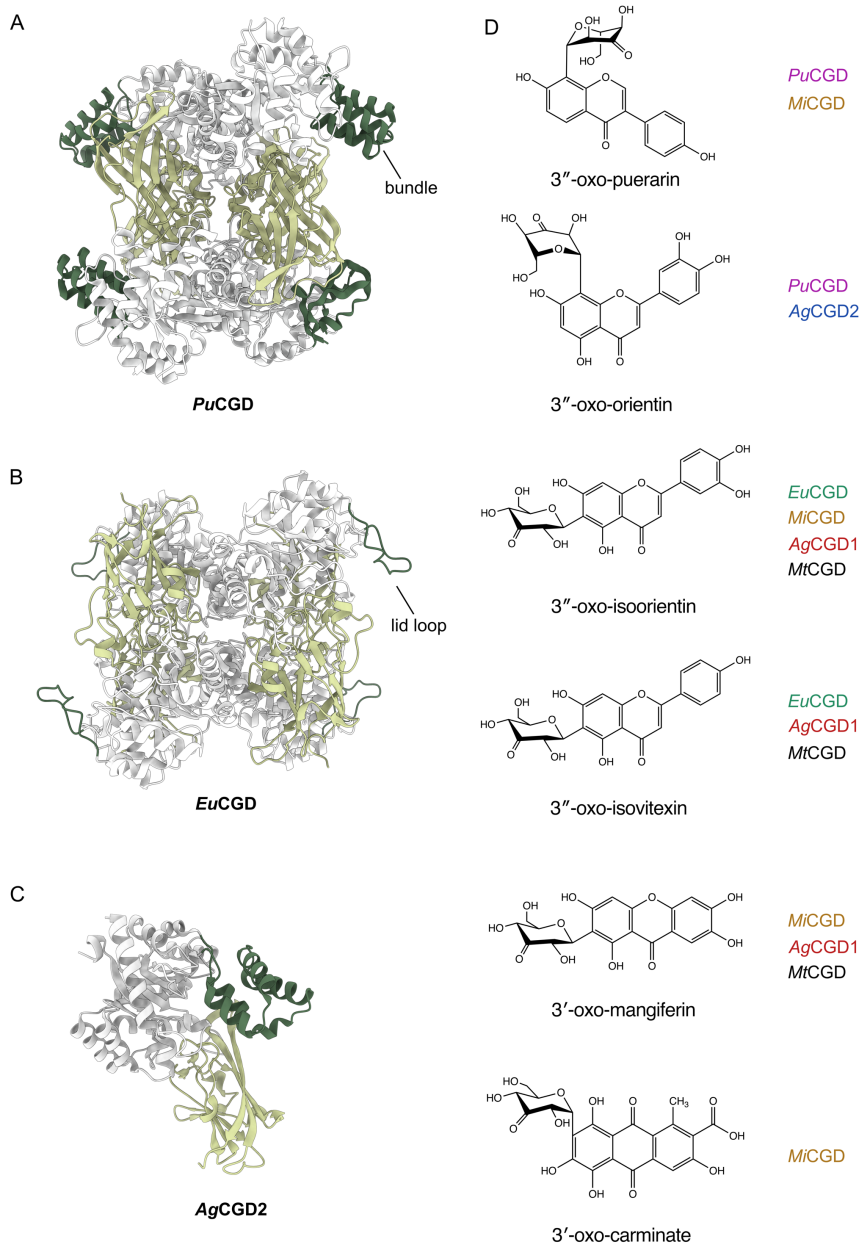


Figure 8: Experimental structures of (A) *PuCGD* (PDB 7EXZ), (B) *EuCGD* (PDB 7EXB) and (C) *AgCGD2* (PDB 7DNM). (D) oxo substrates for the CGDs characterized by Mori et al., 2021. Picture copyright © Valentina Furlanetto 2023.

In contrast to *PuCGD* and *EuCGD*, *AgCGD2* from the soil bacterium *A. globiformis* (PDB 7DNM) is an $\alpha\beta$ heterodimer with the α - and β -subunits with displaying the same overall folds as other CGDs (**Figures 8C; Table 2**) with activity towards 3'-oxo C-8 glycosylated flavonoid 3"-oxo-orientin (**Figure 8D**) (Mori et al., 2021). As for *PuCGD*, *AgCGD2* has a lid structured as a 4-helix bundle.

In the *AgCGD2* complex with isoorientin (PDB 7DNN; Mori et al., 2021), the relative angle between the α - and β -subunits appears different compared to that in *PuCGD*, with only one of the $\alpha\beta$ heterodimers in the asymmetric unit bound to isoorientin. The N-terminal loop of *AgCGD2* resembles that in *EuCGD* but is shorter than the corresponding loop in *PuCGD* that gates the substrate entry. The *AgCGD2*/isoorientin complex confirmed the position of the active site at the interface of the α and β subunits.

The catalytic Mn^{2+} center is conserved in all structures determined by Mori and coworkers, with the only difference being an additional water molecule acting as Mn^{2+} ligand in *PuCGD*, rather than the conserved lysine. While the metal-coordinating ligands are located only in the α -subunit, the *AgCGD2*/isoorientin complex emphasized the importance of the β -subunit for aglycone binding, even though isoorientin did not prove to be a substrate for *AgCGD2*.

Mori and coworkers also characterized biochemically *MtCGD* from *M. trichothecenolyticum* and *AgCGD1*, a homologue of *AgCGD2* produced by the *A. globiformis* (Mori et al., 2021). *MtCGD* and *AgCGD1* are both $\alpha\beta$ heterodimers and active towards 3"-oxo isoorientin, 3"-oxo isovitexin, and 3'-oxo mangiferin (**Figure 8D**). The CGD from *Microbacterium* sp. 5-2b (*MiCGD*) also formed an $\alpha\beta$ heterodimer (CarB, α -subunit; CarC, β -subunit) and displayed activity towards 3"-oxo puerarin, 3"-oxo isoorientin, 3'-oxo mangiferin and weak activity for 3'-oxo carminic acid (**Figure 8D**).

1.4 Pathogenic plant-bacteria interactions

Plants lack a circulating somatically adaptive immune system for self-protection, relying instead on a multifaceted stress response system. When faced with various types of stress, plants trigger alterations in cellular metabolism and gene expression to adapt and survive (Huang et al., 2022).

The specific response depends on factors such as stress intensity, duration, and the combination of different stressors. Plants have evolved diverse survival strategies to navigate environmental challenges, with multiple pathways interconnecting in a complex mechanism involving various factors.

Hormone signaling and the generation of compounds like flavonoids and polyphenols inhibiting reactive oxygen species (ROS) play pivotal roles in initiating gene expression, often involving transcription factors and protein kinases, not only *in planta*, but also in relationship with symbiotic microorganisms, as discussed in Chapter 1.1. In some cases, antagonistic interactions among various plant hormones can occur in response to both biotic and abiotic stressors, adding an extra layer of complexity to the intricate regulatory networks governing plant stress responses (Wang et al., 2011).

In the context of biotic stress in plants, a clear distinction can be made between plant pathogens and plant pests. Plant pests are organisms that feed on vegetative tissue, fruits, and seeds, consuming plant parts directly. In contrast, plant pathogens complete some, or all, of their life cycle inside the plant, frequently resulting in damage or disease. The key factor distinguishing a resistant plant from a susceptible one is its ability to promptly recognize stress or the presence of a pathogen and initiate defense mechanisms with sufficient speed (Buchanan et al., 2015b).

The origins of plant pathogens are still not entirely clear. It is possible that plant ancestors carried parasites with them as they transitioned to land, or ocean-dwelling species may have evolved to exploit land plants. Alternatively, existing pathogenic species might have switched hosts or expanded their host range. Plant pathogens can exhibit different modes of action, categorized as biotrophic, hemibiotrophic, or necrotrophic (Vleeshouwers et al., 2014). Biotrophic species evolve from commensal microbial species and live in symbiosis with the plant. Necrotrophic pathogens cause plant or cellular death and originate from saprophytic microbes that feed off plant tissue in the environment. Hemibiotrophic species initially display biotrophic behavior and later switch to necrotrophic behavior.

Plant pathogens can also be classified based on their host range, with specialists having a narrow range and robust mechanisms to overcome specific host defenses, while generalists have a broader host range, and less is known about their mechanisms for overcoming plant defenses.

Pathogens that attack plants typically generate pathogen-associated molecular patterns (PAMPs), for instance chitin originating from the pathogen's cell wall or components of the peptidoglycan, or lipid anchors of bacterial lipoproteins (LPs) (Boller et al., 2009; Gay et al., 2007). These PAMPs are recognized by the plant through membrane associated pattern recognition receptors (PRRs) presenting Toll-like receptors (TLRs) (Botos et al., 2011; Xia et al., 2021), initiating the PAMP-triggered immunity (PTI) (Janeway, 1989; Zhang et al., 2010) (**Figure 9**).

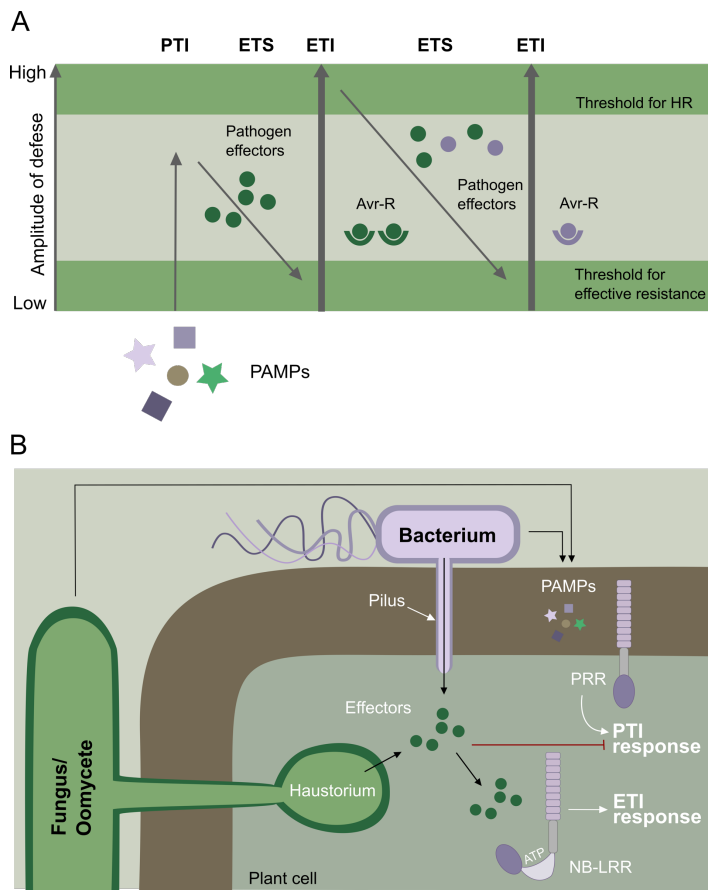


Figure 9: The principles of plant immunity and interaction with plant pathogens. (A) The “zigzag” defense response model illustrates how PAMPs produced by pathogens induce PTI in the plant, which can trigger the pathogen to produce effector molecules. The ETI is initiated by the recognition of effectors (Avr) by plant receptors (R). The amplitude of defense can culminate in the hypersensitive response (HR) leading to programmed cell death. (B) PAMPs generated by

microbial pathogens can be recognized by PRRs located on the cell surface. PRRs have an extracellular leucine-rich repeat (LRR) and an intracellular protein kinase domain and initiate the cascade leading to the PTI. PTI-suppressing bacterial effectors are delivered to the intracellular space for example through a type-III secretion pilus, while fungi and oomycetes effectors can rely on haustoria. Effectors can be recognized by NB-LRR, which have a carboxyl-terminal LRR domain, an ATP- or ADP-binding NB domain and typically an N-terminal Toll-interleukin 1 receptor (TIR) domain. Picture adapted from Buchanan et al., 2015b.

This first step of plant defense can pressure the pathogens to produce effectors to overcome the defense mechanism initiating effector-triggered susceptibility (ETS). Effector molecules help the pathogen to colonize the host and can be cell wall-degrading enzymes, toxins, and transcription factors (Caplan et al., 2008; Wang et al., 2015). These effector molecules can be recognized by immune receptors containing a nucleotide-binding and a leucine-rich repeat domain (NB-LRR), inducing the plant effector triggered immunity (ETI) (**Figure 9**).

From the perspective of bacterial pathogens, outer membrane lipoproteins of Gram-negative bacteria play a key role in pathogenicity and to enhance virulence mechanisms such as host invasion and colonization, immunomodulation, and evasion of the host defenses (Kovacs-Simon et al., 2011; El Rayes et al., 2021a).

1.5 Localization of lipoprotein (Lol) pathway

The cells of Gram-negative bacteria have an inner membrane (IM) and an outer membrane (OM) separated by a periplasmic space that is crowded with different proteins and contains the peptidoglycan (PG) (for an excellent review, see Silhavy et al., 2010). Together, the membranes and peptidoglycan are referred to the bacterial envelope, which plays a critical role for cell growth and survival. The envelope serves as a protective barrier to prevent the diffusion of harmful substances from the external environment into the cell, and maintains cell shape and integrity, for instance. The PG is a thin layer of rigid exoskeleton that determines the shape of the bacterial cell. It is composed of repeating units of the disaccharide N-acetyl glucosamine-N-acetyl muramic acid, crosslinked by pentapeptide side chains. The outer membrane has exposed lipopolysaccharides attached to the outer leaflet.

22 INTRODUCTION

Lipoproteins (LPs) are peripheral membrane proteins attached to the membrane by a lipid anchor covalently attached at the N-terminus. In the well-studied model bacterium *Escherichia coli*, approximately 80 distinct LPs are present in the cell envelope (Brokx et al., 2004), each serving a range of critical functions. LPs can be found attached to the OM or IM of Gram-negative bacteria and play diverse and vital roles in bacterial survival (El Rayes et al., 2021a).

Especially OM LPs offer a wide range of functions critical for cell survival and interaction with the host (**Figure 10**). Some LPs serve as essential stress sensors, actively monitoring the integrity and biogenesis of the bacterial envelope (El Rayes et al., 2021a). For instance, the *E. coli* LPs RcsF and NlpE provide molecular connection between the cell envelope and the cytoplasm acting as stress sensors that transduce the stress to the downstream components of the Rcs and Cpx systems (Laloux et al., 2017; Wall et al., 2018; Delhayé et al., 2019).

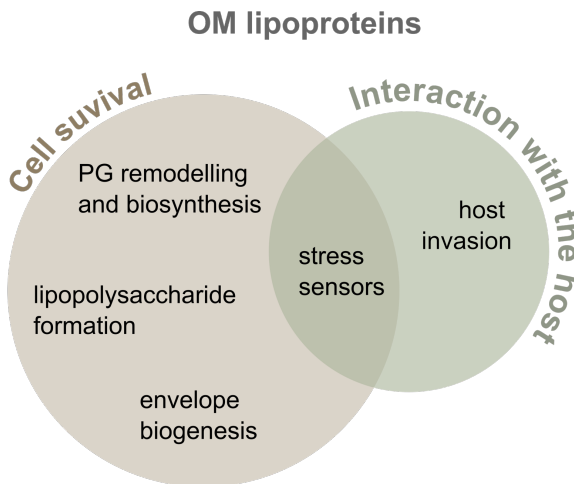


Figure 10: An overview of critical functions played by OM lipoproteins in bacteria. The indispensable roles of OM lipoproteins in cell survival and host interactions underscore their significance in bacterial virulence. Picture copyright © Valentina Furlanetto 2023.

The *E. coli* lipoprotein LptE forms a stable OM complex with LptD, essential for the final stage of the lipopolysaccharide formation and OM insertion (Chimalakonda et al., 2011). The lipoprotein YfiO forming a complex with the lipoproteins YfgL, NlpB and with the outer membrane

protein YaeT is essential for envelope biogenesis, with a key role in the assembly and targeting of outer membrane lipoproteins in *E. coli* (Malinverni et al., 2006).

Bacterial OM LPs also contribute to controlling the architecture of the cell envelope influencing the PG synthesis and remodeling. LpoA and LpoB are required for cell envelope assembly in *E. coli*, forming trans-envelope complexes with polysaccharide polymerases synthesizing PG (Paradis-Bleau et al., 2010; Typas et al., 2010). The *E. coli* Tol-Pal system has been established to play a role in cell division, inducing outer membrane constriction during cell division (Yakhnina et al., 2020).

Moreover, LPs have important functions in bacterial virulence by modulating the bacterium's ability to interact with, and affect, its host and/or environment. The gene *vacJ* coding for an OM LP in *Shigella flexneri* has been identified and characterized as a virulence factor required for intercellular expansion of this pathogen (Suzuki et al., 1994). The lipoprotein FipB from the intracellular pathogen *Francisella tularensis* is required for phagosomal escape and intracellular replication (Qin et al., 2009, 2014). Hence, while the lipid anchor of LPs serves as PAMPs that trigger activation of the innate immune defense system, many LPs also help to evade the PTI mechanisms of the host.

The process of LP biogenesis starts in the cytoplasm where lipoprotein precursors are synthesized (Hayashi et al., 1990) (**Figure 11**). These precursors carry an N-terminal signal peptide comprised of approximately 25 amino acids. Typically, this signal peptide consists of a short segment of positively charged residues, followed by a longer stretch of hydrophobic residues. The final four residues of this signal sequence make up the lipobox, which is a conserved sequence found in lipoproteins (L, A/S, T, G/A, C). The lipobox serves as a determinant for subsequent modifications.

Following the signal sequence is a sorting signal that determines the final localization of the LP, specifying whether it will be retained in the IM or localized to the OM. Approximately half of all OM lipoproteins carry an extensive disordered N-terminal tether that connects the conserved cysteine residue to the folded globular protein (El Rayes, 2021b).

During the translocation of an LP from the cytoplasm to the periplasm, the LP remains anchored in the IM by its hydrophobic signal peptide. The machinery responsible for adding the lipid anchor includes three enzymes: prelipoprotein diacylglycerol transferase (Lgt) (Gennity et al., 1991),

24 INTRODUCTION

prolipoprotein signal peptidase (Lsp) (Innis et al., 1984), and apolipoprotein N-acyltransferase (Lnt). Lgt first attaches a diacylglycerol to the conserved cysteine within the lipobox *via* a thioester bond to form a diacylglyceryl prolipoprotein, after which Lsp cleaves off the signal peptide, leaving the lipid-modified cysteine at the N-terminus of the mature lipoprotein. The role of Lnt is to add a third fatty acid molecule *via* an amide bond to the amino group of the cysteine, resulting in the formation of a triacylated lipid anchor attached to the LP.

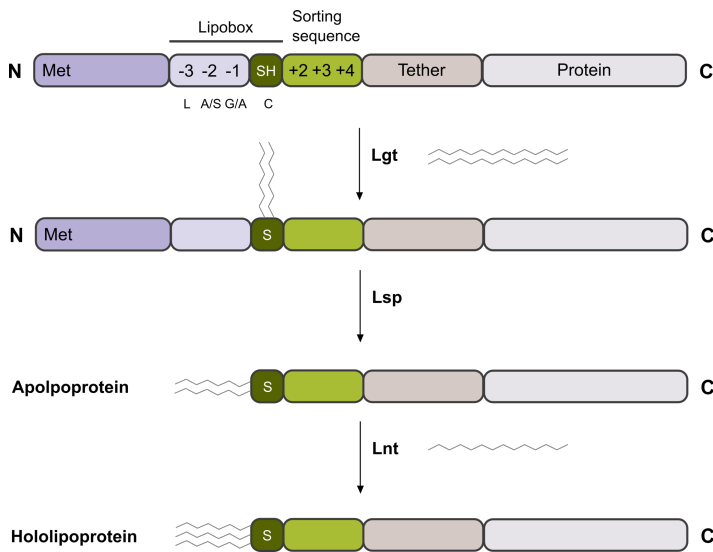


Figure 11: Overview of lipoprotein biogenesis in Gram-negative bacteria. See text for details. Picture copyright © Valentina Furlanetto 2023.

Depending on the sorting signal, the mature LP either remains in IM or is transported to the OM. In the model organism *E. coli*, about 90% of the total LPs are targeted to the OM (Horler et al., 2009). In the case of the latter fate of LPs, the localization of lipoprotein (Lol) pathway of in Gram-negative bacteria is responsible for the extraction of LPs from the IM and transport to the OM (Grabowicz, 2018).

The best studied Lol pathway is that of *E. coli* and includes five proteins, LolA-E (**Figure 12**). LolCDE is an integral membrane protein in IM that belongs to the large family of ATP-binding cassette (ABC) transporters, and specifically the subfamily known as type VII (Thomas et al., 2020). The function of LolCDE is to act as an LP extractor to pull the LP out from IM

by ATP-driven mechanotransmission (Yakushi et al., 2000; Crow et al., 2017), and to deliver the LP to the soluble periplasmic chaperone protein LolA (Matsuyama et al., 1995). LolA interacts with the periplasmic domains of the LolCE heterodimer where it accepts the lipid anchor of the LP and transports it across the periplasm to the OM-attached lipoprotein LolB. LolB binds the lipid anchor of the lipoprotein and inserts it in the OM (Matsuyama et al., 1997).

The function of the OM, including its OM-attached LPs, is closely related to the pathogenicity of Gram-negative bacteria (El Rayes et al., 2021a). Disrupting this system could have a profound impact on the virulence of plant pathogens. Although extensive research has been conducted to understand the mechanisms of the Lol pathway, there is still limited knowledge about the specific molecular transport mechanisms, particularly in non-model organisms like the devastating plant pathogen *Xanthomonas campestris pv. campestris* studied in this thesis (Papers IV and V).

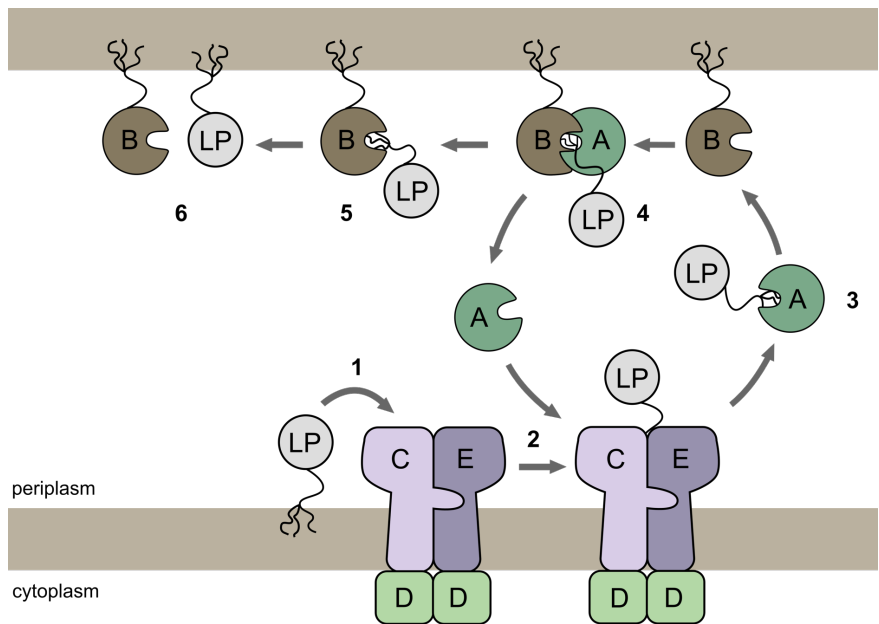


Figure 12: Schematic representation of the *E. coli* Lol pathway. See text for details. Picture copyright © Valentina Furlanetto 2023.

2 Methods

2.1 Cloning and recombinant protein production

2.1.1 Ligation-Independent Cloning (LIC)

Cloning is a process in molecular biology where a gene coding for a target protein is inserted into a plasmid vector, normally a small, circular, double-stranded DNA molecule. This recombinant DNA is then introduced into a host organism, typically *Escherichia coli*, through a process known as transformation, with the plasmid is not integrated into the host's chromosomal DNA. *E. coli* is a convenient expression host since it is well-characterized, grows rapidly to high cell densities, and does not require expensive growth media. It is important to select a vector with suitable characteristics such as antibiotic resistance, recognition sites for promoters, and strategically positioned restriction sites.

While there are many DNA-cloning methods, ligation-independent cloning (LIC) is a versatile and convenient approach that includes several different techniques that do not, unlike traditional cloning methods, depend on the use of restriction enzymes and DNA ligase (Doyle et al., 2005). In comparison to restriction-enzyme cloning, LIC offers a streamlined and highly flexible alternative that does not rely on enzymatic ligation (**Figure13**).

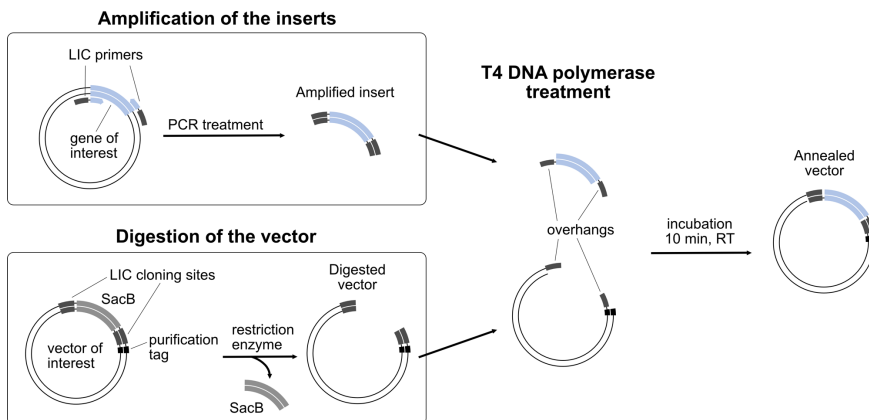


Figure caption on next page.

Figure 13: The LIC technique starts with the amplification of target genes using primers containing overhangs of typically 10-15 base pairs to form a stable association between fragments, which allows for transformation without a ligation step. The target vector is digested at the LIC cloning sites. Once amplified inserts and digested vectors are purified, they are treated separately with T4 DNA polymerase, exploiting its endonuclease activity that removes nucleotides from the exo end. This treatment creates long overhangs that provide enough base pairing to hold the new recombinant DNA molecule when treated inserts and vector are mixed for the annealing, and during the transformation process. Picture adapted from Paper IV (Furlanetto et al., 2023).

2.1.2 Recombinant gene expression

The annealed inserts and vector are typically transformed in *E. coli* through heat shock, where the bacteria are heated at 42°C to allow the new recombinant DNA to enter the bacterial cell, and the immediate transfer to ice restabilizes the bacterial membrane. The recombinant vector typically contains antibiotic resistance gene for clone selection, a specific promoter sequence, and appropriate tags codified for protein purification. The transformed bacteria are plated on agar plates supplemented with the appropriate antibiotic for successful colony selection. The positively transformed bacterial colony is then cultured in the appropriate medium supplemented with the same antibiotic used for colony selection.

The bacterial culture is grown until the exponential phase, at which point the temperature is usually lowered to slow down the bacterial growth before proceeding with the protein production. In the case of the methods used in this thesis, the transcription of the gene of interest is initiated by addition of Isopropyl β -D-1-thiogalactopyranoside (IPTG) that binds the repressor blocking the RNA polymerase bound to the promoter sequence. When IPTG inactivates the repressor, the RNA polymerase can proceed with the transcription of the gene of interest. The obtained mRNA is exported from the nucleus to the cytoplasm where is translated to the target protein. Various factors can be screened to optimize recombinant protein expression, such as transformation method, culture growth condition, choice of vectors, promoters, and protein construct.

2.1.3 Protein purification

The next step is to purify the recombinant protein from other contaminating proteins and macromolecules (**Figure 14**). Harvested bacterial host cells are first lysed to release the cellular content, which can

be achieved by chemical or mechanical means. Sonication that uses ultrasonic waves is a common method for disruption of *E. coli* cells. Once the cells are lysed, the soluble fraction is separated from the membrane debris through ultracentrifugation. In cases where the target is a membrane protein, the membrane fraction is homogenized manually. Subsequently, the membrane proteins are extracted from the lipid fraction using a suitable detergent before protein purification.

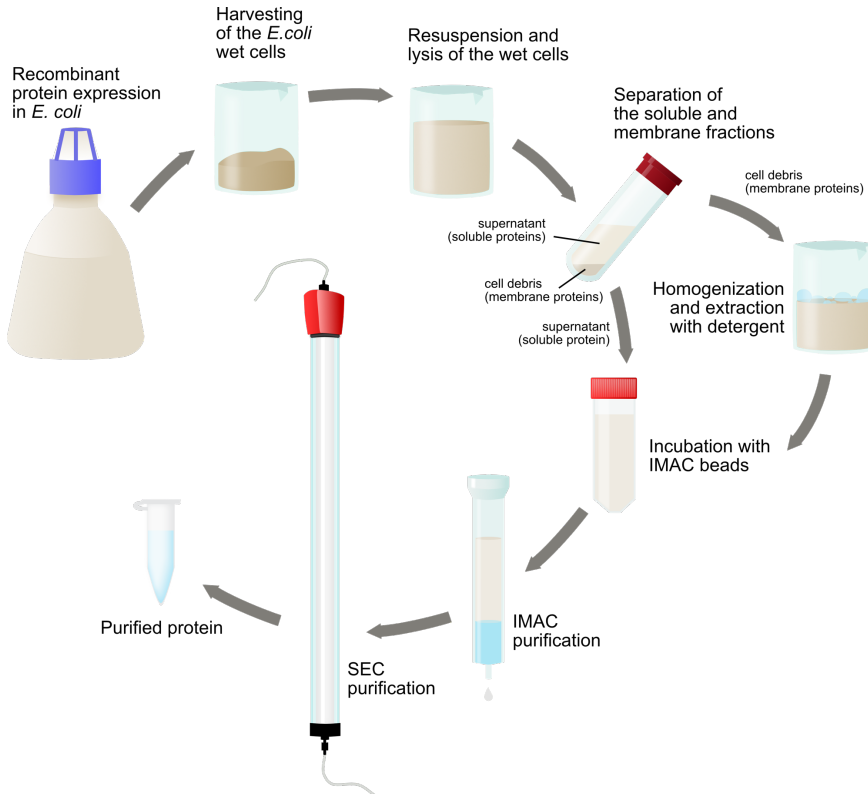


Figure 14: Protein purification workflow. See text for details. Picture adapted from Paper IV (Furlanetto et al., 2023).

Recombinant proteins typically contain an affinity tag that facilitate protein capture from a complex mixture. Immobilized metal-ion affinity chromatography (IMAC) is a common capture technique where the protein is allowed to bind *via* a histidine tag to a metal ion-coupled resin (normally Ni^{2+}). The captured protein can then be eluted by using a buffer containing

e.g., imidazole that competes with the target protein for the resin-bound metal ions to release the target protein.

After capture of the target protein, a suitable second purification step includes size-exclusion chromatography (SEC). As the name implies, SEC separates macromolecules according to size, which is very useful for also determining if a target protein forms higher oligomeric states. The SEC step also serves as a polishing step that eliminates aggregated protein. The method is also useful for buffer exchange. The underlying concept is that small proteins travel more slowly through the column since they can access the inner phase of the porous dextran beads and thereby be delayed to various extents depending on their size, while the larger proteins can not enter the beads and move faster.

2.2 Experimental determination of protein structures

2.2.1 Single-crystal protein X-ray crystallography

Single-crystal protein X-ray crystallography is a powerful experimental technique for determining high-resolution three-dimensional structures of proteins. The method requires that the target protein can arrange itself to form a single ordered three-dimensional lattice.

General requirements for a protein sample to be crystallized include high purity, homogeneity (especially with respect to surface properties), and monodispersity. The experimental setup for protein crystallization can vary, but the key principle is to enable the protein to enter a supersaturated state by adding a suitable precipitant solution to a protein sample at high concentration (McPherson et al., 2014) (**Figure 15A**). The protein molecules will be able to form crystal nuclei only within the zone of supersaturation (a non-equilibrium state where the protein exceeds its solubility limit) and with appropriate kinetics (García-Ruiz, 2003; Rupp, 2015) (**Figure 15B**).

The unit cell is the smallest repetitive unit of the crystal lattice whereas the asymmetric unit is the smallest non-repetitive unit. The number of asymmetric units depends on the crystal's space group, *i.e.*, the unique combination of crystallographic symmetry elements. A typical protein crystal not only contains protein but also a considerable amount of solvent, with an average of ~40-50% of the unit-cell volume, however the solvent

content varies depending on several factors including the space group (Matthews, 1968; Chruszcz et al., 2008).

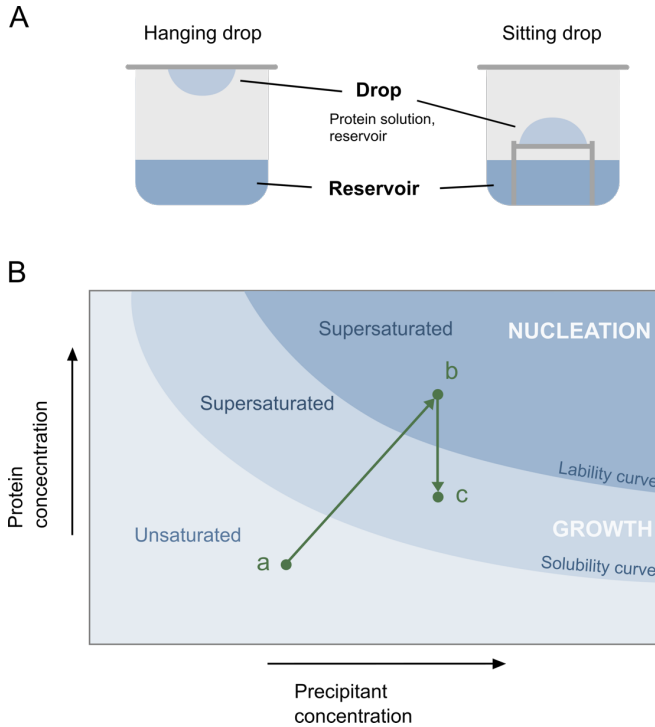


Figure 15: Typical crystallization procedures employ vapor diffusion. A) The crystallization setup can utilize either hanging drop or sitting drop methods, both involving a mixture of protein solution and reservoir solution within the droplet. As the precipitant concentration is higher in the reservoir, vapor diffusion occurs, gradually increasing protein concentration in the drop and promoting crystallization. B) Vapor diffusion begins with an unsaturated protein state (a), and as protein and precipitant concentrations progressively increase, they reach the nucleation zone (b). Nucleation initiates crystal growth, leading to a decrease in protein concentration (c). Picture copyright © Valentina Furlanetto 2023.

The radiation used for macromolecular X-ray crystallography is generated by a synchrotron that accelerates electrons to near the speed of light and filtered to produce a highly intense X-ray beam. Synchrotrons are available to academic researchers, and for the work presented in this thesis the third-generation synchrotron at Diamond Light Source (Oxfordshire, UK) and the fourth-generation synchrotron at MAXIV (Lund, Sweden) have been used.

Bragg's law of diffraction (**Eq. 1**) constitutes the fundamental principle behind the technique and defines the condition, Bragg condition, under which diffraction from a set of lattice planes will occur. Consider a set of parallel lattice planes in real space (in the crystal lattice) with an interplanar spacing d that is exposed to radiation with a wavelength λ that is close to the atomic separation (typically 1 Å, corresponding to 0.1 nm).

$$\text{Eq. 1} \qquad 2 d_{hkl} \sin \theta = n \lambda$$

When (1) the incident angle θ to the plane of two incoming waves is equal to the scattering angle, and (2) the difference in pathlength between the waves scattered from individual planes equals the interplanar distance d , then the Bragg condition is fulfilled, and the waves interfere constructively to give rise to a Bragg reflection. These reflections are recorded on a photon-counting X-ray detector.

The intensity of each recorded Bragg reflection is proportional to the amplitude of the *structure factor* (\mathbf{F}_{hkl}) which is a vector in reciprocal space (diffraction space). The structure factor represents the Fourier transform (FT) of the set of lattice planes from which the reflection originated, and to collect a complete dataset of diffraction intensities, all or at least most of reciprocal space must be sampled. This requires that the crystal is rotated in the synchrotron beam, typically in 0.1° steps, until all possible sets of lattice planes have come in position to fulfill the Bragg condition to generate a reflection (**Figure 16**).

Each structure factor contains information about the entire content of the crystal's unit cell and is a convolution of the FT of the molecule and the FT of the crystal lattice. The presence of the protein molecule inside a crystal lattice has decomposed the continuous FT of the molecule's electron density into a discrete spectrum representing the individual FTs of the plane sets, i.e., the crystal has performed a *Fourier analysis*.

Since the intensities of the recorded reflections only provide the amplitudes of the structure factors, the missing phase component of each structure factor needs to be determined by other methods. This is known to as the *phase problem* in X-ray crystallography. There are different methods to retrieve the missing phase information and the method used in this thesis is known as molecular replacement (MR).

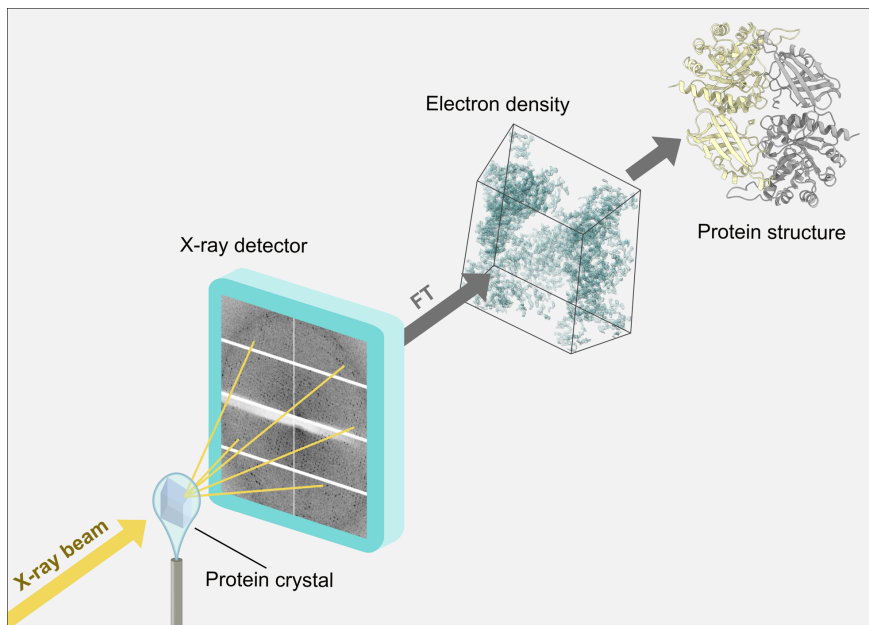


Figure 16: Schematic figure of the data collection experiment. See text for details. Picture copyright © Valentina Furlanetto 2023.

To describe the MR method, the Patterson function (PF) must be introduced. PF is a self-correlation function, *i.e.*, a correlation function of the electron density with itself. It is a convolution product of the electron density in two points separated by a vector, which in **Eq. 2** is expressed for a one-dimensional case where the electron density at a point x and a point $x+u$ is multiplied.

$$\text{Eq. 2} \quad P(\mathbf{u}) = \int \rho(\mathbf{x}) \rho(\mathbf{x} + \mathbf{u}) dV$$

This can be viewed as moving through the unit cell with a vector \mathbf{u} and in every position multiply the electron density at the beginning of \mathbf{u} with the electron density at the end of the vector, followed by taking the integral of all values. The formal expression of PF (**Eq. 3**) shows that while the structure factors themselves contain both amplitude and phase, the PF loses the phase component and only retains the square the amplitudes (which is approximated as the intensity of the recorded reflections), which means the PF can be calculated without phase information.

$$\text{Eq. 3} \quad P(\mathbf{u}, \mathbf{v}, \mathbf{w}) = (1/V) \sum_h \sum_k \sum_l |\mathbf{F}_{hkl}|^2 \cdot \cos [2\pi (hu + kv + lw)]$$

For every unique vector between two points there will be second vector pointing in the opposite direction with the same length, which means PF is centrosymmetric. This gives that PF is a convolution of the electron density in a certain point $\rho(x,y,z)$ and its centrosymmetric image $\rho(-x,-y,-z)$, which is a multiplication of the structure factor with its complex conjugate to give a real number, which is the reason for the removal of the phase component of the electron-density equation.

MR requires an existing protein model with sufficiently high sequence identity to that of the target protein for which the structure will be determined. The model is generated either by classical comparative template-based modeling (for instance Modeller; Webb et al., 2016) or by deep-learning algorithms like AlphaFold2 (Jumper et al., 2021). The resulting homology model is then used as a *search model* for MR calculations.

Patterson functions are calculated for the intensity data collected for the target protein crystal (observed PF) and for the search model (model PF), which result in two vector representations, each with a high level of noise. The traditional approach is to try to find the orientation of the model PF in the asymmetric unit of the observed PF by testing all possible orientations (rotations and translations).

When a possible orientation is found for the search model in the experimental data, model amplitudes (F_c) and phases (φ_c) from the placed model are calculated. The φ_c for each calculated model structure factor is then combined with the corresponding observed structure-factor amplitude obtained from the diffraction experiment.

With observed amplitudes and calculated phases for all possible structure factors (Fourier representations of possible plane sets), the electron-density equation (**Eq. 4**) can be used to perform a *Fourier summation* to produce an electron-density map, which is the actual experimental structure from which a model is built.

$$\text{Eq. 4} \quad \rho_{(xyz)} = 1/V \sum_h \sum_k \sum_l \mathbf{F}_{(hkl)} \cdot e^{-2\pi i (hx + ky + lz)}$$

2.2.2 Single-particle cryo-electron microscopy

The methodology of single-particle cryo-electron microscopy (cryo-EM) is a powerful technique employed in structural biology to determine high-resolution 3D structures of proteins and other biological macromolecules.

Cryo-EM involves the imaging of proteins using an electron microscope, utilizing an electron beam with an energy of approximately 200-300 keV, corresponding to a wavelength (λ) of around 0.025-0.019 Å. In this technique, the specimen, typically in an aqueous solution, is loaded onto grids. For single-particle reconstruction, it undergoes blotting and vitrification in liquid ethane, followed by storage in liquid nitrogen. The process of plunge vitrification requires a rapid cooling rate of approximately 10^5 K/s, resulting in the specimen being vitrified into a thin layer measuring 1-2 μm in thickness. Vitrification is essential for achieving the amorphous phase necessary for this methodology.

Experimental information is carried by unscattered electrons that traverse the specimen, constituting approximately 80% of the total electrons in the beam. Approximately 5% of electrons undergo elastic scattering, retaining their original energy but changing direction and phase upon passing through the specimen. The electrons in the beam interact strongly with the sample, leading to various outcomes, including backscattering, cathodoluminescence generation, and the liberation of secondary electrons, x-rays, and Auger electrons from the sample. Additionally, about 15% of electrons experience inelastic scattering, resulting in energy loss; these electrons are eliminated through an energy filter.

The amorphous phase in which the sample is vitrified enables the collection of multiple projections of individual particles using fast direct detection methods. Modern detectors collect numerous images, or frames, that constitute movies. The dose ($\text{e}/\text{\AA}^2$) increases with time, so most exposed frames are used for the initial low-resolution reconstruction.

The general steps during and after data collection consist of frame collection, that compose the movie. The movie gets aligned and averaged, allowing the defocus determination and contrast-transfer-function (CTF) correction. Defocus refers to the difference in the focal distance between the specimen and the electron microscope's objective lens. It is a crucial parameter in cryo-EM because it affects the contrast and resolution of the images, and accurate defocus determination is essential for achieving high-quality 3D reconstructions. To determine the defocus of an image, a series of images is typically recorded at different focus settings (varying the distance between the specimen and the objective lens). These images exhibit characteristic contrast patterns, known as Thon rings or CTF oscillations, caused by the interference between the focused and defocused

electrons. The spatial frequency and position of these rings provide information about the defocus value. Defocus determination often involves analyzing the contrast transfer function (CTF) of the images. The CTF describes how the microscope modulates the contrast of the specimen's features as a function of spatial frequency. By measuring the positions and characteristics of the Thon rings in the power spectrum of the image, the defocus value can be calculated.

From the averaged movie, the particles can be picked, extracted, and normalized. Particle picking involves identifying and selecting the individual particle images within the raw micrographs. After particle picking, the selected particle images are extracted from the micrographs as individual 2D images, cropping a region of interest around each particle's position in the micrograph. Normalization is typically performed by subtracting the mean pixel value of the particle image and dividing by the standard deviation. This process transforms the pixel values to have a mean of zero and a standard deviation of one.

Once the particles are extracted there is the process of 2D classification, where 2D classes are selected to proceed to build 3D models for the 3D classification. Again, after 3D classification the best models can be selected and undergo refinement. During the refinement various things can be refined, such as defocus, different orientation of different particles, angles and so on. The Fourier Shell Correlation (FSC) gives information about the resolution of the model, that corresponds to an accepted cutoff of 0.143. Information can be drawn also by the local resolution of the model.

In the post processing phase, the refined map is sharpened, enhancing the clarity and resolution of a 3D model, reducing the B-factors, or temperature factors. The lower the B-factor, the sharper the map. The enhancement of the map quality after sharpening is assessed through the evaluation of the FSC function. The sharpened density map is a valuable starting point for model building, where the protein peptide chain is fitted in the experimentally determined density. In the past couple of years, AlphaFold2 is routinely used to obtain starting models that are fitted in the experimental maps.

2.3 Other biophysical methods

2.3.1 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is an analytical technique used to characterize the thermodynamic profiling of molecular interactions, typically those involving macromolecules with other macromolecules or smaller compounds. ITC enables the measurement of the heat changes (or enthalpy, ΔH) associated with these interactions, allowing for the determination of essential thermodynamic parameters such as dissociation constants (K_D), Gibbs free energy changes (ΔG), entropy changes (ΔS), and the stoichiometry of the interaction (n).

In an ITC experiment, the setup consists of two identical cells: one containing water (the reference cell) and the other housing the sample of interest (the sample cell). These cells are constructed from highly conductive materials and are enclosed within an adiabatic jacket. The process involves the titration of a ligand into the sample cell, done in precisely known aliquots while maintaining continuous stirring.

The initial conditions of the experiment depend on the estimated K_D , which is correlated to the concentration in the sample cell ([ligand]) through a parameter known as the C-value (**Eq. 5**). This parameter helps determine the optimal starting conditions for the titration process.

$$\text{Eq. 5} \quad C = \frac{[\text{ligand}]n}{K_D}$$

In ITC experiments, an optimal C-value typically falls within the range of 10 to 50. To achieve this, the concentration of the analyte in the syringe ([analyte]) should be 10 to 20 times higher than the concentration of the ligand in the sample cell ([ligand]).

During the titration process, the ligand is incrementally injected into the sample cell, and the resulting change in heat is measured in comparison to the reference cell. This generates a thermogram, which is a graphical representation of the heat changes over the course of the titration.

The titration curve, which provides valuable insights into the interaction, is derived by integrating the area of each injection peak and plotting these individual values against the molar ratio of the binding event. By analyzing this curve, the thermodynamic parameters associated with the interaction can be determined (**Eq. 6**). ITC's ability to extract these thermodynamic parameters makes it an indispensable tool for

studying molecular interactions and understanding their underlying energetics.

$$\text{Eq. 6} \quad \Delta G = -RT \ln K = \Delta H - T \Delta S$$

2.3.2 Mass spectrometry

Mass spectrometry (MS) stands as a versatile and powerful analytical method; its primary function is to determine the chemical composition of molecules by precisely measuring their mass-to-charge ratio (m/z).

In the context of MS, the first step is ionization, where molecules are converted into ions suitable for mass analysis. Various techniques, such as direct injection, vaporization, or electrospray ionization (ESI) for liquid samples and laser ablation or thermal desorption for solid samples, introduce these samples into the mass spectrometer.

One such ionization method is Electrospray Ionization (ESI) is used for analyzing polar and thermally labile compounds, such as complex protein mixtures or smaller molecules. It operates by generating ions from a sample in solution via an electrospray. Following ionization, the released ions with distinct mass-to-charge ratios (m/z) are propelled into the mass spectrometer for subsequent analysis. ESI-MS was used in Paper I for the analysis of the products of the enzymatic reaction (Chapter 3.2).

Matrix Assisted Laser Desorption/Ionization (MALDI), used for the analysis of large biomolecules such as proteins, peptides, and nucleic acids. During MALDI, the analyte of interest is combined with a matrix material, typically a crystalline organic compound. The matrix's role is twofold: it assists in vaporizing the analyte, preparing it for ionization, and shields it from the potentially destructive effects of the laser radiation employed in the process, thereby minimizing fragmentation. The matrix-analyte mixture is then irradiated with a laser beam, which initiates desorption and ionization, for further MS analysis. This method has been used in Papers I, II, III for the protein m/z analysis (Chapter 3.2, 3.3, 3.4). In Paper III the product analysis of the enzymatic reaction was performed using the quadrupole Time-of-Flight (TOF) MS.

Quadrupoles use four parallel rods with variable voltage for the selective transmission ions according to their m/z . The ions are then propelled in the TOF mass spectrophotometer that features a component known as an "ion mirror" or reflectron that employs an electric field to reflect ions back toward the detector. The critical advantage of this setup is that ions with

different m/z values experience varying flight times, primarily due to their different velocities. As a result, the ions with higher mass take longer paths and ultimately arrive later at the detector. This feature greatly enhances the resolution of the mass spectrometer, enabling the precise measurement of m/z values even for molecules with very similar masses.

The resulting spectrum consists of different peaks, where the base peak is the peak presenting the highest intensity and can corresponds to the molecular peak (or parent peak) that reflects the m/z of the ionized molecule. The ionized molecule can undergo fragmentation, detected in MS as multiple peaks at lower molecular weight.

2.3.3 High-performance liquid chromatography

High-performance liquid chromatography (HPLC) is a technique in analytical chemistry used for the separation, identification, and quantification of compounds in a mixture. This is accomplished by exploiting the interactions of the compounds with both a liquid mobile phase and a solid stationary phase, which is packed into a column. Notably, HPLC differs from traditional liquid chromatography by the application of high pressures, typically ranging between 50 and 350 bar.

Various mobile phases, usually comprising solvents or solvent mixtures, can be applied in HPLC, and diverse types of columns are available to accommodate different analytical needs. The choice of stationary phase within the column is contingent upon factors such as the chemical properties of the material and the size and shape of the particles it contains.

In the HPLC process, the sample solution is actively pumped through a column filled with solid adsorbent material. Each constituent within the mixture interacts uniquely with this material, resulting in distinct flow rates and, consequently, varied retention times. These differential retention times manifest in the form of peaks within the resulting chromatogram, each peak corresponding to an individual compound within the mixture.

HPLC was used in Papers I, II and III for the analysis of the enzymatic reaction products, utilizing silica gel as stationary phase, detecting the products via UV-Visible analysis (Chapter 3.2, 3.3, 3.4).

2.4 Phylogenetic analysis

Protein sequences for the proteins studied were analyzed with respect to their relationships and distribution. Genome, sequence and structure databases were used: GenBank (<https://www.ncbi.nlm.nih.gov/genbank>); Uniprot (<https://www.uniprot.org>); PDB (<https://www.rcsb.org>); AlphaFold database (<https://alphafold.ebi.ac.uk>). Bacterial phyla were annotated based on NCBI Taxonomy Browser (Schoch et al., 2020). MicrobesOnline (<https://www.microbesonline.org>) and the NCBI Sequence Viewer (<https://www.ncbi.nlm.nih.gov/projects/sviewer/>) were used to analyze and confirm the operon architecture.

Homologous sequences were identified using tools such as BLAST® (Basic Local Alignment Search Tool; Altschul et al., 1990) or fold recognition through FoldSeek (<https://search.foldseek.com/search>; van Kempen et al. 2023) against predicted protein models in AlphaFoldDB (<https://alphafold.ebi.ac.uk>). MSA is performed with software like Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>), which calculates the phylogenetic tree.

Multiple sequence alignments (MSA) were made using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>; Sievers et al., 2011) and visualized using ESPript 3.0 (Robert et al., 2014). MSAs for phylogenetic analyses were generated using MUSCLE as implemented in MEGAX (Kumar et al., 2018) followed by editing and visualization using Interactive Tree of Life, iTOL (Letunic et al., 2021).

3 Present investigation

3.1 Aim of the thesis

The aim of this thesis is to provide new insights into the molecular mechanisms underlying communication between bacteria and plants, focusing on both symbiotic and pathogenic interactions.

The first part of the thesis focuses on bacterial deglycosylation of glycosylated aromatic polyketides produced of plant origin, specifically by the beneficial soil bacteria, including *Deinococcus aerius*, *Streptomyces canus*, and *Microbacterium testaceum*. Results from biochemical, functional, structural, and phylogenetic analyses contribute to new knowledge on enzymatic C-glycosyl deglycosylation in bacteria and thereby sheds further light on important enzymatic mechanisms that support symbiotic communication between plants and beneficial soil bacteria.

The second part of the thesis focuses on characterization of the lipoprotein-localization system of the devastating bacterial plant pathogen *Xanthomonas campestris* *pv.* *campestris*. The rationale for studying this system is based on the importance of lipoproteins and their proper localization for bacterial viability, and for the pathogenicity and virulence capacity of the bacteria. The unique aspects of the Xcc Lol system discussed in this thesis offer new insights that can help to address improved strategies for targeting the pathogen.

3.2 Paper I. C-glycosyl deglycosylation in *D. aerius*

3.2.1 *Deinococcus aerius*

The bacterium *Deinococcus aerius* is an orange-pigmented, coccoid bacterial strain isolated from the high atmosphere above the main island of Japan (Yang et al., 2009). *Deinococcus aerius* is known for its tolerance to desiccation, UV-C and gamma radiation and strains of the *Deinococcus* family have been isolated and grown from a variety of sites, including radioactive and arid soils, animal feces, and meat (Wang et al., 2010; Peng et al., 2009; Rainey et al., 2005; Ito et al., 1983; Davis et al., 1963).

While it is challenging to determine the natural habitat of this organism, studies have been performed on the bacterial population of soil in extreme environments as in the Sonoran Desert, detecting bacteria belonging to the genus *Deinococcus* (Rainey et al., 2005). *Deinococcus* has been also isolated as soil bacteria from plant rhizosphere (Lai et al., 2006). Bacteria belonging to this genus are ionizing-radiation-resistant bacteria thanks to their heightened DNA repair capacity, that provides them with a selective advantage in arid environments (Rainey et al., 2005). Studies on *Nicotiana attenuata* highlighted the increased colonization of *Deinococcus* species under UVB radiation exposure (Santhanam et al., 2017).

Since *Deinococcus* is an aerobic and chemoorganotrophic bacterium it is likely to live in rich organic habitats but also resist extreme conditions, for instance on dispersed dust particles. Although *D. aerius* has been tested negative to nitrate reduction, the studies conducted on extreme soil environments demonstrated that *Deinococcus* species have developed the ability to perform enhanced DNA repair when subjected to arid and radioactive soils (Mattimore et al., 1996; Rainey et al., 2005).

Analysis of the *D. aerius* TR0125 genome revealed a gene cluster with similarities to previously identified gene clusters that code for enzymes that catalyze deglycosylation of plant C-glycosyl flavonoids produced by plants (Chapter 1.3). This cluster included the gene DAERI_200051 coding for a GMC-type oxidoreductase, *cgo* (Chapter 1.3.1) that we refer to as DaCGO1, and two additional genes, DAERI_200053 (*cgdα*) and DAERI_200052 (*cgdβ*) predicted to correspond to the α- and β-subunits of a CGD complex, DaCGD (Chapter 1.3.2). To confirm the predicted functions of these genes, all three gene products were characterized.

3.2.2 Biochemical characterization of the CGD gene cluster

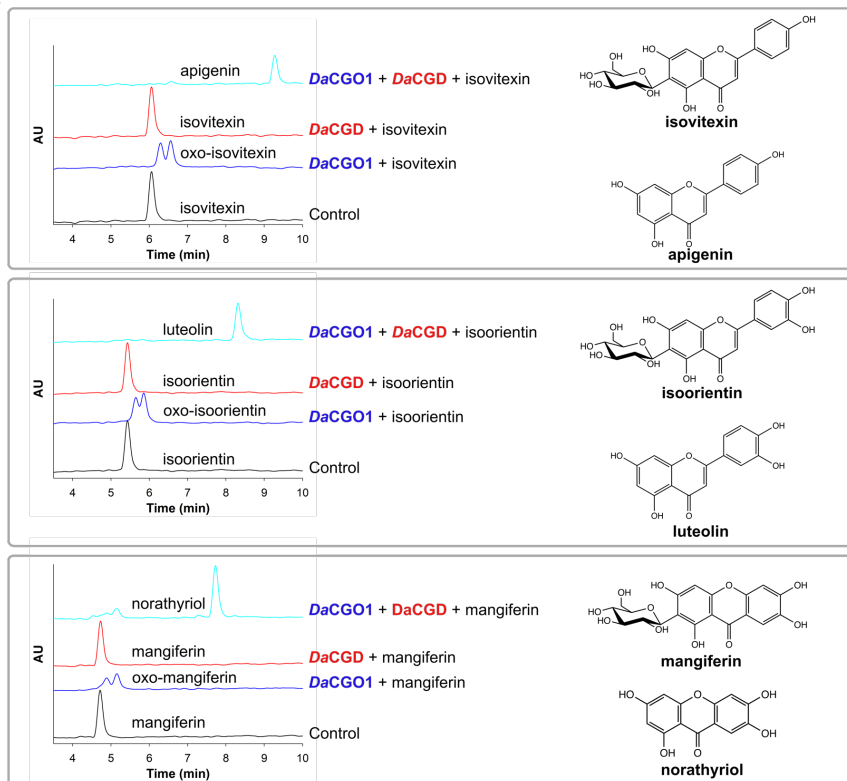
DaCGO1 and *DaCGD* were expressed recombinantly in *E. coli* and the purified proteins characterized biochemically. SEC, SDS-PAGE, spectroscopic and activity analyses showed that *DaCGO1* is a monomeric metal-independent and FAD-dependent enzyme with the FAD non-covalently bound, and that *DaCGD* is a metal-dependent enzyme that assembly as an $\alpha_2\beta_2$ complex.

Several potential electron-donor substrates were tested for *DaCGO1*: the O-glucoside flavonoid genistin; the C-glucosyl flavonoids puerarin, orientin, vitexin, isovitexin, and isoorientin; the C-glucosyl xanthone mangiferin; and the C-glucosyl hydroxyanthrapurin carminic acid. The reactions with *DaCGO1* and oxygen as electron acceptor for the reductive half-reaction revealed specificity for the 6-C-glucosyl flavonoids isovitexin, isoorientin and for the 2-C-glucosyl xanthone mangiferin, with the highest activity for isovitexin. HPLC analyses of the oxidation products for isovitexin, isoorientin and mangiferin showed the progressive formation of two distinct peaks, possibly isomers of the oxidized substrates (**Figure 17A**; Nakamura et al., 2019).

The kinetic characterization of the combined *DaCGO1/DaCGD* reaction was performed using isovitexin and isoorientin as substrates (**Figure 17B**). A series of divalent metal cations were tested in one-pot *DaCGO1/DaCGD* assays, and only minor variation in activity was observed, but with a small advantage for Mn^{2+} . Based on HPLC, oxidation products from the *DaCGO1* reaction were gradually cleaved by *DaCGD* and eventually converted to a single peak corresponding to the free aglycone. The *DaCGD* reaction was considerably slower than the *DaCGO1* reaction, and hence, C-C bond cleavage determines the overall reaction rate for the one-pot reaction.

DaCGD was not able to cleave the C-C bond in non-oxidized C-glycosylated substrates, thus confirming the results from previous studies (Chapter 1.3) that a glucosyl 3-keto group is required for deglycosylation by *DaCGD*. The substrate preference of *DaCGD* resembles that of *AgCGD1* and *MtCGD* from the soil bacteria *M. trichothecenolyticum* and *A. globiformis*, and to *EuCGD* from intestinal bacterium *E. cellulosolvens* (Mori et al., 2021).

A



B

Enzyme, substrate	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($s^{-1}mM^{-1}$)
DaCGO1 + isovitexin	0.047 ± 0.008	171 ± 10.6	3625 ± 225
DaCGO1 + isoorientin	0.021 ± 0.002	33 ± 0.86	1237 ± 32
DaCGO1 + mangiferin	0.064 ± 0.01	13 ± 0.81	204 ± 12
DaCGO1 + DaCGD + isoorientin	0.083 ± 0.014	0.42 ± 0.02	4.99 ± 1.3
DaCGO1 + DaCGD + isovitexin	0.108 ± 0.03	0.29 ± 0.021	2.69 ± 0.7

Figure 17: (A) Results from HPLC analysis of the reaction products from *DaCGO1*, *DaCGD* and *DaCGO1/DaCGD* reactions (left) with the tested substrates and aglycone products (right). (B) Kinetic constants for the *DaCGO1* and *DaCGO1/DaCGD* reactions based on nonlinear regression fitting to the Michaelis-Menten equation (triplicate measurements). Picture copyright © Valentina Furlanetto 2023.

3.2.3 Structure determination of *DaCGO1* and *DaCGD*

The crystal structure of *DaCGO1* showed that the enzyme is monomeric and shares a monomer fold like those of other GMC oxidoreductases of CAZy family AA3 (www.cazy.org). The family AA3 structures are characterized by a *p*-hydroxybenzoate hydroxylase (PHBH)-like fold (Wierenga et al 1979; Mattevi, 1998) with an FAD-binding domain having a Rossmann-like $\beta\alpha\beta$ mononucleotide-binding motif, and a substrate-binding domain with a central 6-stranded β -sheet and two α -helices.

The GMC oxidoreductases also have a loop-and-lid region typically involved in oligomerization (Hallberg et al., 2004; Hallberg et al., 2002). In the monomeric *DaCGO1* this region does not have a function in oligomerization and includes two α -helices in the loop region and two β -strands in the lid region (**Figure 18A**). The active site, located at the interface between the FAD-binding and substrate-binding domains, is gated by an active-site loop. In the experimental structure of *DaCGO1* this loop is in a closed state and has an unusual proline signature with the sequence PIPIP.

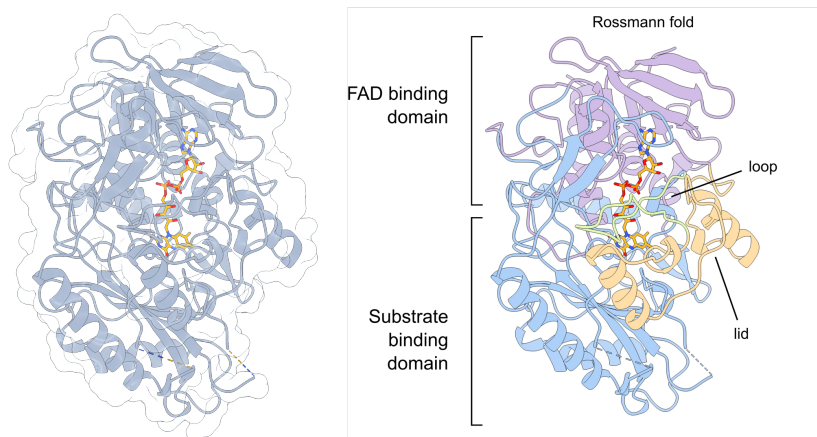
The structure of *DaCGD* was determined experimentally using both cryo-EM and X-ray crystallography, and the results confirmed a novel $\alpha_2\beta_2$ oligomerization state of this CGD (**Figure 18B**). As for the structures of *PuCGD* and *AgCGD2* (Chapter 1.3.2; Mori et al., 2021), the lid domain in the α -subunit of *DaCGD* folds as a 4-helix bundle. Interestingly, in the structures of *DaCGD*, the lid domains are located at the interface between the two heterodimers, and the β -subunits are pointing outwards.

The *DaCGD* structures produced by X-ray crystallography and cryo-EM are very similar except for minor displacements of the β -subunit and the 4-helix bundles in the α -subunits. This emphasizes that these regions are dynamic to some degree, which is likely to be required for substrate binding. The corresponding regions in other CGDs have been shown to be flexible, as has been discussed in Chapter 1.3.2. The difference however is that in *DaCGD* the 4-helix bundle and β -subunit are at the interface between two heterodimers, which is likely to restrict the extent of conformational changes compared with CGDs that have these regions facing the surrounding solvent.

The coordination sphere for the catalytic metal is slightly different in *DaCGD* compared with that in other structurally determined CGDs, which may indicate some differences relating to the reaction mechanism. Like in

PuCGD discussed in Chapter 1.3.2, the metal is coordinated by two water molecules instead of only one as in *AgCGD2* and *EuCGD*. In *DaCGD*, one of these two metal-coordinated water molecules is also coordinated by a conserved lysine side chain forming an unusually short interaction of 2.07 Å. This interaction is probably required for generating the catalytic hydroxide needed to initiate proton abstraction from the aglycone moiety.

A



B

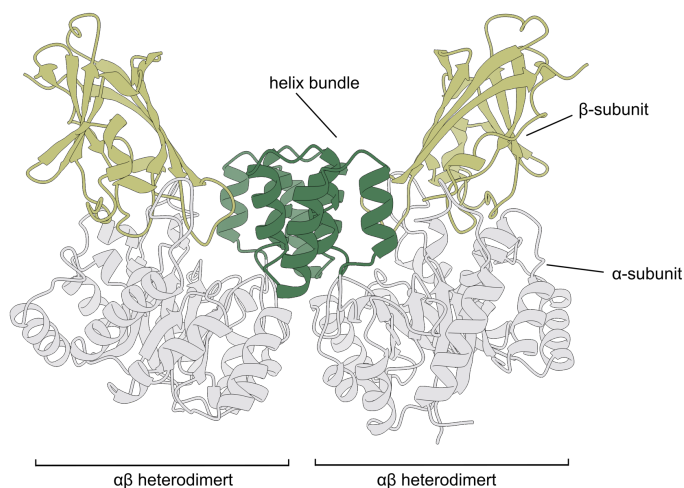


Figure 18: Experimental structures of *DaCGO1* and *DaCGD*. (A) The monomeric *DaCGO1* shows the typical GMC-oxidoreductase fold but with a more compact loop-and-lid region. (B) Crystal structure of *DaCGD* showing the novel $\alpha_2\beta_2$ assembly. The 4-helix bundle in the α -subunit is located at the interface of the two heterodimers.

3.2.4 C-glycosyl deglycosylation mechanism

To evaluate possible structural determinants of substrate binding in *DaCGO1* and *DaCGD*, the substrates isovitexin, isoorientin and mangiferin were docked manually in the experimental structures to produce theoretical hypotheses of substrate binding.

To dock substrates in *DaCGO1*, the closed active-site loop had to be removed, and the docking of the substrates was guided by the experimental structures available for P2Os with bound monosaccharides (Kujawa et al., 2006; Bannwarth et al., 2006; Tan et al., 2010; Spadiut et al., 2010; Tan et al., 2011; Hassan et al., 2013; Tan et al., 2014), and by available mechanistic data (Sucharitakul et al., 2011; Prongjit et al., 2013; Wongnate et al., 2011; 2014; 2019) (**Figure 19A**). A productive docking pose requires that the C3'' atom, *i.e.*, the site of oxidative attack, is oriented in front of and slightly below the FAD N5 atom, and that the catalytic histidine (His440) is positioned suitably for proton abstraction from the keto sugar product. While the aglycone in isoorientin and isovitexin have some degree of rotational freedom around the C1'' (sp^3)–C6 (sp^2) bond, the most favorable orientation placed the C5-OH/C4=O pair of the aglycone pointing towards the flavin isoalloxazine.

The oxidoreductive mechanism performed by GMC enzymes can be divided into a reductive half-reaction, where the FAD is reduced and the substrate is oxidized, and an oxidative half-reaction, where FADH⁻ is re-oxidized (**Figure 19B**). The glycosyl-carbon position to be oxidized depends on the specificity of the individual GMC enzyme. The mechanism relies on a conserved His-Asn pair at the *re* side of the isoalloxazine ring (His440 and Asn484 in *DaCGO1*). Extensive studies have been performed to elucidate the reaction mechanisms of these two half-reactions, most importantly the studies performed on the fungal pyranose 2-oxidase *ToP2O* discussed in Chapter 1.3.1 (Sucharitakul et al., 2011; Prongjit et al., 2013; Wongnate et al., 2011; 2014; 2019).

The most recent mechanistic data for *ToP2O* (Chapter 1.3.1; Wongnate et al., 2019) state that the reductive half-reaction is initiated by hydride transfer from the C2 position to be oxidized to the N5 atom of the flavin with the generation of a protonated ketone intermediate. The conserved histidine proceeds with the abstraction of the proton from the protonated ketone intermediate, producing the keto-sugar product. The oxidative half-reaction requires the conserved histidine to activate molecular oxygen by offering it a proton with the concomitant transfer of an electron from

FADH^- to O_2 forming flavin semiquinone- OOH radical pair, formation of a C_4a -hydroperoxyflavin intermediate, and finally the elimination of H_2O_2 to obtain the re-oxidized FAD (Wongnate et al., 2014).

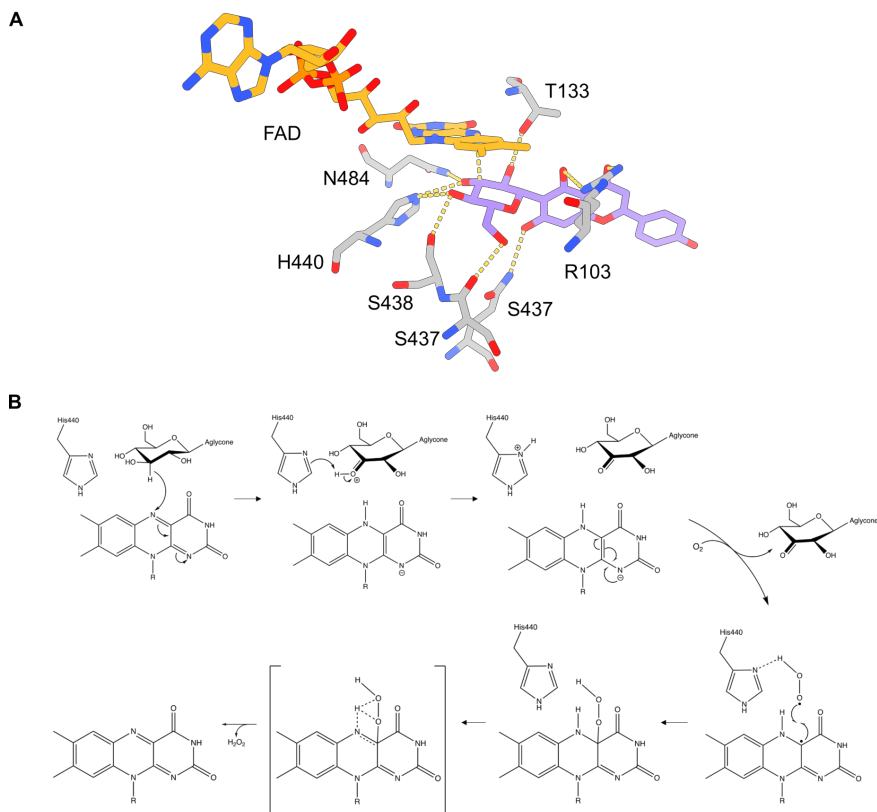


Figure 19: Docking of substrate to *DaCGO1*. (A) Isovitexin docked in the *DaCGO1* active site. (B) Proposed mechanism of *DaCGO1* for C_3'' oxidation of the sugar moiety in isovitexin, based on the mechanism of *ToP2O* by Wongnate et al., 2014.

The proposed C-deglycosylation mechanism performed by CGDs suggests that a metal-coordinated hydroxide abstracts a hydrogen from the carbon atom in the aglycone placed in *ortho* or *para* position of the bond to be cleaved. This event initiates the deprotonation of the catalytic histidine that then acts as a base to abstract a hydrogen from the glucosyl in the slightly acidic C_2'' position. The negatively charged glucosyl species and the broken aromaticity in the aglycone moiety enable C-C cleavage through a β -elimination-like reaction (**Figure 20A**).

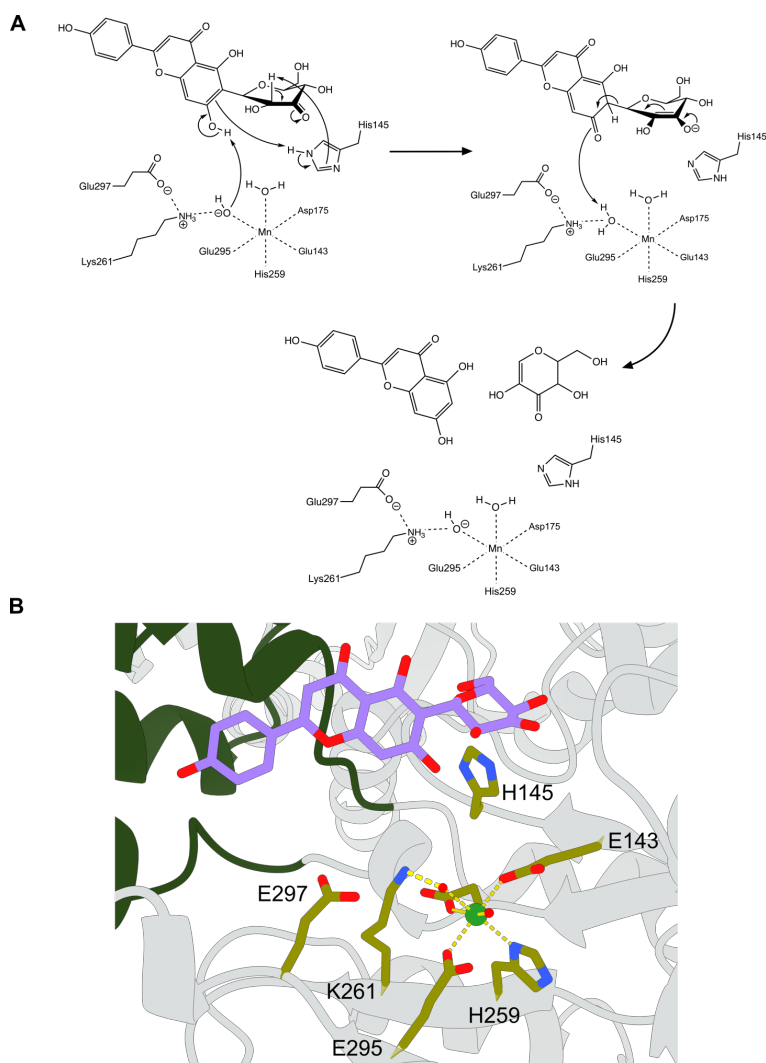


Figure 20: CGD mechanism. (A) Proposed C-C bond cleavage mechanism of *DaCGD*. (B) 3''-oxo-isovitexin docked at the *DaCGD* active site.

A recent study from Choi and co-workers challenges this mechanism in the context of the C-C cleavage of 3''-oxo-puerarin from *PuCGD*, based on DFT calculations (Choi et al., 2023). According to their study the mechanism does not undergo the formation of the quinoid intermediate, and the Michael-type β -elimination of the dehydro-3''-oxo-puerarin causes the C-

C cleavage and the formation of the final products. While Choi et al.'s calculations offer valuable insights into an alternative potential C-C cleavage mechanism of 3"-oxo-puerarin from *PuCGD*, their considerations may not directly apply to the cleavage of 3"-oxo 6C-glycosyl compounds.

The experimental structures of *DaCGD* were found to be in the open conformation, thus facilitating docking of the substrates (**Figure 20B**). The crystallographic electron density shows bound water molecules and metal ions that traced out the sugar-binding subsite, which aided positioning of the glucosyl moiety. The glucosyl group could be oriented such that C2" would be coordinated by the catalytic histidine (His145), as required for C-C bond cleavage of C3"-oxo compounds in the previously proposed mechanism.

While the docked sugar would be stabilized by interactions in the active site, the open conformation does not support proximity of the metal-coordinated hydroxyl group and the aglycone moiety, implying that a conformational change, possibly by movements of both the 4-helix bundle and the β -subunit, is needed for a catalytically competent active-site conformation. Without additional experimental data, it is not possible to predict the extent or effects of such a conformational change.

As already mentioned, the substrate preference of *DaCGD* resembles that of *AgCGD1*, *MtCGD* and to some extent that of *EuCGD*. Of these enzymes, only the structure of *EuCGD* has been experimentally determined (Mori et al., 2021). *EuCGD* does not have a lid that forms a 4-helix bundle but instead a long, disordered loop. The 4-helix bundle is however present in *PuCGD* and *AgCGD2* that are active towards the oxo forms of C-8 glycosyl flavonoids such as 3-oxo puerarin and 3-oxo orientin.

Further considerations can be made about the specificity of *DaCGD* with respect to the reactions performed by *DaCGO1*. As mentioned above, the docking of substrates in *DaCGO1* allowed positioning of the sugar moiety to enable oxidation in C3". A docking pose with C2" as the site of oxidative attack would cause steric clashes. HPLC analysis of the *DaCGO1* reaction products showed a progressive evolution of two distinct product peaks, a feature of the CGO reaction observed also for other GMC-type CGOs (Nakamura et al., 2019; Mori et al., 2021).

The possibility that the two peaks would correspond to a C2"- and a C3"-oxo product is incompatible with the results from substrate docking in *DaCGO1*, even if the active-site loop is omitted. Nakamura and co-workers

analyzed the identities of the two peaks by performing NMR on chemically synthesized 3''-oxo puerarin (Nakamura et al., 2019). Their results showed that the C3''-product undergoes non-enzymatic tautomerization to generate the 2''-oxo product.

Whether the non-enzymatic tautomerization has any effect on subsequent C-C bond cleavage is not known. The two peaks disappear progressively as the *DaCGD* reaction proceeds, suggesting two scenarios: either *DaCGD* is selective only for the 3''-oxo product, and the tautomerization is fast enough to re-equilibrate the 3''-oxo product; or *DaCGD* is not selective towards one specific product, but both 3''- and 2''-oxo products are deglycosylated.

To better evaluate these possible scenarios, some considerations need to be made. The manual docking of oxo-products in *DaCGD* was aided by electron density for to firmly bound water molecules and metal ions, which provided a probable hypothesis for how to orient the glucosyl C2'' relative the catalytic histidine to support C-C bond cleavage of *C*-glucosyl compounds oxidized at the C3'' position.

Furthermore, the *in vitro* one-pot reaction does not provide the exact same condition of the *in vivo* reaction, and it is not known how the interactions between CGO and CGD, environmental conditions, or the presence of other proteins may affect the tautomerization of the 3''-oxo to 2''-oxo product. Further studies are needed to investigate this phenomenon and to determine in more detail the substrate specificity and selectivity of *DaCGD*.

Our findings that *D. aerius* can deglycosylate *C*-glycosyl aromatic polyketides such as isoorientin, isovitexin and mangiferin provide insights into how this bacterium may interact with plants. The presence of *Deinococcus* species in extreme soil environments and in the rhizosphere of certain plants suggests that also *Deinococcus aerius* could act as a plant symbiont.

3.3 Paper II. C-glycosyl oxidoreductase in *S. canus*

3.3.1 *Streptomyces canus*

Streptomyces canus is a Gram-positive soil bacterium belonging to the Actinobacteria class. Actinobacteria are the most abundant class of symbiotic soil bacteria and have been extensively studied for their production of antibacterial and antifungal natural compounds (Baltz, 2008). In the case of *S. canus*, various strains of this bacterium have been investigated for their ability to produce antifungal metabolites like resistomycin and tetracenomycin D (Zhang et al., 2013).

Especially *S. canus* strain CA-091830 has been found to produce antibacterial potentiators against methicillin-resistant *Staphylococcus aureus*, such as the natural cyclic peptides krisynomycins B and C (Perez-Bonilla et al., 2020), which enhance the effect of imipenem in a synergistic manner. Notably, *S. canus* strains not only demonstrate the production of antibacterial and antifungal metabolites but also exhibit the ability to degrade neonicotinoid insecticide acetamiprid that pollute soil and water (Guo et al., 2019). To expand the knowledge about possible benefits of *S. canus*, it is particularly interesting to address mechanisms of how *S. canus* interacts with plants, and for the scope of this thesis, in the context of C-glycosyl aromatic polyketides (Chapter 1.2).

Phylogenetic analyses of sequences for putative pyranose oxidases (POxs) sequences revealed the presence of a possible POx in *S. canus* (ScPOx). ScPOx was produced and characterized biochemically and found to be an FAD-dependent GMC oxidoreductase belonging to CAZY subfamily AA3_4 (Chapter 1.3.1). This subfamily includes fungal POx members that have been extensively studied for their biotechnological applications. It was only recently that two bacterial AA3_4 POx members were biochemically characterized, namely *Arthrobacter siccitolerans* (AsPOx; Mendes et al., 2016) and *Kitasatospora aureofaciens* (KaPOx, previously known as *Streptomyces aureofaciens*) (Herzog et al., 2017).

AsPOx was found to be closely related to the recently characterized CGO MtCarA, which suggests that pyranose oxidases and CGOs are close homologs (Kumano et al., 2021). Results from the biochemical characterization of ScPOx shed light on its catalytic activity and relationship with other bacterial POx and CGO members and places the

52 PRESENT INVESTIGATION

enzyme in the larger context of symbiotic plant-bacteria interactions mediated by glycosylated plant aromatic polyketides.

3.3.2 Biochemical characterization of ScPOx

The recombinantly expressed and purified ScPOx appeared as a monomeric protein with a molecular weight of 52 kDa. The monomeric state of ScPOx is shared with other bacterial monomeric oxidases like AsPOx, KaPOx and the CGOs from *Microbacterium sp. 5-2b* (CarA), *M. trichothecenolyticum* (MtCarA) (Kumano et al., 2021), and DaCGO1 (Paper I), and distinguishes it from the homotetrameric fungal P2Os (Hallberg et al., 2004).

As shown by TCA precipitation and MALDI-TOF experiments the FAD cofactor is non-covalently bound to ScPOx, which is another difference from fungal P2Os that have covalently bound FAD. The UV-Vis spectrum of ScPOx showed the characteristic FAD-absorption bands at 390 and 450 nm, which disappeared upon reduction with 4 M D-xylose. Furthermore, the blueshift of the flavin peak upon oxidation of the TCA-treated ScPOx further indicated the presence of non-covalently bound FAD.

When comparing the sequence of ScPOx with those of fungal ToP2O and bacterial KaPOx, it became evident that ScPOx lacks the XTHW sequence motif required for covalent flavinylation. As for ScPOx, the enzymes AsPOx, CarA, and MtCarA also lack the flavinylation motif and have non-covalently bound FAD cofactors.

3.3.3 Catalytic activity of ScPOx

To characterize the catalytic activity of ScPOx, different electron donors and acceptors have been tested, including aldopyranoses that serve as electron-donor substrates for fungal POx members where the site of oxidative attack on the sugar differs depending on the specificity of the enzyme. ScPOx demonstrated activity for D-glucose, D-galactose, D-xylose, D-ribose, and L-arabinose, but performed considerably worse on these substrates compared with KaPOx or fungal POxs (Herzog et al., 2017; Abrera et al., 2020). The Michaelis Menten constants (K_m) for the substrates were in the millimolar range, which strongly suggested that they are not natural substrates for ScPOx.

Next, glycosylated aromatic polyketide were tested as substrates, *e.g.*, the C-glycosyl compounds carminic acid, mangiferin and puerarin, and the O-glycosides naringenin and rutin. Of these, puerarin performed best, and some activity was observed also for carminic acid and mangiferin, while no activity was detected with the O-glycosides. The catalytic efficiency for puerarin was considerably higher than for the tested monosaccharides, suggesting that ScPOx may in fact be a CGO (**Figure 21A**).

To put the results from activity studies in a structural context, a theoretical model of ScPOx was generated and compared to the experimental structures of *ToP2O* and *MtCarA*. ScPOx showed the typical GMC fold with an FAD-binding domain and a substrate-binding domain (**Figure 21B**). As for *DaCGO1*, ScPOx has a more compact lid-and-loop structure compared to oligomeric GMC members. The “head” domain of unknown function in *ToP2O* is not present in ScPOx. The N-terminal domain in *ToP2O* involved in oligomerization is also absent in the bacterial members, which is consistent with GMC-type bacterial pyranose oxidases and CGOs being monomers (**Figure 21C**).

The active site in ScPOx includes the conserved catalytic His-Asn pair (His438-Asn482). The largest difference between the predicted model of ScPOx and the experimental structures of *ToP2O* and *DaCGO1* is seen for active site loop. This loop in *ToP2O* is highly dynamic and has been observed experimentally to adopt several conformations where a fully closed and an open conformation dominate the conformational ensemble depending on the enzyme state. The sequence for this loop in ScPOx is different and due to the uncertainty of the predicted model, it is difficult to make any assumptions about the conformation except that it is likely to be different compared to *ToP2O*, *DaCGO1* and *MtCarA*.

54 PRESENT INVESTIGATION

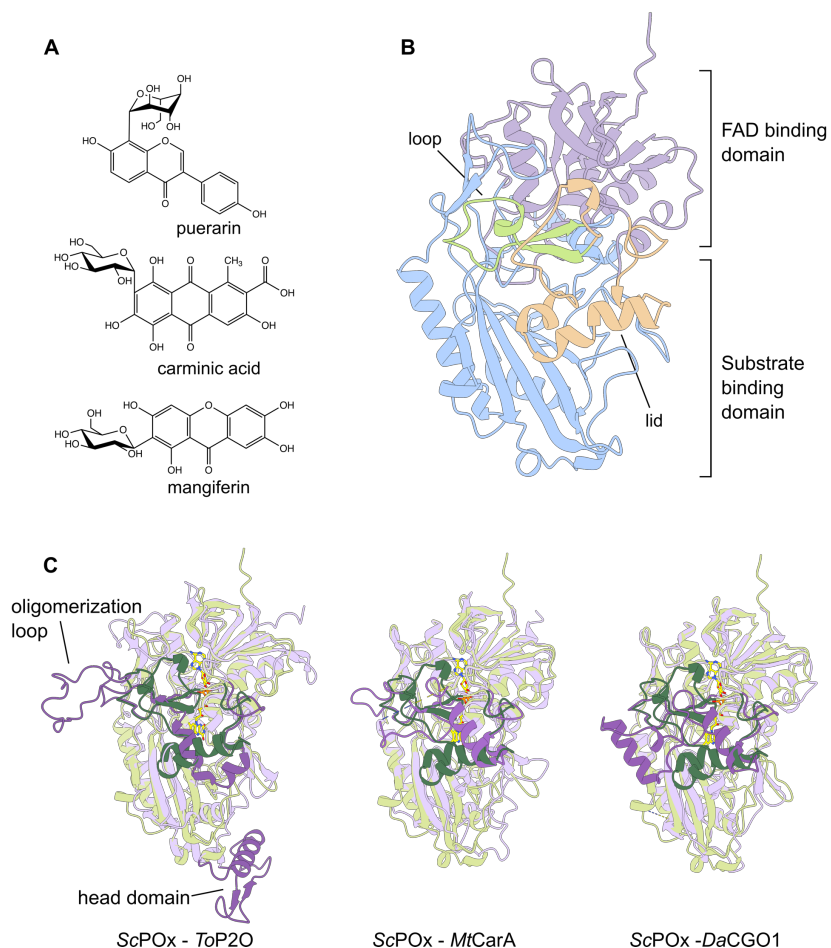


Figure 21: (A) ScPOx showed activity towards the C-glycosyl plant compounds puerarin, carminic acid and mangiferin, with higher specific activity for puerarin. (B) The predicted structure of ScPOx folds as a GMC oxidoreductase, with a FAD domain, a substrate binding domain, and a compact lid-and-loop domain (AlphaFold2 model (Jumper et al., 2021) C) Superimposition of the predicted model of ScPOx with the experimental structures of CGOs discussed in Chapter 1.3.1 ToP2O (PDB 1TT0), MtCarA (PDB 7DVE) and with the experimental structure of DaCGO1 (PDB 8QVE; Paper I).

3.4 Paper III. C-glycosyl deglycosylation in *M. testaceum*

3.4.1 *Microbacterium testaceum*

The genus *Microbacterium* belongs to the class of Actinobacteria, and various studies have highlighted this genus as a promising microbial inoculant in agriculture based on its biostimulant activity on crops, antibiosis against pathogens, and production of volatile compounds that are beneficial for plants (Barnett et al., 2006; Cordovez et al., 2018).

Microbacterium testaceum is a Gram-positive endophytic soil bacterium known to elicit plant defense mechanisms leading to suppression of blast disease in the rice cultivar Pusa Basmati-1, and to suppress fungal infections (Sahu et al., 2021; Patel et al., 2022). *M. testaceum* shows typical characteristics of beneficial endophytic plant bacteria and offer several mechanisms that promote plant growth, including phosphate-solubilizing activity (Fretes et al., 2018).

M. testaceum has been isolated from potato leaves shown to degrade N-acyl-L-homoserine lactone (AHL) compounds (Morohoshi et al., 2009). AHL compounds are used by pathogenic Gram-negative species as virulence regulators, and co-inoculation of potato slices with *M. testaceum* and *Pectobacterium carotovorum* subsp. *carotovorum* attenuated the soft rot symptoms caused by the pathogenic bacteria (Morohoshi et al., 2009).

Our genome analysis of *M. testaceum* strain NS220 revealed the presence of a *cgd* gene that was not present in a *cgd* gene cluster, which unlike traditional *cgd* gene clusters, encoded the α - and β -subunits in the same polypeptide chain. While the precise classification of this strain is still unclear (Midha et al., 2016), the sequence of the studied *cgd* gene is conserved in other *M. testaceum* strains. A phylogenetic analysis of CGD sequences for symbiotic and pathogenic soil bacteria was performed to shed light on the distribution of CGDs among bacteria. Moreover, the experimental structural determination of the fused *M. testaceum* CGD, *MteCGD*, and a comparison with previously characterized CGDs from soil and intestinal bacteria revealed new features and further insights into the mechanism of C-glycosyl deglycosylation of aromatic plant polyketides.

3.4.2 Biochemical characterization of *MteCGD*

The putative *cgd* gene NS220_14145 identified in the genome of *M. testaceum* NS220 was cloned and expressed, and the gene product

recovered and purified. SDS-PAGE and MALDI-TOF confirmed a monomer mass of 50.2 kDa and SEC analysis showed that the enzyme is a homodimer of 100 kDa. The GMC oxidoreductase *DaCGO1* (see Paper I and Chapter 3.2) was used as CGO to produce oxo substrates for *MteCGD* reactions and investigate the deglycosylation activity. One-pot reactions of *DaCGO1* and *MteCGD* were performed with isoorientin, isovitexin, and mangiferin as model substrates. The results showed that *MteCGD* can deglycosylate all three 3''-oxo compounds by cleaving the C-C bond between aglycone and sugar moiety (**Figure 22**). While *DaCGO1* prefers isovitexin over isoorientin (and *DaCGD* mirrors this preference), *MteCGD* performed best on 3''-oxo isoorientin, slightly slower on 3''-oxo isovitexin, and only very slowly on 3''-oxo mangiferin. Both products of the reactions, 1,5-anhydro-D-erythro-hex-1-en-3-ulose and aglycone, could be identified using reverse-phase LC-electrospray ionization (LC-ESI)-MS/MS analysis.

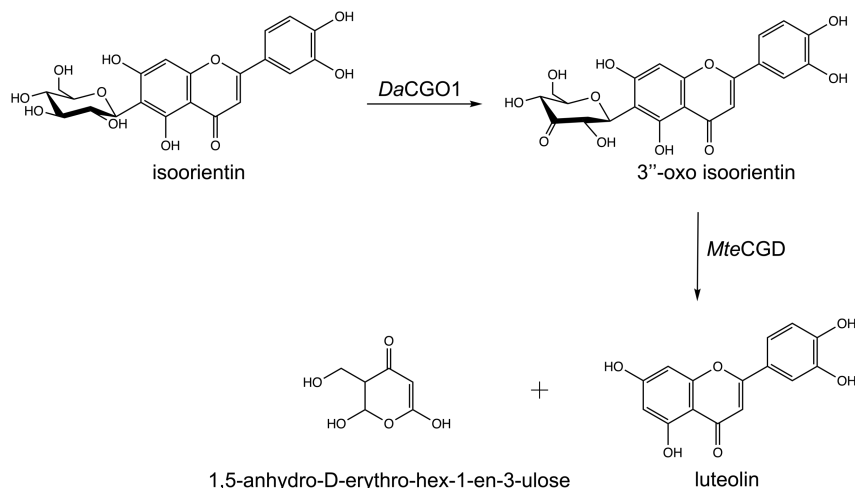


Figure 22: Products from CGO/CGD reactions. Oxidation of isoorientin by *DaCGO1* yields 3''-oxo isoorientin, which is cleaved by *MteCGD* to produce aglycone (luteolin) and 1,5-anhydro-D-erythro-hex-1-en-3-ulose. The products support the mechanism proposed earlier for C-C-bond cleavage by CGDs (Mori et al. 2021, Chapter 3.2.4)

3.4.3 Structure determination of *MteCGD*

The structure of *MteCGD* was determined by single-crystal X-ray crystallography (2.0 Å) and single-particle cryo-EM (2.3 Å). The homodimeric structure (**Figure 23A**) is essentially identical in the two experimental models with an r.m.s.d. of 0.53 Å for 866 aligned C α atoms.

The homodimer was confirmed independently by several methods: SEC, X-ray crystallography, cryo-EM and interface analysis using PDBePISA that returned a complexation significance score (CSS) of 1.0 (<https://www.ebi.ac.uk/pdbe/pisa/>; Krissinel et al., 2007).

The *MteCGD* monomer consists of two domains, an N-terminal β -domain (residues 1-136) and a C-terminal α -domain (residues 137-433). The α - and β -domains correspond to the separate α - and β -subunits in previously characterized CGDs (Chapter 1.3.2 and 3.2.3; Mori et al., 2021) (**Figure 23B**). As observed for other CGDs, the α -domain has a xylose isomerase-like TIM-barrel fold, while the β -domain contains a DUF6379-type β -sandwich composed of a 5-stranded and a 4-stranded antiparallel β -sheet. Unlike most of the previously characterized CGDs (Chapter 1.3.2 and 3.2.3; Mori et al., 2021), *MteCGD* does not have a lid domain that forms a 4-helix bundle, but instead has an extensive loop like that found in *EuCGD* (Mori et al., 2021).

The α - and β -domains in each chain are connected by a short linker and their relative orientation in the homodimeric assembly resembles a “yin-and-yang” arrangement. The active site with the catalytic metal center is located at the C-terminal end of the parallel β -strands as expected for a TIM barrel. The substrate-binding pocket is covered by the lid loop and β_{4a}/α_{4a} loop in the α -domain of one monomer and the β_{6b}/β_{7b} loop in the β -domain of the opposite monomer (**Figure 23C**).

A comparison of *MteCGD* and the structure of *EuCGD* (Mori et al., 2021) reveals that the yin-and-yang homodimer of *MteCGD* corresponds to one half of the *EuCGD* heterooctamer, but with a different angle between of the α - and β -domains compared with the α - and β - subunits in *EuCGD* (**Figure 24**).

The shielding of the active site is different in *MteCGD* compared with *EuCGD*, which is mainly due to the eight-residue longer lid loop in *MteCGD*. The large helical lid domains in *PuCGD*, *DaCGD* and *AgCGD2* contribute extensive shielding of their active sites compared with the lid loop in *MteCGD* and *EuCGD*. The β_{6b}/β_{7b} loop has the same length in *MteCGD* and *EuCGD* but is considerably shorter in *PuCGD*, *DaCGD* and *AgCGD2*.

The presence of a helical lid *versus* a loop lid is intriguing. *MteCGD* and *EuCGD* catalyze deglycosylation of the model substrates equally well as do *PuCGD*, *DaCGD* and *AgCGD2*. It is tempting to speculate that the helical

lid domain might have a function other than to just bind or shield the substrate.

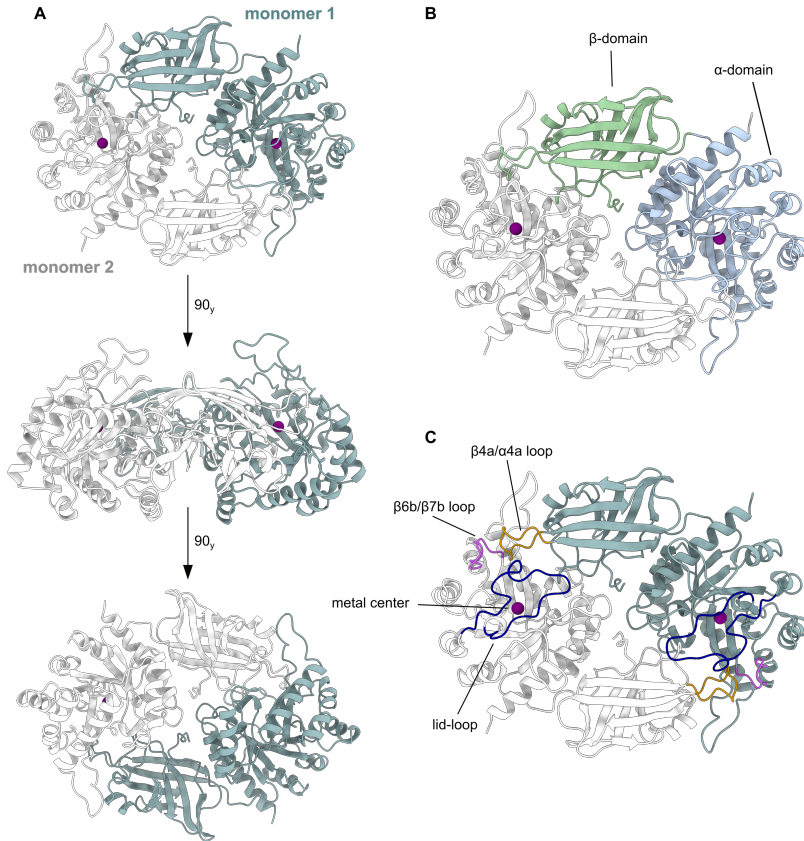


Figure 23: Structure of *MteCGD*. (A) Overall structure of the homodimer shown in different orientations (X-ray structure; PDB QRZ). (B) The α - and β - domains in the *MteCGD* monomer highlighted. (C) The two active sites are located at the interface of the monomers. The structural elements that gate the active site are highlighted.

As mentioned in Chapter 1.3.2, the helical lid domain in the cryo-EM structure of *PuCGD* (PDB 7DRD) was not visible in the map, and a *PuCGD* deletion mutant lacking the lid could still be expressed, folded, and assembled into an $\alpha_4\beta_4$ heterooctamer, but was inactive. This makes the role of the lid domain even more enigmatic since its sequence, location and

orientation differ between the enzymes; for instance, in *DaCGD*, the helical lids are buried at the interface between two $\alpha\beta$ heterodimers whereas in other CGD the lid is oriented outwards from the core and exposed to the surrounding environment. There are also few common structural denominators between the loop lids and the helical lids that may suggest a common function. A comprehensive mutagenesis study is required to further investigate the significance of a loop *versus* a helical bundle.

The metal coordinating residues in *MteCGD* are Glu292, Asp324, Glu405, and Gln369, and two water molecules, one of which is interacting closely with Lys371, like the observed metal coordination in *DaCGD* (Chapter 3.2.3). The metal coordinating residue Gln369 corresponds to a histidine in *PuCGD*, *EuCGD*, *AgCGD2* and *DaCGD*. His294 and Glu407 are the catalytically important residues responsible for the C-C cleavage mechanism (Mori et al., 2021).

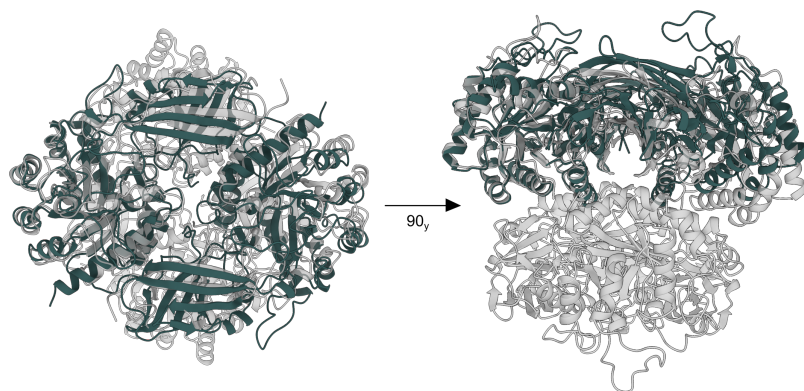


Figure 24: Comparison of the oligomeric state in *MteCGD* and *EuCGD*. (A) Overlay of the crystal structures of *MteCGD* (dark green, PDB 8QRZ) and *EuCGD* (transparent grey ribbons; PDB 7EXB).

3.4.4 Catalytic activity of *MtCGD*

The theoretical docking of the substrate 3"-oxo-isovitexin was guided by the structure of the isoorientin-bound *AgCGD2* (PDB 7DNN, Mori et al., 2021). Given the open conformation of *MteCGD*, the ligand could be positioned approximately with the same orientation as isoorientin in the *AgCGD2* structure.

MteCGD presents the same reactivity of *EuCGD* as described by Mori and co-workers (Chapter 1.3.2; Mori et al., 2021), and the analysis of the residues interacting with the substrate in the theoretical docking revealed similarities of the binding sites of the two CGDs here discussed (**Figure 25**). In particular, the position of the active histidine (His294 in *MteCGD* and His157A in *EuCGD*) is retained, and the Lys257 coordinating the sugar in position O4'' in the *MteCGD* α -domain is conserved in *EuCGD* (Lys120A). On the other side, the lid loop is considerably longer in *MteCGD* than in *EuCGD* and it is difficult to predict how differently the aglycone moiety is interacting with this structural element. In *MteCGD* Lys336 and Arg344 can establish bonds with the hydroxyl groups in the aglycone B ring, and these residues are not conserved in the *EuCGD* loop.

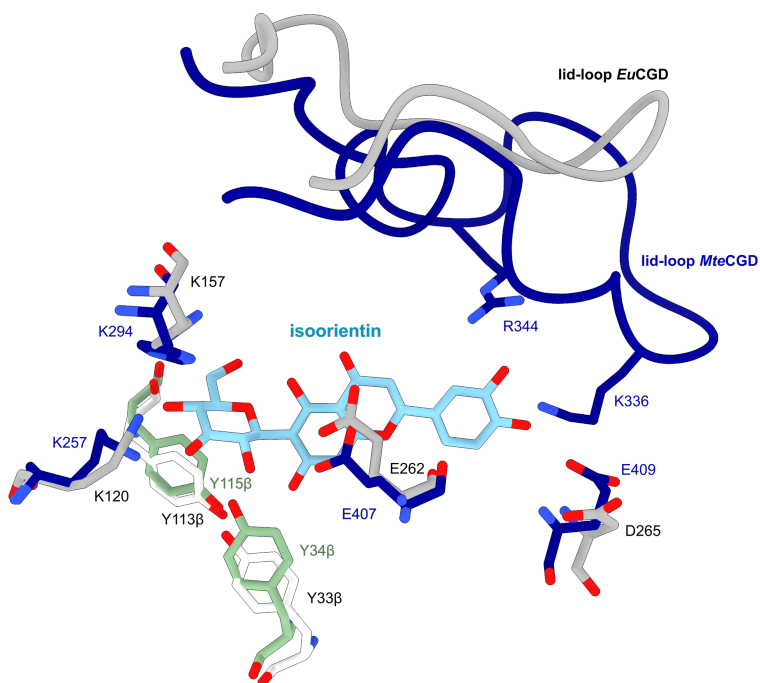


Figure 25: Docking of 3''-oxo isoorientin in the active site of *MteGCD* (PDB 8QRZ) and overlay with *EuCGD* (PDB 7EXB). The *MteCGD* α -domain residues are shown in blue, and the β -domain residues in green. The *EuCGD* α -subunit residues are shown in grey, and the β -domain residues in white. The lid loops are shown as ribbons.

3.4.5 Phylogenetic analysis of CGDs in Prokaryota

The phylogenetic analysis of CGD enzymes provided insights on the distribution of these enzymes in bacteria. The sequence of the *DaCGD* β -subunit were used in BLAST searches to find related sequences. CGD sequences could not be found in eukaryotes but were widespread in prokaryotes, especially the bacterial phyla: *Actinomycetota* (previous *Actinobacteria*; G+) to which *M. testaceum* and *S. canus* belong; *Pseudomonadota* (prev. *Proteobacteria*; G-); and *Bacillota* (prev. *Firmicutes*; G+). Representatives from the phyla *Bacteroidota* (prev. *Bactroidetes*; G-); *Chloroflexota* (prev. *Chloroflexi*; G+/G-) and *Deinococcota* (prev. *Deinococcus-Thermus*; G+/G-) were also included. The new names of phyla (Oren et al., 2021) and Gram status are given in parenthesis.

These results provide several new insights. First, CGD $\alpha\beta$ -fusions (such as *MteCGD*) occur more frequently in *Actinobacteria*. Second, CGD enzymes appear to be more wide-spread among bacterial phyla first anticipated.

Third, CGD enzymes are not present in well-studied bacterial plant pathogens such as *Pseudomonas syringae*, *Ralstonia solanacearum*, *Agrobacterium tumefaciens*, *Xanthomonas oryzae*, *Xanthomonas campestris*, *Xanthomonas axonopodis*, *Erwinia amylovora*, *Xylella fastidiosa*, *Dickeya dadantii*, *Dickeya solani*, *Pectobacterium carotovorum*, and *Pectobacterium atrosepticum* (Mansfield et al., 2012).

Fourth, CGD homologs are common in bacteria that are beneficial for plants, and other organisms. Some of the bacteria that have CGD enzymes are PGPRs such as *Rhizobium* species, which are known for their beneficial effects on plant growth and productivity (Alemneh et al., 2020; Palai et al., 2021; Matse et al., 2020; Hayat et al., 2010). *Novosphingobium* species find applications in rhizoremediation (Lee et al., 2023). *Pseudomonas* species are known for their role in phytohormone synthesis and nutrient solubilization (Costa-Gutierrez et al., 2022). *Deinococcus*, *Streptomyces* and *Microbacterium* have shown beneficial effects on plants and have been discussed in Chapter 3.2, 3.3 and 3.4.

To conclude, our results strongly indicate that CGDs are widespread in beneficial bacteria and absent from pathogens, at least those that are confirmed and well-studied plant pathogens. We found one *Erwinia* species, *Erwinia* sp. JUb26, that at first seemed to challenge this

hypothesis. The genus *Erwinia* of the order *Enterobacterales* is known to include many species that are plant pathogens. However, further examination of the literature revealed that *E. sp. JUb26* is a non-virulent gut endosymbiont of the soil nematode *Caenorhabditis elegans* (Samuel, 2017).

3.5 Papers IV and V. Lipoprotein localization in *Xcc*

3.5.1 *Xanthomonas campestris* pv. *campestris*

Xanthomonas campestris pv. *campestris* (*Xcc*) is a Gram-negative seed-borne bacterium that causes black rot disease in *Brassica* crops. The disease is spread worldwide since the 1970s when in the United States approximately 70% of the million transplants were systematically infected from one single seedbed (Hakalová et al., 2022). Recurrent black rot epidemics were reported in the 1990s in Italy, India, Japan, and Russia, with losses between 23 and 90% in susceptible *Brassica* cultivars (CABI, 2022), as well as later in Brazil (Hakalová et al., 2022) and in the Republic of South Africa (Mandiriza-Mukwirimba et al., 2016) and New Zealand (Braithwaite et al., 2008). This makes *Xcc* the most important cruciferous pathogen worldwide (Vicente et al., 2013; Cruz et al., 2017), remaining the most serious bacterial problem of agricultural production worldwide.

This pathogen can be seed-borne or enter the leaves through the hydathodes or wounds in the plant, manifesting as typical V-shaped lesion. The bacteria move *via* the plant's vascular system affecting growth and causing plant death (Lu et al., 2021). Despite the efforts to produce and use *Xcc*-free seeds, the pathogen is still spreading. The current methods to manage *Xcc* are mainly physical, *e.g.*, removing *Xcc* from the seeds' surface, hot water treatment or dry heat (Hakalová et al., 2022), however, these techniques cannot eliminate deeper infections in the seed.

Studies on the virulence of *Xcc* in the model plant *Arabidopsis thaliana* revealed that *Xcc* effectors cause protein modifications in the plant's NB-LRR, specifically the effector AvrAC/XopAC uridylylates and inhibits several plant PRR complexes (Chapter 1.4; Xu et al., 2008). Multiple other virulence factors also need to be considered. Since *Xcc* is a Gram-negative bacterium, its virulence is also correlated to the production of protective biofilm, structural integrity, and transport of nutrient, together with the production and localization of outer membrane lipoproteins (Chapter 1.5).

The studies of Liao and co-workers showed that LolA and LolB were important for *Xcc* virulence (Liao et al., 2019; 2022). Possible virulence factors have been identified for several species belonging to the Xanthomonadaceae family, including *Xcc* (Sidhu et al., 2008), *Xanthomonas citri* subsp. *citri* (Ferreira et al., 2016; Yamazaki et al., 2008; Zimaro et al., 2013), *Xanthomonas oryzae* (Wang et al., 2013;

González et al., 2012), and *Xylella fastidiosa* (Silva et al., 2011). It is not known whether *Xcc* lipoprotein peptides play a role in PTI signaling and what PRRs would be engaged in that case.

3.5.2 LolA-LolB complex formation and interactions

In Paper IV, *Xcc* LolA and a periplasmic variant of *Xcc* LolB were recombinantly produced and purified, and biochemical, biophysical, and structural characterization were performed. SEC analysis of a 1:1 molar mixture of LolA and LolB resulted in a single monodisperse species corresponding to a LolA-LolB heterodimer. This shows that LolA and LolB can form a stable, non-obligate heterodimer in absence of the lipoprotein.

ITC studies were performed on *Xcc* LolA and LolB to determine the thermodynamic profile of their interaction in the absence of lipoprotein anchor. The interaction was characterized by a 1:1 binding model and a K_D value of $\sim 15 \mu\text{M}$, consistent with the micromolar range K_D value determined for the *E. coli* LolA-LolB interaction (Kaplan et al., 2022) and for *Vibrio cholerae* LolA-LolB (Jaiman et al., 2023). In contrast to the exothermic interaction reported for *E. coli* and *V. cholerae* LolA-LolB, the titration curve showed that the interaction between *Xcc* LolA and LolB is endothermic.

To further investigate the *Xcc* LolA-LolB interaction, the crystal structure of *Xcc* LolA-LolB was determined and analyzed in Paper IV. The structure revealed two LolA-LolB heterodimers in the asymmetric unit. Importantly, the residues stabilizing the heterodimeric interaction do not coincide with regions expected to be involved in binding of the lipid anchor (**Figure 26A**).

Xcc LolA has a β -barrel-like fold with an 11-stranded antiparallel β -sheet (β_{1A} - β_{11A}) and a C-terminal β -strand (β_{12A}) parallel with β_{11A} . Of the four α -helices (α_{1A} - α_{4A}), α_{2A} - α_{4A} form the lid domain closing over the hydrophobic cavity where a lipid anchor would bind.

The residue Arg43 in *E. coli* LolA has been assigned important roles in affecting the conformation of the lid, binding of lipid anchor, and in the transfer of the lipid anchor from LolA to LolB (Kaplan et al., 2022). Despite its predicted importance, this residue is not conserved in *Xcc* LolA and instead Thr43 occupies this position, and the threonine does not form interactions controlling the helical lid. Despite the low sequence similarity between *Xcc* LolA to LolB, they share a similar fold, the latter having an

antiparallel β -barrel-like fold defined by 11 β -strands (β_1B - $\beta_{11}B$) and three α -helices α_1B - α_3B , where α_2B - α_3B pack as a lid over the hydrophobic cavity (**Figure 26A**). The lids in LolA and LolB are stabilized by several hydrogen bonds and salt bridges in what appears to be closed conformations, but with the hydrophobic cavities shielded differently from the solvent.

In both heterodimers of the asymmetric unit, LolA and LolB interact in a “mouth-to-mouth” orientation that connects the two hydrophobic cavities to form a continuous hydrophobic channel, probably required to ensure successful lipid transfer (**Figure 26B**). The LolB hook (residues 74-79) docks on the concave side of the LolA cavity but shows slightly different conformations in the two heterodimers. The different hook conformations induce small but distinct differences in the LolA-LolB interaction, and thereby affecting the precise association mode between the two proteins. This probably serve to modulate the stability of the heterodimeric state. Although the interactions between the two heterodimers appear to be maintained, the different hook conformations cause a tighter association between LolA and LolB in heterodimer 1 compared to heterodimer 2. The LolB hook is located at the bottom of an annulus at the surface between LolA and LolB, which is probably where the tether of the lipoprotein would protrude (**Figure 26B**).

The key interactions between LolA and LolB are located on the external part of the LolB β -barrel (β_1B , β_2B , $\beta_{10}B$ and $\beta_{11}B$) that forms salt bridges with LolA residues located in the β_{5A}/β_{6A} loop (**Figure 26A**). Additionally, the C-terminal strand β_{12A} in LolA constitutes a key spot for heterodimer interaction, forming several parallel β -type backbone hydrogen bonds with LolB β_1B . The relevance of the LolA C-terminal strand in the LolA-LolB interaction has been highlighted also by mutagenesis studies for studies of the *E. coli* LolA-LolB complex, where the β_{12A} strand was found to be essential for the release and transfer of LPs (Okuda et al., 2008). The hydrogen bonds and salt links that stabilize the Xcc LolA-LolB heterodimer are further accompanied by many hydrophobic interactions that add further specificity and stability.

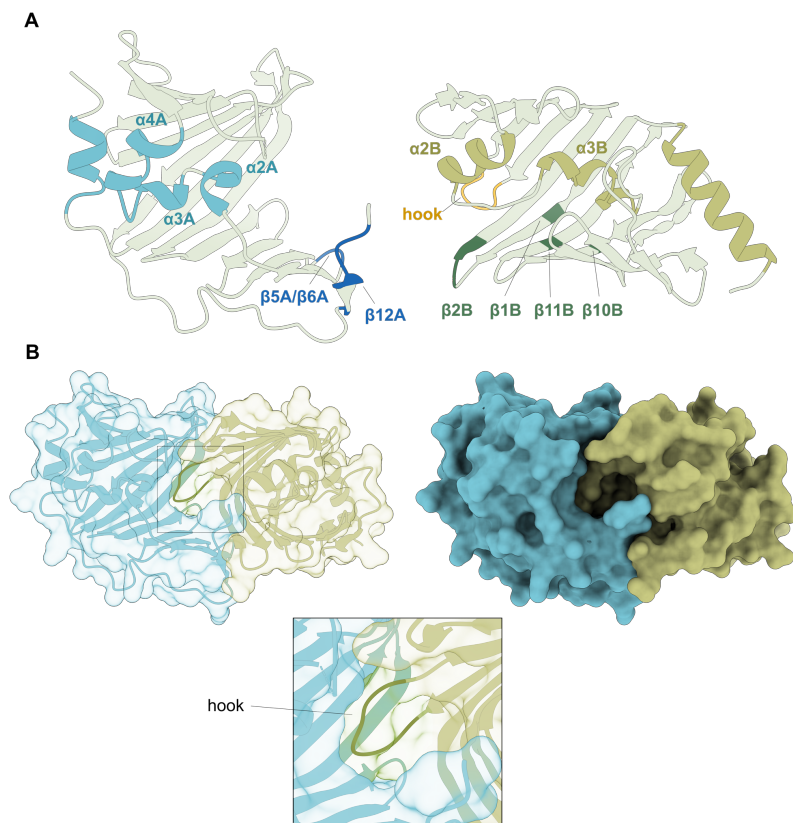


Figure 26: Experimental structure of the *Xcc* LolA-LolB complex (PDB 8ORN). (A) *Xcc* LolA (left) with highlighted α -helices (light blue) and structural elements interacting with LolB (dark blue). *Xcc* LolB (right) with highlighted α -helices (light green) and structural elements interacting with LolA (dark green). The LolB hook is here indicated in yellow. (B) Ribbon representation of the heterodimeric structure of the *Xcc* LolA-LolB (LolA represented in blue and LolB in green) with overlaid semitransparent surface (left). On the right the same structure but with solid surface represented to highlight the formation of the annulus at the interface of the two monomers. The inset shows the hook laying at the bottom of the annulus.

It was possible to model the lipid anchor in the hydrophobic cavities of LolA and LolB without disrupting the interactions securing the complex, which provided insights on a hypothetical transfer mechanism (**Figure 27A**). For modeling the lipid anchor in LolA to simulate a pre-transfer state, it was necessary to lift the LolA helical lid and move the $\beta 8A/\beta 9A$ loop. The conformation of the LolB hook needed to be slightly shifted

towards the inner β -sheet 2 in LolA, to insert the hook between the R2 acyl chain of the modeled lipid anchor, which supports the proposed importance of the hook for stabilizing the interaction between LolA and LolB. Upon docking of LolA with LolB, the LolB hook would chisel in between the inner LolA β -sheet and the lipid anchor to initiate lipid transfer and further conformational changes in LolB to enable lipid binding and release of LolA.

A lipid anchor was modeled in LolB to simulate the post-transfer state. As for LolA, it was required to lift the LolB helical lid to accommodate the lipid anchor. Opening of the LolB lid suggests that as transfer proceeds, the LolA-LolB interaction would be weakened as structural changes induced by the hook and lid movements propagate across the interface (**Figure 27A**). The LolB would basically “scoop out” the lipid anchor from LolA, which induces conformational changes that eventually leads to dissociation of the proteins (**Figure 27B**).

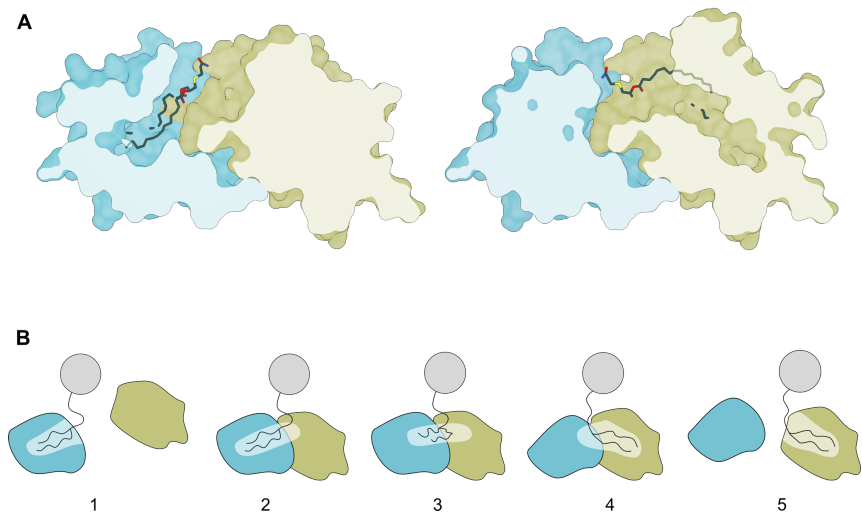


Figure 27: Docking of a diacyl lipid anchor to the Xcc LolA-LolB complex. (A) Lipid anchor modeled in LolA (left) and LolB (right). (B) Proposed steps for the transfer of a lipoprotein from LolA to LolB.

The canonical Lol model does not explain what happens after LolA has delivered the lipoprotein to LolB. Our results that LolA and LolB can form a stable lipoprotein-free heterodimer challenges some aspects of the model

proposed for *E. coli* and other Gram-negative bacteria. Once LolA has delivered the lipoprotein to LolB, the resulting LP-free LolA could be bound by LP-free LolB molecules in OM. Additionally, LP-free LolA may traverse the periplasm and associate with LP-free LolB. We cannot identify any mechanism that would prevent either of these events to take place. In both cases, LP-free LolA would be accumulating at the inner face of OM, especially during high LP expression, resulting in LP-free LolA-LolB complexes tethered to the inner face of the OM.

Thus, both LP-free forms of LolB and LolA-LolB complexes are expected to be present in the OM. It is likely that LP-bound LolA has a thermodynamic advantage over LP-free LolA in the interaction with LP-free LolB or LP-free LolA-LolB complexes and still be able to transfer LP to LolB. This hypothesis could be tested using LP-bound LolA and LolB provided that *Xcc* LolA can be loaded using the cognate ABC extractor charged with an LP carrying the correct acyl anchor.

If, however, LP-bound LolA fails to replace LP-free LolA in LolA-LolB complexes, this would strongly argue against the canonical Lol paradigm, and rather suggest that LolA is not recycled after lipoprotein delivery but captured by LolB to form LP-free “dead-end” LolA-LolB species. In that case we would expect that as lipoprotein transfer progresses, such species would accumulate in the OM and eventually lead to shut down of Lol-mediated LP transfer and degradation of dead-end complexes.

We performed additional ITC studies that were not included in Paper IV. These experiments aimed at investigating the effects of potential inhibitors that have been proposed for the *E. coli* proteins. In addition to *Xcc*, we also wanted to include another plant pathogen, *Erwinia amylovora* in this study. For this purposed the genes coding for *E. amylovora* LolA (EAMY_1336) and LolB (EAMY_1605) were cloned and expressed using the same methodology as for *Xcc*, and the proteins purified.

3.5.3 LolFD, a novel ABC transporter for lipoprotein extraction

As mentioned in Chapter 1.5, an ABC transporter is required for extracting the anchor of the lipoprotein from the inner membrane. In the canonical Lol system defined by *E. coli*, the ABC extractor LolCDE is built up by of two transmembrane subunits, LolC and LolE, that form a heterodimeric asymmetric membrane domain. On the cytoplasmic side, the membrane

domain is then attached to a LolD homodimer which provides the ATPase activity. However, it has been noted that some Gram-negative bacteria lack one or several genes coding for the five Lol proteins of the canonical system. One study reported that most *Gammaproteobacteria* would have three genes (*lolc*, *lold*, *lole*) for the ABC transporter, while *Alphaproteobacteria*, *Betaproteobacteria*, and some *Gammaproteobacteria* only have one gene (*lof*) for the transmembrane domain. The LolF sequence showed sequence similarities to both LolC and LolE and was hypothesized to form a LolF homodimer (LoVullo et al., 2015).

In paper V, we expanded the sequence analysis reported by LoVullo and coworkers to further investigate this hypothesis. Sequence analyses of LP extractors, and inspection of predicted structural models of the Xcc ABC extractor revealed interesting differences and similarities between the canonical LolCDE and hypothetical LolFD extractors.

In our bioinformatic analysis focused on the phylum *Pseudomonadota* (previous *Proteobacteria*), which is the taxonomic home to many pathogenic genera including *Xcc*. *Deltaproteobacteria* was earlier a class of *Pseudomonadota* but have been reclassified to form two new phyla, *Thermodesulfobacteriota* and *Deferribacteriota* (Waite et al., 2020). Since the previous study (LoVullo et al., 2015) had included *Deltaproteobacteria* as a class of *Pseudomonadota*, we also included these representatives.

Interestingly, the distribution of LolC, LolE and LolF proteins across *Pseudomonadota* showed that most members have operons with only one gene (*lof*) for the transmembrane domain (**Figure 28**). *Alphaproteobacteria* and *Betaproteobacteria*, and species of the phyla *Thermodesulfobacteriota* and *Deferribacteriota* were all found to have operons with only *lof* and *lold*.

In *Gammaproteobacteria*, *lof-lolD* operons were found most order: *Chromatiales*, *Oceanospirillales*, *Thiotrichales*, *Xanthomonadales*, *Pseudomonadales*, *Methylococcales*, *Legionellales*, *Nevskiales*, and *Moraxellales*. The “canonical” *lolC-lolD-lolE* operon was however found only in *Pasteurellales*, *Aeromonadales*, *Alteromonadales*, *Vibrionales* and *Enterobacterales*. These results show that most *Gammaproteobacteria* are likely to use a symmetrical LolFD-type ABC extractor whereas only a subset use the LolCDE type. Our hypothesis based on these results was that the “canonical” LolCDE extractor is the exception rather than the rule and that a LolFD-type extractor is the true canonical form.

The LolA sequences show a similar distribution as the LolFD- and LolCDE-type ABC-extractors where LolAs with cognate asymmetric LolCDE-type extractors cluster on a separate branch. The LolB sequences show a less well-defined relationship pattern, although also in this case LolBs associated with LolCDE are located on a separate branch. It has been noted that *Alphaproteobacteria* and *Deltaproteobacteria* (now phyla *Thermodesulfobacteriota* and *Deferribacterota*) do not have *lolB* genes (Okuda & Tokuda, 2011; Wilson et al., 2016; Smith et al., 2023), which agrees with our results.

Further analysis of the presence of all three genes *lgt*, *lsp* and *lnt* required to synthesize triacylated lipid anchors showed that the presence or absence of Lnt did not correlate with whether the system uses LolFD or LolCDE. In the case of the order *Xanthomonadales*, members can be roughly divided into two groups, one having only Lgt and Lsp (i.e., the genera *Xanthomonas*, *Xylella*, *Stenotrophomonas*, *Dyella* and *Frateriella*), and one group with Lgt, Lsp and a new, uncharacterized type of Lnt (i.e., *Ahniella*, *Aquimonas*, *Lysobacter* and *Thermomonas*).

To summarize, the sequence analyses suggest distinct relationships between the type of ABC extractor and the specific LolA and/or LolB. Regarding the different occurrence and distribution of Lgt-LsP-Lnt enzymes, it is likely that different types of LP acylation require different recognition mechanisms and, hence, different Lol systems.

The theoretical 3D model of the *Xcc* LolFD extractor allowed structural comparisons with experimental models of *E. coli* LolCDE, which helped to identify key differences between these transporters. The *E. coli* LolCDE (Yakushi et al., 2000) belongs to the type-VII ABC transporters (Thomas et al., 2020; Bilsing et al., 2023), and different cryo-EM structures of this complex have been determined so far (Tang et al., 2021; Sharma et al., 2021; Bei et al., 2022). The asymmetric transmembrane domain (TMD) of LolCDE, which is responsible for the extraction of the LPs from the IM (Okuda et al., 2009; Kaplan et al., 2018), is composed of eight transmembrane helices, four from LolC and four from LolE. On the cytoplasmic side, the TM domain interacts with two LolD ATPase subunits.

We used the experimental structures of LolCDE reconstituted in nanodiscs (Bei et al., 2022) to model the different conformational states of *Xcc* LolFD (**Figure 29**). These three templates were apo LolCDE (PDB 7V8M), LP-bound LolCDE (PDB 7V8L) and LP-free AMP-PNP-inhibited state (PDB 7V8I).

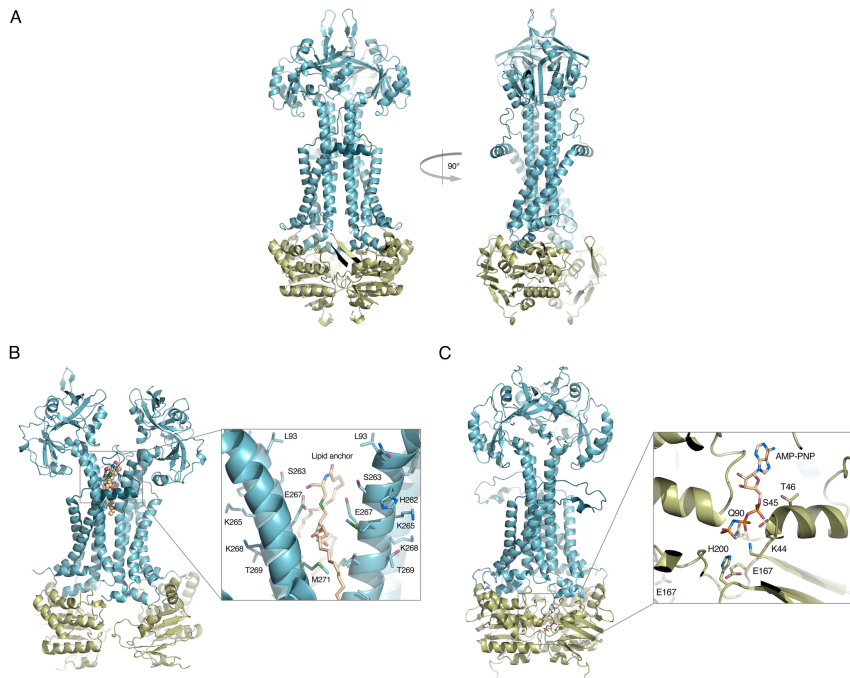


Figure 29: Overall structures of the predicted models of Xcc LolFD, predicted by AlphaFold2. (A) Model based on the apo state of the previously determined *E. coli* LolCDE complex (PDB 7V8M; Bei et al., 2022); (B) Model based on the lipoprotein-bound state of the previously determined *E. coli* LolCDE complex (PDB 7V8L; Bei et al., 2022); (C) Model based on the AMP-PNP-bound state of the previously determined *E. coli* LolCDE complex (PDB 7V8I; Bei et al., 2022).

It was straightforward to generate homology models of Xcc LolFD based on the three LolCDE templates. The LolF N-terminal α -helix interacts with the LolD cytosolic subunits, and following the N-terminal α -helix, TM1 extends towards the periplasmic side of the membrane ending in the LolF periplasmic domain (PD) (**Figure 29A**).

After the PD, TM2 forms a helical stalk with TM1, extending to the cytoplasmic coupling helix (CH). Together with the N-terminal α -helix, the CH interacts with the LolD subunit. TM3, following the CH helix, crosses the membrane and forms a “shoulder” at the membrane-periplasm interface. In LolCDE, the LolC and LolE shoulders have different lengths and define the “front pocket” and the “back pocket”, respectively, for the LPs binding (Sharma et al., 2021). In LolF, the shoulder is more similar to

the longer LolE shoulder, with a conserved U-loop that in LolE penetrates the V-shaped LP-binding cavity to translate the LP to LolA (Bei et al., 2022). The binding pockets in the symmetrical LolF homodimer are identical, and the lipoprotein can be presented to LolA by either of the two LolF subunits, probably depending on which PD domain LolA binds to. The shoulder loop is followed by TM4, which ends at the cytoplasmic side where the short C-terminal α -helical segment that interacts with LolD. Different considerations can be made regarding the implications of a homodimeric state of the LolFD as opposed to a heterodimeric state.

LolFD would, at least theoretically, be able to present two identical binding pockets for binding of LP. The residues that could interact with the R1, R2 and R3 acyl chains in LolC and LolE (Bei et al., 2022) are mostly conserved in LolF (being essentially aliphatic residues), as well as the residues binding the +2, +3 and +4 residues of the modeled lipoprotein RcsF from the structure 7V8L. Interestingly, the +1 cysteine binds in the negatively charged substrate binding pocket in LolCE, where it interacts with Glu263 in the LolC TM2 and with Asp264 in the LolE TM2, corresponding to Glu267 and Ser263 in Xcc LolF. Being The LolFD TMD homodimeric, the two Glu267 are positioned symmetrically and are both suitable for the triacyl cysteine binding (**Figure 29B**). Hence, both the U-loops of the LolF monomers can present the LP to LolA.

In LolCDE, only the LolC PD is responsible for the interaction with LolA (Okuda et al., 2009; Kaplan et al., 2018), while being the TMD of LolF a homodimer, it presents two identical PD domains. Structure and mutagenesis studies on the binding of LolA to LolCDE in *E. coli* identified key residues for the interaction with LolA in the LolC PD hook (Kaplan et al., 2018). Among these key residues Thr173, Ile178 and Arg182 are quite conserved in LolF (Thr175, Met180, Lys184). In LolC, the mutations of the hook F175A and R177A impair the LolA binding, and interestingly these residues are replaced by alanine in LolF (Ala174 and Ala179). On the other side, the Arg185 in the neighboring β -strand in the LolF hook might compensate the absence of the arginine in position 179.

When it comes to the cytoplasmic LolD domains in LolFD, the experimental AMP-PNP bound LolCDE structure (7V8I) has been compared with the corresponding theoretical Xcc LolFD model (**Figure 29C**). The ATPase activity of the LolC subunits induces conformational changes on the transporter via mechanotransmission (Crow et al., 2017). These conformational changes push the LP up the V-shapes binding pocket

to the PD hook where LolA is bound. The AMP-PNP bound features tight LolD-LolD association, tighter association of TMDs and compression of the LP-binding cavity, with displacement of the U-loop away from the cavity.

The *Xcc* LolD shows high sequence identity to the *E. coli* one (54.2%), presenting the typical Walker A motif for the phosphate-binding loop, where the conserved catalytic residues in the ATP-Mn²⁺ binding sites are located at the interface between the two subunits. Furthermore, the regions interacting with the TMDs are conserved in the *Xcc* and *E. coli* LolD homologs. While the ATPase activity and LP recognition and binding are predicted to be similar among *Xcc* and *E. coli* LP extractors, the implications of the homodimeric state of the TMD are difficult to predict, especially given that LolF presents features from both LolC and LolE.

These observations highlight the uniqueness of the Lol system in *Xcc* and underscore the need for further investigation to unravel the specific mechanisms governing lipoprotein transport in this plant pathogen. Understanding these mechanisms could offer valuable insights into the virulence strategies of *Xcc* and may have implications for developing targeted interventions against this pathogen.

4 Concluding remarks and outlook

The objective of this thesis was to investigate two distinct mechanisms that underlie the establishment of both symbiotic and pathogenic interactions between bacteria and plants.

Biochemical, functional, and structural characterization of the *cgo-cgd* gene cluster of *D. aerius* revealed the ability of this organism to deglycosylate C-glycosyl aromatic polyketides of plant origin. The GMC oxidoreductase *DaCGO1* combined with *DaCGD* could deglycosylate isovitexin, isoorientin, and mangiferin. The 3D structures of *DaCGO1* and *DaCGD* were determined and showed that *DaCGO1* is FAD-dependent GMC oxidoreductase and *DaCGD* has novel $\alpha_2\beta_2$ heterotetrametric assembly. The GMC oxidoreductase *ScPOx* from the soil bacterium *S. canus* was identified and biochemically and functionally characterized. This enzyme functions as a CGO by oxidizing C-glycosyl plant compounds, including puerarin, carminic acid, and mangiferin. This places *S. canus* in the context of symbiotic interaction relevant to plant hosts.

Additional results for the soil bacterium *M. testaceum* revealed a *cgd* gene with fused α - and β -subunits. *MteCGD* was able to cleave 3''-oxidized C-glycosyl compounds provided by *DaCGO1*, with the highest rate for 3''-isoorientin. The identification of products supports the previously proposed mechanism for C-deglycosylation by CGDs. The structure of *MteCGD* revealed yet another novel oligomeric state, a homodimer of two $\alpha\beta$ -fused chains. Results from phylogenetic analysis demonstrated the presence of CGD homologs in beneficial bacteria while lacking in pathogenic bacteria, which supports our hypothesis that deglycosylation of C-glycosylated aromatic polyketides has a role in establishing symbiotic interactions between plants and bacteria, and possibly to modulate soil suppressiveness and plant defense. Further biochemical, functional, bioinformatic, and plant-microbe interaction studies will shed further light on the role and importance of these enzyme systems. Such knowledge is likely to be useful in strategies to improve growth conditions for plants.

The second part of this thesis investigated the Lol pathway the pathogenic Gram-negative bacterium *X. campestris* pv. *campestris*. The work provided the first characterization of a physical complex between the two key proteins in lipoprotein localization, LolA and LolB. The structure corresponded to the lipoprotein-free LolA-LolB complex, and provided important information about how the enzymes interact, to ensure

successful lipoprotein transfer. Bioinformatic studies on the Lol lipoprotein extractor provided new information that challenges the established model of how lipoproteins are extracted from the inner membrane. These results underscore that a homodimeric lipoprotein extractor is the most common type among Lol systems in bacteria of the *Pseudomonadota* phylum. To prove the hypothesis of homodimeric lipoprotein extractors, *Xcc* LolFD was produced and a homodimeric oligomeric state confirmed.

These results shed new light on the Lol pathway of the plant pathogen *Xcc*, offering a more detailed picture of this system, which may also lead to new strategies to fight *Xcc*. Continued studies of interest would for instance be to model possible inhibitors using the experimental structures of *Xcc* LolA and LolB and the theoretical model of LolFD.

5 Acknowledgements

First of all, I would like to express my gratitude to my main supervisor, **Christina Divne**. From the very beginning, you believed in me, hired me, provided guidance, and equipped me with the tools to become a proficient scientist and researcher. You not only supported my strengths but also offered grounding when needed and motivation when required. You have been more than a supervisor; you have been a trustworthy friend, and I will forever be thankful for having met you.

I'd like to extend my thanks to my friends and colleagues in the Divne group. **Dayanand**, it has been an incredible (and productive) joy working alongside you. I appreciate the lengthy scientific discussions and the deep conversations after work. **Markus**, you are the most talkative Finn I know, and I'm grateful for the insightful discussions and the laughter we've shared. I also want to express my gratitude to **Tom**, former PhD student in the group, who guided me during the initial phase of my PhD studies. We shared travels and endured long days at the synchrotron, which, despite being tiring, were undeniably fun.

I'm deeply thankful to **Ines Ezcurra**, my co-supervisor, whose expertise in plant biotechnology was crucial in adding a three-dimensional aspect to this thesis (no pun intended). It has been a pleasure collaborating with you.

A special thanks goes to our collaborators at the 3D-EM facility at KI, particularly to **Martin Hällberg**, whose incredible competence in cryo-EM and his patience in teaching me data elaboration at unusual hours were essential for this thesis. Thanks to my academic sister, **Amy Bondy**, whom I met at the facility, and whose friendship is one of the most wonderful things that has happened since I moved to Sweden. Thank you for preparing me food during my challenging moments, for listening and providing support, and for being the best podcast co-host.

I would like to express my gratitude to the PSF facility at KI, especially to **Martin Moche** for his assistance with crystallization experiments and synchrotron data collection, and to **Thomas Nyman** for his support with the seemingly endless ITC experiments.

Thanks to our collaborators from BOKU, Vienna, including **Dietmar Haltrich**, **Anja Kostelac**, and **Jolanta Puc**. This collaboration led to an unexpected and exciting shift in our research topic. Thanks also to **Vaibhav Srivastava** for the fruitful collaboration.

Thanks to all my past and present co-workers at DIB. The Kick-off days, celebrations, after-work gatherings, and seminar sessions have been memorable. I'm grateful to the PhD students with whom I shared the PhD seminars for the thought-provoking discussions. A special shoutout to

Teun, who is almost as talkative as me and made organizing the lunch seminars easy and almost fun. **Johannes**, with whom I practiced my Swedish and had captivating scientific and non-scientific discussions. **Gustav**, whose kindness and support meant a lot, being always willing to help and discuss research. Thanks to **Kasra**, that in the last year brought his laughter and energy to the floor and in my life. Lastly, to all the DIB colleagues who shared Floor 2 with me, making our workplace a fun environment.

A heartfelt thanks to the past and present colleagues at the Glycoscience Department, especially to my cool gym and bouldering buddy **Kun**, to **Alma** that made me appreciate how warm and sunny Stockholm can be, the strong and inspiring **Amparo**, the kind and not-so-smiley (lol) **Susie**, and all the members of the *Language Lunch Club*. Lunchtimes were always lively and enjoyable with you all. To **Zsuzska**, my comfort human, the Hungarian sister I needed, the person I could talk to endlessly: thank you for your support, and I'll do my best to support you back, especially with my knowledge on how to survive the PhD journey.

To my friend and ex-floor 3 colleague **Caterina**, you brought a piece of our homeland to Stockholm, and I'm grateful for all the meaningful, amazing moments we've shared. You, along with **Serge** and **little Letizia**, have brought warmth to my heart. Thank you for everything.

Thanks to my chosen Swedish family, **José**, **Laia**, **Rodolfo**, **Sawsan**, and **Tamara**. All the fun activities, lazy evenings on the couch, Sunday family dinners, reciprocal therapy sessions, love, and support have meant the world to me. We've witnessed each other's transformations over the years and cherished every version of ourselves. I don't know how I would have survived this PhD journey without you.

Thanks to **Marica**, who always blindly believes in me, and who's talent brought to the creation of a dance concept that feels like home and kept me sane during these years; special thank you to the strong and caring **Olivia** and to all the dancing queens **Antea**, **Ari**, **Nele**, **Marian**, **Rebecka**, **Lilian**, **Nataja**, and **Vendela**. You are my safety net and my happy place. Dancing with you has enriched the quality of my life.

To **Marit**, who taught me mindfulness, self-compassion, and equipped me with the tools to navigate through adversities and cherish the moments of joy. I feel incredibly fortunate to have crossed paths with you.

Thanks to my Italian friends **Susanna**, **Paola**, **Aurora**, **Enrico**, and **Luca**. Our mandatory meetups when I return to Italy keep our friendship strong despite the distance.

Grazie alla mia famiglia. Ai miei genitori **Michela** e **Ivano**, so che sareste stati fieri di me qualunque cosa fossi diventata, che fosse un'artista di strada, una tatuatrice, una parrucchiera, una suora, un medico o una

scienziata. Alla fine sono diventata PhD (perché lavorare per davvero... no). Vi voglio bene, grazie per avermi cresciuta come avete fatto e per avere sempre sostenuto le mie decisioni, qualsiasi esse fossero. A mia sorella **Annamaria**, ormai cresciuta, mi spiace non essere stata a casa con te, ma spero essere riuscita a sostenerti anche da qui. Condividere momenti ed esperienze con te é una delle parti migliori delle mie vacanze. Mi manchi.

And finally, a special thanks to **Tintin**, my perfect ginger cat and the latest addition to my family. You've spent almost the entire time I wrote this thesis sleeping on my lap, and every time I look at you, my heart is filled with love.

6 References

- Abbrera, A. T., Sützl, L., & Haltrich, D. (2020). Pyranose oxidase: A versatile sugar oxidoreductase for bioelectrochemical applications. *Bioelectrochemistry*, 132, 107409. <https://doi.org/10.1016/j.bioelechem.2019.107409>
- Alemneh, A. A., Zhou, Y., Ryder, M. H., & Denton, M. D. (2020). Mechanisms in plant growth-promoting rhizobacteria that enhance legume–rhizobial symbioses. *Journal of Applied Microbiology*, 129(5), 1133–1156. <https://doi.org/10.1111/jam.14754>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Bag, S., Mondal, A., Majumder, A., Mondal, S. K., & Banik, A. (2022). Flavonoid mediated selective cross-talk between plants and beneficial soil microbiome. *Phytochemistry Reviews*, 21(5), 1739–1760. <https://doi.org/10.1007/s11101-022-09806-3>
- Baltz, R. H. (2008). Renaissance in antibacterial discovery from actinomycetes. *Current Opinion in Pharmacology*, 8(5), 557–563. <https://doi.org/10.1016/j.coph.2008.04.008>
- Bannwarth, M., Heckmann-Pohl, D., Bastian, S., Giffhorn, F., & Schulz, G. E. (2006). Reaction geometry and thermostable variant of pyranose 2-oxidase from the white-rot fungus *Peniophora* sp. *Biochemistry*, 45(21), 6587–6595. <https://doi.org/10.1021/bi052465d>
- Barnett, S. J., Roget, D. K., & Ryder, M. H. (2006). Suppression of Rhizoctonia solani AG-8 induced disease on wheat by the interaction between *Pantoea*, *Exiguobacterium*, and *Microbacteria*. *Soil Research*, 44(4), 331–342. <https://doi.org/10.1071/SR05113>
- Bei, W., Luo, Q., Shi, H., Zhou, H., Zhou, M., Zhang, X., & Huang, Y. (2022). Cryo-EM structures of LolCDE reveal the molecular mechanism of bacterial lipoprotein sorting in *Escherichia coli*. *PLoS Biology*, 20(10), e3001823. <https://doi.org/10.1371/journal.pbio.3001823>
- Bilsing, F. L., Anlauf, M. T., Hachani, E., Khosa, S., & Schmitt, L. (2023). ABC Transporters in Bacterial Nanomachineries. *International Journal of Molecular Sciences*, 24(7), 6227. <https://doi.org/10.3390/ijms24076227>
- Boller, T., & He, S. Y. (2009). Innate immunity in plants: an arms race between pattern recognition receptors in plants and effectors in microbial pathogens. *Science*, 324(5928), 742–744. <https://doi.org/10.1126/science.1171647>
- Botos, I., Segal, D. M., & Davies, D. R. (2011). The structural biology of Toll-like receptors. *Structure*, 19(4), 447–459. <https://doi.org/10.1016/j.str.2011.02.004>
- Bowsher, A. W., Evans, S., Tiemann, L. K., & Friesen, M. L. (2018). Effects of soil nitrogen availability on rhizodeposition in plants: a review. *Plant and Soil*, 423, 59–85. <https://doi.org/10.1007/s11104-017-3497-1>

- Braithwaite, B. E., Roush, T. L., Stewart, A., & Hampton, J. G. (2008). Opportunities for developing value-added brassica seed. *NZGA: Research and Practice Series*, 14, 139-146. <https://doi.org/10.33584/rps.14.2008.3174>
- Braune, A., Engst, W., & Blaut, M. (2016). Identification and functional expression of genes encoding flavonoid O- and C-glycosidases in intestinal bacteria. *Environmental Microbiology*, 18(7), 2117-2129. <https://doi.org/10.1111/1462-2920.12864>
- Brokx, S. J., Ellison, M., Locke, T., Bottorff, D., Frost, L., & Weiner, J. H. (2004). Genome-wide analysis of lipoprotein expression in *Escherichia coli* MG1655. *Journal of bacteriology*, 186(10), 3254-3258. <https://doi.org/10.1128/jb.186.10.3254-3258.2004>
- Bucar, F., Xiao, J., & Ochensberger, S. (2021). Flavonoid C-glycosides in diets. *Handbook of Dietary Phytochemicals*, 117-153. https://doi.org/10.1007/978-981-13-1745-3_6-1
- Buchanan, B. B., Gruissem, W., Jones, R. L. (2015a). Symbiotic nitrogen fixation. Buchanan, B. B. (Ed), Gruissem, W. (Ed), Jones, R. L. (Ed), *Biochemistry and Molecular Biology of Plants*. (2nd ed., 718-735). John Wiley & Sons.
- Buchanan, B. B., Gruissem, W., Jones, R. L. (2015b). Responses to Plant Pathogens. Buchanan, B. B. (Ed), Gruissem, W. (Ed), Jones, R. L. (Ed), *Biochemistry and Molecular Biology of Plants*. (2nd ed., 984-1051). John Wiley & Sons.
- CABI (2022). Black rot. Available at: <https://www.plantwise.org/knowledgebank/datasheet/56919>; <https://doi.org/10.1079/pwkb.species.56919>
- Caplan, J., Padmanabhan, M., & Dinesh-Kumar, S. P. (2008). Plant NB-LRR immune receptors: from recognition to transcriptional reprogramming. *Cell Host & Microbe*, 3(3), 126-135. <https://doi.org/10.1016/j.chom.2008.02.010>
- Casas, M. I., Falcone-Ferreira, M. L., Jiang, N., Mejía-Guerra, M. K., Rodríguez, E., Wilson, T., ... & Grotewold, E. (2016). Identification and characterization of maize salmon silks genes involved in insecticidal maysin biosynthesis. *The Plant Cell*, 28(6), 1297-1309. <https://doi.org/10.1105/tpc.16.00003>
- Cavener, D. R. (1992). GMC oxidoreductases: a newly defined family of homologous proteins with diverse catalytic activities. *Journal of Molecular Biology*, 223(3), 811-814. [https://doi.org/10.1016/0022-2836\(92\)90992-S](https://doi.org/10.1016/0022-2836(92)90992-S)
- Chimalakonda, G., Ruiz, N., Chng, S. S., Garner, R. A., Kahne, D., & Silhavy, T. J. (2011). Lipoprotein LptE is required for the assembly of LptD by the β -barrel assembly machine in the outer membrane of *Escherichia coli*. *Proceedings of the National Academy of Sciences USA*, 108(6), 2492-2497. <https://doi.org/10.1073/pnas.1019089108>
- Choi, J., Kim, Y., Eser, B. E., & Han, J. (2023). Theoretical study on the glycosidic C-C bond cleavage of 3''-oxo-puerarin. *Scientific Reports*, 13(1), 16282. <https://doi.org/10.1038/s41598-023-43379-1>
- Chruszcz, M., Potrzebowski, W., Zimmerman, M. D., Grabowski, M., Zheng, H., Lasota, P., & Minor, W. (2008). Analysis of solvent content and oligomeric states

82 REFERENCES

- in protein crystals—does symmetry matter? *Protein Science*, 17(4), 623-632.
<https://doi.org/10.1110/ps.073360508>
- Cocking, E. C. (2003). Endophytic colonization of plant roots by nitrogen-fixing bacteria. *Plant and Soil*, 252, 169-175.
<https://doi.org/10.1023/A:1024106605806>
- Cordovez, V., Schop, S., Hordijk, K., Dupré de Boulois, H., Coppens, F., Hanssen, I., ... & Carrión, V. J. (2018). Priming of plant growth promotion by volatiles of root-associated *Microbacterium* spp. *Applied and Environmental Microbiology*, 84(22), e01865-18. <https://doi.org/10.1128/AEM.01865-18>
- Costa-Gutierrez, S. B., Adler, C., Espinosa-Urgel, M., & de Cristóbal, R. E. (2022). *Pseudomonas putida* and its close relatives: Mixing and mastering the perfect tune for plants. *Applied Microbiology and Biotechnology*, 106(9-10), 3351-3367.
<https://doi.org/10.1007/s00253-022-11881-7>
- Crow, A., Greene, N. P., Kaplan, E., & Koronakis, V. (2017). Structure and mechanotransmission mechanism of the MacB ABC transporter superfamily. *Proceedings of the National Academy of Sciences USA*, 114(47), 12572-12577.
<https://doi.org/10.1073/pnas.1712153114>
- Cruz, J., Tenreiro, R., & Cruz, L. (2017). Assessment of diversity of *Xanthomonas campestris* pathovars affecting cruciferous plants in Portugal and disclosure of two novel *X. campestris* pv. *campestris* races. *Journal of Plant Pathology*, 403-414. <https://www.jstor.org/stable/44686785>
- Dai, S., Chen, Q., Jiang, M., Wang, B., Xie, Z., Yu, N., ... & Tian, B. (2021). Colonized extremophile *Deinococcus radiodurans* alleviates toxicity of cadmium and lead by suppressing heavy metal accumulation and improving antioxidant system in rice. *Environmental Pollution*, 284, 117127.
<https://doi.org/10.1016/j.envpol.2021.117127>
- Davis, N. S., Silverman, G. J., & Masurovsky, E. B. (1963). Radiation-resistant, pigmented coccus isolated from haddock tissue. *Journal of Bacteriology*, 86(2), 294-298. <https://doi.org/10.1128/jb.86.2.294-298.1963>
- Delhaye, A., Collet, J. F., & Laloux, G. (2019). A fly on the wall: how stress response systems can sense and respond to damage to peptidoglycan. *Frontiers in Cellular and Infection Microbiology*, 9, 380.
<https://doi.org/10.3389/fcimb.2019.00380>
- Dias, M. C., Pinto, D. C., & Silva, A. M. (2021). Plant flavonoids: Chemical characteristics and biological activity. *Molecules*, 26(17), 5377.
<https://doi.org/10.3390/molecules26175377>
- Dixon, R. A., & Paiva, N. L. (1995). Stress-induced phenylpropanoid metabolism. *The Plant Cell*, 7(7), 1085. <https://doi.org/10.1105/tpc.7.7.1085>
- Doyle, S. A. (2005). High-throughput cloning for proteomics research. *Methods in Molecular Biology*, 310, 107-113. https://doi.org/10.1007/978-1-59259-948-6_7
- El Rayes, J., Rodríguez-Alonso, R., & Collet, J. F. (2021a). Lipoproteins in Gram-negative bacteria: new insights into their biogenesis, subcellular targeting and

- functional roles. *Current Opinion in Microbiology*, 61, 25-34.
<https://doi.org/10.1016/j.mib.2021.02.003>
- El Rayes, J., Szewczyk, J., Deghelt, M., Csoma, N., Matagne, A., Iorga, B. I., ... & Collet, J. F. (2021b). Disorder is a critical component of lipoprotein sorting in Gram-negative bacteria. *Nature Chemical Biology*, 17(10), 1093-1100.
<https://doi.org/10.1038/s41589-021-00845-z>
- Ferreira, R. M., Moreira, L. M., Ferro, J. A., Soares, M. R., Laia, M. L., Varani, A. M., ... & Ferro, M. I. T. (2016). Unravelling potential virulence factor candidates in *Xanthomonas citri* subsp. *citri* by secretome analysis. *PeerJ*, 4, e1734.
<https://doi.org/10.7717/peerj.1734>
- Frete, C. D., Suryani, R., Purwestri, Y. A., Nuringtyas, T. R., & Widiyanto, D. (2018). Diversity of endophytic bacteria in sweet sorghum (*Sorghum bicolor* (L.) Moench) and their potential for promoting plant growth. *Indian Journal of Science and Technology*, 11(11), 1-10.
<https://doi.org/10.17485/ijst/2018/v11i11/120283>
- Furlanetto V, Divne C (2023). LolA and LolB from the plant-pathogen *Xanthomonas campestris* forms a stable heterodimeric complex in the absence of lipoprotein. *Frontiers in Microbiology*, 14, 1216799.
<https://doi.org/10.3389/fmicb.2023.1216799>
- García-Ruiz, J. M. (2003). Counterdiffusion methods for macromolecular crystallization. In *Methods in Enzymology* (Vol. 368, pp. 130-154). Academic Press. [https://doi.org/10.1016/S0076-6879\(03\)68008-0](https://doi.org/10.1016/S0076-6879(03)68008-0)
- Gay, N. J., & Gangloff, M. (2007). Structure and function of Toll receptors and their ligands. *Annual Reviews of Biochemistry*, 76, 141-165.
<https://doi.org/10.1146/annurev.biochem.76.060305.151318>
- Gennity, J. M., & Inouye, M. (1991). The protein sequence responsible for lipoprotein membrane localization in *Escherichia coli* exhibits remarkable specificity. *Journal of Biological Chemistry*, 266(25), 16458-16464.
[https://doi.org/10.1016/S0021-9258\(18\)55322-9](https://doi.org/10.1016/S0021-9258(18)55322-9)
- González, J. F., Degrassi, G., Devescovi, G., De Vleeschauwer, D., Höfte, M., Myers, M. P., & Venturi, V. (2012). A proteomic study of *Xanthomonas oryzae* pv. *oryzae* in rice xylem sap. *Journal of Proteomics*, 75(18), 5911-5919.
<https://doi.org/10.1016/j.jprot.2012.07.019>
- Grabowicz, M. (2018). Lipoprotein Transport: Greasing the Machines of Outer Membrane Biogenesis: Re-examining lipoprotein transport mechanisms among diverse Gram-negative bacteria while exploring new discoveries and questions. *BioEssays*, 40(4), 1700187. <https://doi.org/10.1002/bies.201700187>
- Guo, L., Fang, W. W., Guo, L. L., Yao, C. F., Zhao, Y. X., Ge, F., & Dai, Y. J. (2019). Biodegradation of the neonicotinoid insecticide acetamiprid by actinomycetes *Streptomyces canus* CGMCC 13662 and characterization of the novel nitrile hydratase involved. *Journal of Agricultural and Food Chemistry*, 67(21), 5922-5931. <https://doi.org/10.1021/acs.jafc.8b06513>
- Hakalová, E., Čechová, J., Tekielska, D. A., Eichmeier, A., & Pothier, J. F. (2022). Combined effect of thyme and clove phenolic compounds on *Xanthomonas*

84 REFERENCES

- campestris* pv. *campestris* and biocontrol of black rot disease on cabbage seeds. *Frontiers in Microbiology*, 13, 1007988. <https://doi.org/10.3389/fmicb.2022.1007988>
- Hallberg, B. M., Henriksson, G., Pettersson, G., & Divne, C. (2002). Crystal structure of the flavoprotein domain of the extracellular flavocytochrome cellobiose dehydrogenase. *Journal of Molecular Biology*, 315(3), 421-434. <https://doi.org/10.1006/jmbi.2001.5246>
- Hallberg, B. M., Leitner, C., Haltrich, D., & Divne, C. (2004). Crystal structure of the 270 kDa homotetrameric lignin-degrading enzyme pyranose 2-oxidase. *Journal of Molecular Biology*, 341(3), 781-796. <https://doi.org/10.1016/j.jmb.2004.06.033>
- Han, J. W., Choi, G. J., & Kim, B. S. (2018). Antimicrobial aromatic polyketides: a review of their antimicrobial properties and potential use in plant disease control. *World Journal of Microbiology and Biotechnology*, 34, 1-10. <https://doi.org/10.1007/s11274-018-2546-0>
- Hartmann, A., Rothballer, M., & Schmid, M. (2008). Lorenz Hiltner, a pioneer in rhizosphere microbial ecology and soil bacteriology research. *Plant and Soil*, 312, 7-14. <https://doi.org/10.1007/s11104-007-9514-z>
- Hartwig, U. A., & Phillips, D. A. (1991). Release and modification of nod-gene-inducing flavonoids from alfalfa seeds. *Plant Physiology*, 95(3), 804-807. <https://doi.org/10.1104/pp.95.3.804>
- Hassan, S., & Mathesius, U. (2012). The role of flavonoids in root–rhizosphere signalling: opportunities and challenges for improving plant–microbe interactions. *Journal of Experimental Botany*, 63(9), 3429-3444. <https://doi.org/10.1093/jxb/err430>
- Hassan, N., Tan, T. C., Spadiut, O., Pisanelli, I., Fusco, L., Haltrich, D., ... & Divne, C. (2013). Crystal structures of Phanerochaete chrysosporium pyranose 2-oxidase suggest that the N-terminus acts as a propeptide that assists in homotetramer assembly. *FEBS Open Bio*, 3, 496-504. <https://doi.org/10.1016/j.fob.2013.10.010>
- Hattori, M., Kanda, T., Shu, Y. Z., Akao, T., Kobashi, K., & Namba, T. (1988a). Metabolism of barbaloin by intestinal bacteria. *Chemical and Pharmaceutical Bulletin*, 36(11), 4462-4466. <https://doi.org/10.1248/cpb.36.4462>
- Hattori, M., Shu, Y. Z., El-Sedawy, A. I., Namba, T., Kobashi, K., & Tomimori, T. (1988b). Metabolism of homoorientin by human intestinal bacteria. *Journal of Natural Products*, 51(5), 874-878. <https://doi.org/10.1021/np50059a010>
- Hattori, M., Shu, Y. Z., Tomimori, T., Kobashi, K., & Namba, T. (1989). A bacterial cleavage of the C-glucosyl bond of mangiferin and bergenin. *Phytochemistry*, 28(4), 1289-1290. [https://doi.org/10.1016/0031-9422\(89\)80239-0](https://doi.org/10.1016/0031-9422(89)80239-0)
- Hayashi, S., & Wu, H. C. (1990). Lipoproteins in bacteria. *Journal of Bioenergetics and Biomembranes*, 22, 451-471. <https://doi.org/10.1007/BF00763177>

- Hayat, R., Ali, S., Amara, U., Khalid, R., & Ahmed, I. (2010). Soil beneficial bacteria and their role in plant growth promotion: a review. *Annals of Microbiology*, 60, 579-598. <https://doi.org/10.1007/s13213-010-0117-1>
- He, P., Wang, S., Li, S., Liu, S., Zhou, S., Wang, J., ... & Ma, W. (2023). Structural mechanism of a dual-functional enzyme DgpA/B/C as both a C-glycoside cleaving enzyme and an O-to C-glycoside isomerase. *Acta Pharmaceutica Sinica B*, 13(1), 246-255. <https://doi.org/10.1016/j.apsb.2022.05.022>
- Herzog, P. L., Sützl, L., Eisenhut, B., Maresch, D., Haltrich, D., Obinger, C., & Peterbauer, C. K. (2019). Versatile oxidase and dehydrogenase activities of bacterial pyranose 2-oxidase facilitate redox cycling with manganese peroxidase in vitro. *Applied and Environmental Microbiology*, 85(13), e00390-19. <https://doi.org/10.1128/AEM.00390-19>
- Horler, R. S., Butcher, A., Papangelopoulos, N., Ashton, P. D., & Thomas, G. H. (2009). Echo LOCATION: an in silico analysis of the subcellular locations of *Escherichia coli* proteins and comparison with experimentally derived locations. *Bioinformatics*, 25(2), 163-166. <https://doi.org/10.1093/bioinformatics/btn596>
- Huang, C. Y., & Jin, H. (2022). Coordinated epigenetic regulation in plants: a potent managerial tool to conquer biotic stress. *Frontiers in Plant Science*, 12, 795274. <https://doi.org/10.3389/fpls.2021.795274>
- Innis, M. A., Tokunaga, M., Williams, M. E., Loranger, J. M., Chang, S. Y., Chang, S., & Wu, H. C. (1984). Nucleotide sequence of the *Escherichia coli* prolipoprotein signal peptidase (lsp) gene. *Proceedings of the National Academy of Sciences USA*, 81(12), 3708-3712. <https://doi.org/10.1073/pnas.81.12.3708>
- Ito, H., Watanabe, H., Takehisa, M., & Iizuka, H. (1983). Isolation and identification of radiation-resistant cocci belonging to the genus *Deinococcus* from sewage sludges and animal feeds. *Agricultural and Biological Chemistry*, 47(6), 1239-1247. <https://doi.org/10.1271/bbb1961.47.1239>
- Jaiman, D., Nagampalli, R., & Persson, K. (2023). A comparative analysis of lipoprotein transport proteins: LolA and LolB from *Vibrio cholerae* and LolA from *Porphyromonas gingivalis*. *Scientific Reports*, 13(1), 6605. <https://doi.org/10.1038/s41598-023-33705-y>
- Janeway, C. A. (1989, January). Approaching the asymptote? Evolution and revolution in immunology. In Cold Spring Harbor symposia on quantitative biology (Vol. 54, pp. 1-13). Cold Spring Harbor Laboratory Press. <https://doi.org/10.1101/SQB.1989.054.01.003>
- Ji, Y., Li, B., Qiao, M., Li, J., Xu, H., Zhang, L., & Zhang, X. (2020). Advances on the in vivo and in vitro glycosylations of flavonoids. *Applied Microbiology and Biotechnology*, 104, 6587-6600. <https://doi.org/10.1007/s00253-020-10667-z>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kaplan, E., Greene, N. P., Crow, A., & Koronakis, V. (2018). Insights into bacterial lipoprotein trafficking from a structure of LolA bound to the LolC periplasmic

86 REFERENCES

- domain. *Proceedings of the National Academy of Sciences USA*, 115(31), E7389-E7397. <https://doi.org/10.1073/pnas.1806822115>
- Kaplan, E., Greene, N. P., Jepson, A. E., & Koronakis, V. (2022). Structural basis of lipoprotein recognition by the bacterial Lol trafficking chaperone LolA. *Proceedings of the National Academy of Sciences USA*, 119(36), e2208662119. <https://doi.org/10.1073/pnas.2208662119>
- Kenrick, P., & Crane, P. R. (1997). The origin and early evolution of plants on land. *Nature*, 389(6646), 33–39. <https://doi.org/10.1038/37918>
- Kikuchi, K., Matsushita, N., Suzuki, K., & Hogetsu, T. (2007). Flavonoids induce germination of basidiospores of the ectomycorrhizal fungus *Suillus bovinus*. *Mycorrhiza*, 17, 563–570. <https://doi.org/10.1007/s00572-007-0131-8>
- Kingston, R. L., Scopes, R. K., & Baker, E. N. (1996). The structure of glucose-fructose oxidoreductase from *Zymomonas mobilis*: an osmoprotective periplasmic enzyme containing non-dissociable NADP. *Structure*, 4(12), 1413–1428. [https://doi.org/10.1016/s0969-2126\(96\)00149-9](https://doi.org/10.1016/s0969-2126(96)00149-9)
- Kovacs-Simon, A., Titball, R. W., & Michell, S. L. (2011). Lipoproteins of bacterial pathogens. *Infection and Immunity*, 79(2), 548–561. <https://doi.org/10.1128/iai.00682-10>
- Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology*, 372(3), 774–797. <https://doi.org/10.1016/j.jmb.2007.05.022>
- Kujawa, M., Ebner, H., Leitner, C., Hallberg, B. M., Prongjit, M., Sucharitakul, J., ... & Divne, C. (2006). Structural basis for substrate binding and regioselective oxidation of monosaccharides at C3 by pyranose 2-oxidase. *Journal of Biological Chemistry*, 281(46), 35104–35115. <https://doi.org/10.1074/jbc.M604718200>
- Kumano, T., Hori, S., Watanabe, S., Terashita, Y., Yu, H. Y., Hashimoto, Y., ... & Kobayashi, M. (2021). FAD-dependent C-glycoside–metabolizing enzymes in microorganisms: Screening, characterization, and crystal structure analysis. *Proceedings of the National Academy of Sciences USA*, 118(40), e2106580118. <https://doi.org/10.1073/pnas.2106580118>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35:1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Lai, W. A., Kämpfer, P., Arun, A. B., Shen, F. T., Huber, B., Rekha, P. D., & Young, C. C. (2006). *Deinococcus ficus* sp. nov., isolated from the rhizosphere of *Ficus religiosa* L. *International Journal of Systematic and Evolutionary Microbiology*, 56(4), 787–791. <https://doi.org/10.1099/ijs.0.64007-0>
- Laloux, G., & Collet, J. F. (2017). Major Tom to ground control: how lipoproteins communicate extracytoplasmic stress to the decision center of the cell. *Journal of Bacteriology*, 199(21), 10–1128. <https://doi.org/10.1128/jb.00216-17>
- Lee, S. Y., Lee, Y. Y., & Cho, K. S. (2023). Effect of *Novosphingobium* sp. CuT1 inoculation on the rhizoremediation of heavy metal-and diesel-contaminated

- soil planted with tall fescue. *Environmental Science and Pollution Research*, 30(6), 16612-16625. <https://doi.org/10.1007/s11356-022-23339-4>
- Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49(W1), W293-W296. <https://doi.org/10.1093/nar/gkab301>
- Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., & Henrissat, B. (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels*, 6(1), 1-14. <https://doi.org/10.1186/1754-6834-6-41>
- Liao, C. T., Chiang, Y. C., & Hsiao, Y. M. (2019). Functional characterization and proteomic analysis of lolA in *Xanthomonas campestris* pv. *campestris*. *BMC Microbiology*, 19(1), 1-18. <https://doi.org/10.1186/s12866-019-1387-9>
- Liao, C. T., Li, C. E., Chang, H. C., Hsu, C. H., Chiang, Y. C., & Hsiao, Y. M. (2022). The lolB gene in *Xanthomonas campestris* pv. *campestris* is required for bacterial attachment, stress tolerance, and virulence. *BMC Microbiology*, 22(1), 1-13. <https://doi.org/10.1186/s12866-021-02416-7>
- Liu, X., Shen, J., & Zhu, K. (2022). Antibacterial activities of plant-derived xanthones. *RSC Medicinal Chemistry*, 13(2), 107-116. <https://doi.org/10.1039/d1md00351h>
- Lodewyckx, C., Vangronsveld, J., Porteous, F., Moore, E. R., Taghavi, S., Mezgeay, M., & der Lelie, D. V. (2002). Endophytic bacteria and their potential applications. *Critical Reviews in Plant Sciences*, 21(6), 583-606. <https://doi.org/10.1080/0735-260291044377>
- Long, S. R. (1996). Rhizobium symbiosis: nod factors in perspective. *The Plant Cell*, 8(10), 1885. <https://doi.org/10.1105/tpc.8.10.1885>
- LoVullo, E. D., Wright, L. F., Isabella, V., Huntley, J. F., & Pavelka Jr, M. S. (2015). Revisiting the Gram-negative lipoprotein paradigm. *Journal of Bacteriology*, 197(10), 1705-1715. <https://doi.org/10.1128/jb.02414-14>
- Lugtenberg, B., & Kamilova, F. (2009). Plant-growth-promoting rhizobacteria. *Annual Review of Microbiology*, 63, 541-556. <https://doi.org/10.1146/annurev.micro.62.081307.162918>
- Lu, L., Monakhos, S. G., Lim, Y. P., & Yi, S. Y. (2021). Early Defense Mechanisms of Brassica oleracea in Response to Attack by *Xanthomonas campestris* pv. *campestris*. *Plants*, 10(12), 2705. <https://doi.org/10.3390/plants10122705>
- Ma, C., Hua, J., Li, H., Zhang, J., & Luo, S. (2022). Inoculation with carbofuran-degrading rhizobacteria promotes maize growth through production of IAA and regulation of the release of plant-specialized metabolites. *Chemosphere*, 307, 136027. <https://doi.org/10.1016/j.chemosphere.2022.136027>
- Mandiriza-Mukwirimba, G., Kritzing, Q., and Aveling, T. (2016). A survey of brassica vegetable smallholder farmers in the Gauteng and Limpopo provinces of South Africa. *Journal of Agriculture and Rural Development in the Tropics*, 117, 35-44. <https://hdl.handle.net/2263/53080>

88 REFERENCES

- Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., ... & Foster, G. D. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology*, 13(6), 614-629.
<https://doi.org/10.1111/j.1364-3703.2012.00804.x>
- Matse, D. T., Huang, C. H., Huang, Y. M., & Yen, M. Y. (2020). Effects of coinoculation of Rhizobium with plant growth promoting rhizobacteria on the nitrogen fixation and nutrient uptake of *Trifolium repens* in low phosphorus soil. *Journal of Plant Nutrition*, 43(5), 739-752.
<https://doi.org/10.1080/01904167.2019.1702205>
- Matsuyama, S. L., Tajima, T., & Tokuda, H. (1995). A novel periplasmic carrier protein involved in the sorting and transport of *Escherichia coli* lipoproteins destined for the outer membrane. *The EMBO Journal*, 14(14), 3365-3372.
<https://doi.org/10.1002/j.1460-2075.1995.tb07342.x>
- Matsuyama, S. I., Yokota, N., & Tokuda, H. (1997). A novel outer membrane lipoprotein, LolB (HemM), involved in the LolA (p20)-dependent localization of lipoproteins to the outer membrane of *Escherichia coli*. *The EMBO Journal*, 16(23), 6947-6955. <https://doi.org/10.1093/emboj/16.23.6947>
- Mattevi, A. (1998). The PHBH fold: not only flavoenzymes. *Biophysical Chemistry*, 70(3), 217-222. [https://doi.org/10.1016/S0301-4622\(97\)00126-9](https://doi.org/10.1016/S0301-4622(97)00126-9)
- Matthews, B. W. (1968). Solvent content of protein crystals. *Journal of Molecular Biology*, 33(2), 491-497. [https://doi.org/10.1016/0022-2836\(68\)90205-2](https://doi.org/10.1016/0022-2836(68)90205-2)
- Mattimore, V., & Battista, J. R. (1996). Radioresistance of *Deinococcus radiodurans*: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation. *Journal of Bacteriology*, 178(3), 633-637.
<https://doi.org/10.1128/jb.178.3.633-637.1996>
- McPherson, A., & Gavira, J. A. (2014). Introduction to protein crystallization. *Acta Crystallographica Section F: Structural Biology Communications*, 70(1), 2-20.
<https://doi.org/10.1107/S2053230X13033141>
- Mendes, S., Banha, C., Madeira, J., Santos, D., Miranda, V., Manzanera, M., ... & Martins, L. O. (2016). Characterization of a bacterial pyranose 2-oxidase from *Arthrobacter siccitolerans*. *Journal of Molecular Catalysis B: Enzymatic*, 133, S34-S43. <https://doi.org/10.1016/j.molcatb.2016.11.005>
- Malinverni, J. C., Werner, J., Kim, S., Sklar, J. G., Kahne, D., Misra, R., & Silhavy, T. J. (2006). YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in *Escherichia coli*. *Molecular Microbiology*, 61(1), 151-164.
<https://doi.org/10.1111/j.1365-2958.2006.05211.x>
- Mori, T., Kumano, T., He, H., Watanabe, S., Senda, M., Moriya, T., ... & Kobayashi, M. (2021). C-Glycoside metabolism in the gut and in nature: Identification, characterization, structural analyses and distribution of CC bond-cleaving enzymes. *Nature Communications*, 12(1), 6294.
<https://doi.org/10.1038/s41467-021-26585-1>
- Morohoshi, T., Someya, N., & Ikeda, T. (2009). Novel N-acylhomoserine lactone-degrading bacteria isolated from the leaf surface of *Solanum tuberosum* and

- their quorum-quenching properties. *Bioscience, Biotechnology, and Biochemistry*, 73(9), 2124-2127. <https://doi.org/10.1271/bbb.90283>
- Morris, J. L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., ... & Donoghue, P. C. (2018). The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences USA*, 115(10), E2274-E2283. <https://doi.org/10.1073/pnas.1719588115>
- Mus, F., Crook, M. B., Garcia, K., Garcia Costas, A., Geddes, B. A., Kouri, E. D., ... & Peters, J. W. (2016). Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Applied and Environmental Microbiology*, 82(13), 3698-3710. <https://doi.org/10.1128/AEM.01055-16>
- Nakamura, K., Zhu, S., Komatsu, K., Hattori, M., & Iwashima, M. (2019). Expression and characterization of the human intestinal bacterial enzyme which cleaves the C-glycosidic bond in 3"-oxo-puerarin. *Biological and Pharmaceutical Bulletin*, 42(3), 417-423. <https://doi.org/10.1248/bpb.b18-00729>
- Nakamura, K., Zhu, S., Komatsu, K., Hattori, M., & Iwashima, M. (2020). Deglycosylation of the isoflavone C-glucoside puerarin by a combination of two recombinant bacterial enzymes and 3-oxo-glucose. *Applied and Environmental Microbiology*, 86(14), e00607-20. <https://doi.org/10.1128/AEM.00607-20>
- Okuda, S., Watanabe, S., & Tokuda, H. (2008). A short helix in the C-terminal region of LolA is important for the specific membrane localization of lipoproteins. *FEBS Letters*, 582(15), 2247-2251. <https://doi.org/10.1016/j.febslet.2008.05.022>
- Okuda, S., & Tokuda, H. (2009). Model of mouth-to-mouth transfer of bacterial lipoproteins through inner membrane LolC, periplasmic LolA, and outer membrane LolB. *Proceedings of the National Academy of Sciences USA*, 106(14), 5877-5882. <https://doi.org/10.1073/pnas.0900896106>
- Okuda, S., & Tokuda, H. (2011) Lipoprotein sorting in bacteria. *Annual Review of Microbiology*, 65, 239-259. <https://doi.org/10.1146/annurev-micro-090110-102859>
- Oren, A. & Garrity, G. M. (2021). Valid publication of the names of forty-two phyla of prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*, 71(10), <https://doi.org/10.1099/ijsem.0.005056>
- Palai, J. B., Malik, G. C., Maitra, S., & Banerjee, M. (2021). Role of rhizobium on growth and development of groundnut: A review. *International Journal of Agriculture Environment and Biotechnology*, 14(1), 63-73. <https://doi.org/10.30954/0974-1712.01.2021.7>
- Pankiewicz, V. C. S., Delaux, P. M., Infante, V., Hirsch, H. H., Rajasekar, S., Zamora, P., ... & Ané, J. M. (2022). Nitrogen fixation and mucilage production on maize aerial roots is controlled by aerial root development and border cell functions. *Frontiers in Plant Science*, 13, 977056. <https://doi.org/10.3389/fpls.2022.977056>
- Paradis-Bleau, C., Markovski, M., Uehara, T., Lupoli, T. J., Walker, S., Kahne, D. E., & Bernhardt, T. G. (2010). Lipoprotein cofactors located in the outer membrane activate bacterial cell wall polymerases. *Cell*, 143(7), 1110-1120. <https://doi.org/10.1016/j.cell.2010.11.037>

90 REFERENCES

- Patel, A., Sahu, K. P., Mehta, S., Balamurugan, A., Kumar, M., Sheoran, N., ... & Kumar, A. (2022). Rice leaf endophytic *Microbacterium testaceum*: Antifungal actinobacterium confers immunocompetence against rice blast disease. *Frontiers in Microbiology*, 13, 1035602. <https://doi.org/10.3389/fmicb.2022.1035602>
- Peng, F., Zhang, L., Luo, X., Dai, J., An, H., Tang, Y., & Fang, C. (2009). *Deinococcus xinjiangensis* sp. nov., isolated from desert soil. *International Journal of Systematic and Evolutionary Microbiology*, 59(4), 709-713. <https://doi.org/10.1099/ijs.O.004564-0>
- Perez-Bonilla, M., Oves-Costales, D., Gonzalez, I., de la Cruz, M., Martin, J., Vicente, F., ... & Reyes, F. (2020). Krisynomycins, imipenem potentiators against methicillin-resistant *Staphylococcus aureus*, produced by *Streptomyces canus*. *Journal of Natural Products*, 83(9), 2597-2606. <https://doi.org/10.1021/acs.jnatprod.0c00294>
- Prongjit, M., Sucharitakul, J., Palfey, B. A., & Chaiken, P. (2013). Oxidation mode of pyranose 2-oxidase is controlled by pH. *Biochemistry*, 52(8), 1437-1445. <https://doi.org/10.1021/bi301442x>
- Qin, A., Scott, D. W., Thompson, J. A., & Mann, B. J. (2009). Identification of an essential *Francisella tularensis* subsp. *tularensis* virulence factor. *Infection and Immunity*, 77(1), 152-161. <https://doi.org/10.1128/iai.01113-08>
- Qin, A., Zhang, Y., Clark, M. E., Rabideau, M. M., Millan Barea, L. R., & Mann, B. J. (2014). FipB, an essential virulence factor of *Francisella tularensis* subsp. *tularensis*, has dual roles in disulfide bond formation. *Journal of Bacteriology*, 196(20), 3571-3581. <https://doi.org/10.1128/jb.01359-13>
- Rainey, F. A., Ray, K., Ferreira, M., Gatz, B. Z., Nobre, M. F., Bagaley, D., ... & da Costa, M. S. (2005). Extensive diversity of ionizing-radiation-resistant bacteria recovered from Sonoran Desert soil and description of nine new species of the genus *Deinococcus* obtained from a single soil sample. *Applied and Environmental Microbiology*, 71(9), 5225-5235. <https://doi.org/10.1128/AEM.71.9.5225-5235.2005>
- Remali, J., Sahidin, I., & Aizat, W. M. (2022). Xanthone biosynthetic pathway in plants: A review. *Frontiers in Plant Science*, 13, 809497. <https://doi.org/10.3389/fpls.2022.809497>
- Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42(W1), W320-W324. <https://doi.org/10.1093/nar/gku316>
- Rossmann, M. G., Liljas, A., Brändén, C. I., & Banaszak, L. J. (1975). 2 Evolutionary and structural relationships among dehydrogenases. In *The enzymes* (Vol. 11, pp. 61-102). Academic Press. [https://doi.org/10.1016/S1874-6047\(08\)60210-3](https://doi.org/10.1016/S1874-6047(08)60210-3)
- Rowland, P., Basak, A. K., Gover, S., Levy, H. R., & Adams, M. J. (1994). The three-dimensional structure of glucose 6-phosphate dehydrogenase from *Leuconostoc mesenteroides* refined at 2.0 Å resolution. *Structure*, 2(11), 1073-1087. [https://doi.org/10.1016/s0969-2126\(94\)00110-3](https://doi.org/10.1016/s0969-2126(94)00110-3)

- Rupp, B. (2015). Origin and use of crystallization phase diagrams. *Acta Crystallographica Section F: Structural Biology Communications*, 71(3), 247-260. <https://doi.org/10.1107/S2053230X1500374X>
- Ruszcza, E. M., Beerling, D. J., Franks, P. J., Chater, C., Casson, S. A., Gray, J. E., & Hetherington, A. M. (2011). Land plants acquired active stomatal control early in their evolutionary history. *Current Biology*, 21(12), 1030-1035. <https://doi.org/10.1016/j.cub.2011.04.044>
- Sahu, K. P., Patel, A., Kumar, M., Sheoran, N., Mehta, S., Reddy, B., ... & Kumar, A. (2021). Integrated metabarcoding and culturomic-based microbiome profiling of rice phyllosphere reveal diverse and functional bacterial communities for blast disease suppression. *Frontiers in Microbiology*, 12, 780458. <https://doi.org/10.3389/fmicb.2021.780458>
- Samuel, B. (2017). Genomic analyses of the natural microbiome of *Caenorhabditis elegans* (No. 503338). USDOE Joint Genome Institute (JGI), Berkeley, CA (United States). <https://doi.org/10.25585/1488295>
- Santhanam, R., Oh, Y., Kumar, R., Weinhold, A., Luu, V. T., Groten, K., & Baldwin, I. T. (2017). Specificity of root microbiomes in native-grown *Nicotiana attenuata* and plant responses to UVB increase *Deinococcus* colonization. *Molecular Ecology*, 26(9), 2543-2562. <https://doi.org/10.1111/mec.14049>
- Scervino, J. M., Ponce, M. A., Erra-Bassells, R., Vierheilig, H., Ocampo, J. A., & Godeas, A. (2005). Flavonoids exhibit fungal species and genus specific effects on the presymbiotic growth of *Gigaspora* and *Glomus*. *Mycological Research*, 109(7), 789-794. <https://doi.org/10.1017/S0953756205002881>
- Schoch CL, et al. (2020). NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database (Oxford). 2020:baaa062. <https://doi.org/10.1093/database/baaa062>
- Sharma, S., Zhou, R., Wan, L., Feng, S., Song, K., Xu, C., ... & Liao, M. (2021). Mechanism of LolCDE as a molecular extruder of bacterial triacylated lipoproteins. *Nature Communications*, 12(1), 4687. <https://doi.org/10.1038/s41467-021-24965-1>
- Sharma, D., Shree, B., Kumar, S., Kumar, V., Sharma, S., & Sharma, S. (2022). Stress induced production of plant secondary metabolites in vegetables: Functional approach for designing next generation super foods. *Plant Physiology and Biochemistry*, 192, 252-272. <https://doi.org/10.1016/j.plaphy.2022.09.034>
- Shen, B. (2003). Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology*, 7(2), 285-295. [https://doi.org/10.1016/S1367-5931\(03\)00020-6](https://doi.org/10.1016/S1367-5931(03)00020-6)
- Sidhu, V. K., Vorhölter, F. J., Niehaus, K., & Watt, S. A. (2008). Analysis of outer membrane vesicle associated proteins isolated from the plant pathogenic bacterium *Xanthomonas campestris* pv. *campestris*. *BMC Microbiology*, 8(1), 1-16. <https://doi.org/10.1186/1471-2180-8-87>
- Sievers, F., Wilm, A., Dineen, D. G., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments

92 REFERENCES

- using Clustal Omega. *Molecular Systems Biology*, 7, 539.
<https://doi.org/10.1038/msb.2011.75>
- Silhavy, T. J., Kahne, D., & Walker, S. (2010). The bacterial cell envelope. *Cold Spring Harbor Perspectives in Biology*, 2(5), a000414.
<https://doi.org/10.1101/cshperspect.a000414>
- Silva, M. S., De Souza, A. A., Takita, M. A., Labate, C. A., & Machado, M. A. (2011). Analysis of the biofilm proteome of *Xylella fastidiosa*. *Proteome Science*, 9(1), 1-10. <https://doi.org/10.1186/1477-5956-9-58>
- Siqueira, J. O., Safir, G. R., & Nair, M. G. (1991). Stimulation of vesicular-arbuscular mycorrhiza formation and growth of white clover by flavonoid compounds. *New Phytologist*, 118(1), 87-93. <https://doi.org/10.1111/j.1469-8137.1991.tb00568.x>
- Smith, H. C., May, K. L., & Grabowicz, M. (2023). Teasing apart the evolution of lipoprotein trafficking in gram-negative bacteria reveals a bifunctional LolA. *Proceedings of the National Academy of Sciences USA*, 120(6), e2218473120. <https://doi.org/10.1073/pnas.2218473120>
- Spadiut, O., Tan, T. C., Pisanelli, I., Haltrich, D., & Divne, C. (2010). Importance of the gating segment in the substrate-recognition loop of pyranose 2-oxidase. *The FEBS Journal*, 277(13), 2892-2909. <https://doi.org/10.1111/j.1742-4658.2010.07705.x>
- Steinkellner, S., Lendzemo, V., Langer, I., Schweiger, P., Khaosaad, T., Toussaint, J. P., & Vierheilig, H. (2007). Flavonoids and strigolactones in root exudates as signals in symbiotic and pathogenic plant-fungus interactions. *Molecules*, 12(7), 1290-1306. <https://doi.org/10.3390/12071290>
- Sucharitakul, J., Wongnate, T., & Chaiyen, P. (2011). Hydrogen peroxide elimination from C4a-hydroperoxyflavin in a flavoprotein oxidase occurs through a single proton transfer from flavin N5 to a peroxide leaving group. *Journal of Biological Chemistry*, 286(19), 16900-16909. <https://doi.org/10.1074/jbc.M111.222976>
- Suzuki, T., Murai, T., Fukuda, I., Tobe, T., Yoshikawa, M., & Sasakawa, C. (1994). Identification and characterization of a chromosomal virulence gene, vacJ, required for intercellular spreading of *Shigella flexneri*. *Molecular Microbiology*, 11(1), 31-41. <https://doi.org/10.1111/j.1365-2958.1994.tb00287.x>
- Taberman, H., Parkkinen, T., & Rouvinen, J. (2016). Structural and functional features of the NAD (P) dependent Gfo/Idh/MocA protein family oxidoreductases. *Protein Science*, 25(4), 778-786. <https://doi.org/10.1002/pro.2877>
- Tan, T. C., Pitsawong, W., Wongnate, T., Spadiut, O., Haltrich, D., Chaiyen, P., & Divne, C. (2010). H-bonding and positive charge at the N (5)/O (4) locus are critical for covalent flavin attachment in *Trametes* pyranose 2-oxidase. *Journal of Molecular Biology*, 402(3), 578-594. <https://doi.org/10.1016/j.jmb.2010.08.011>
- Tan, T. C., Haltrich, D., & Divne, C. (2011). Regioselective control of β -D-glucose oxidation by pyranose 2-oxidase is intimately coupled to conformational degeneracy. *Journal of Molecular Biology*, 409(4), 588-600. <https://doi.org/10.1016/j.jmb.2011.04.019>

- Tan, T. C., Spadiut, O., Gandini, R., Haltrich, D., & Divne, C. (2014). Structural basis for binding of fluorinated glucose and galactose to *Trametes multicolor* pyranose 2-oxidase variants with improved galactose conversion. *PLoS One*, 9(1), e86736. <https://doi.org/10.1371/journal.pone.0086736>
- Tang, X., Chang, S., Zhang, K., Luo, Q., Zhang, Z., Wang, T., ... & Dong, H. (2021). Structural basis for bacterial lipoprotein relocation by the transporter LolCDE. *Nature Structural & Molecular Biology*, 28(4), 347-355. <https://doi.org/10.1038/s41594-021-00573-x>
- Thomas, C., Aller, S. G., Beis, K., Carpenter, E. P., Chang, G., Chen, L., ... & Tampé, R. (2020). Structural and functional diversity calls for a new classification of ABC transporters. *FEBS Letters*, 594(23), 3767-3775. <https://doi.org/10.1002/1873-3468.13935>
- Thomas, R. (2001). A biosynthetic classification of fungal and streptomycete fused-ring aromatic polyketides. *ChemBioChem*, 2(9), 612-627. [https://doi.org/10.1002/1439-7633\(20010903\)2:9<612::AID-CBIC612>3.0.CO;2-Z](https://doi.org/10.1002/1439-7633(20010903)2:9<612::AID-CBIC612>3.0.CO;2-Z)
- Typas, A., Banzhaf, M., van Sappero, B. V. D. B., Verheul, J., Biboy, J., Nichols, R. J., ... & Vollmer, W. (2010). Regulation of peptidoglycan synthesis by outer-membrane proteins. *Cell*, 143(7), 1097-1109. <https://doi.org/10.1016/j.cell.2010.11.038>
- van Kempen, M., Kim, S. S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C. L., ... & Steinegger, M. (2023). Fast and accurate protein structure search with Foldseek. *Nature Biotechnology*, 1-4. <https://doi.org/10.1038/s41587-023-01773-0>
- Vicente, J. G., & Holub, E. B. (2013). *Xanthomonas campestris* pv. *campestris* (cause of black rot of crucifers) in the genomic era is still a worldwide threat to brassica crops. *Molecular Plant Pathology*, 14(1), 2-18. <https://doi.org/10.1111/j.1364-3703.2012.00833.x>
- Vleeshouwers, V. G., & Oliver, R. P. (2014). Effectors as tools in disease resistance breeding against biotrophic, hemibiotrophic, and necrotrophic plant pathogens. *Molecular Plant-Microbe Interactions*, 27(3), 196-206. <https://doi.org/10.1094/MPMI-10-13-0313-IA>
- Waite, D. W., Chuvochina, M., Pelikan, C., Parks, D. H., Yilmaz, P., Wagner, M., ... & Hugenholtz, P. (2020). Proposal to reclassify the proteobacterial classes *Deltaproteobacteria* and *Oligoflexia*, and the phylum *Thermodesulfobacteria* into four phyla reflecting major functional capabilities. *International Journal of Systematic and Evolutionary Microbiology*, 70(11), 5972-6016. <https://doi.org/10.1099/ijsem.0.004213>
- Wall, E., Majdalani, N., & Gottesman, S. (2018). The complex Rcs regulatory cascade. *Annual Review of Microbiology*, 72, 111-139. <https://doi.org/10.1146/annurev-micro-090817-062640>
- Wang, W., Mao, J., Zhang, Z., Tang, Q., Xie, Y., Zhu, J., ... & Goodfellow, M. (2010). *Deinococcus wulumuqiensis* sp. nov., and *Deinococcus xibeiensis* sp. nov., isolated from radiation-polluted soil. *International Journal of Systematic and*

94 REFERENCES

- Evolutionary Microbiology*, 60(9), 2006-2010.
<https://doi.org/10.1099/ijs.o.015917-0>
- Wang, Y. H., & Irving, H. R. (2011). Developing a model of plant hormone interactions. *Plant Signaling & Behavior*, 6(4), 494-500.
<https://doi.org/10.4161/psb.6.4.14558>
- Wang, Y., Kim, S. G., Wu, J., Huh, H. H., Lee, S. J., Rakwal, R., ... & Kim, S. T. (2013). Secretome analysis of the rice bacterium *Xanthomonas oryzae* (Xoo) using *in vitro* and *in planta* systems. *Proteomics*, 13(12-13), 1901-1912.
<https://doi.org/10.1002/pmic.201200454>
- Wang, G., Roux, B., Feng, F., Guy, E., Li, L., Li, N., ... & Zhou, J. M. (2015). The decoy substrate of a pathogen effector and a pseudokinase specify pathogen-induced modified-self recognition and immunity in plants. *Cell Host & Microbe*, 18(3), 285-295. <https://doi.org/10.1016/j.chom.2015.08.004>
- Webb, B., & Sali, A. (2016). Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics*, 54(1), 5-6.
<https://doi.org/10.1002/cpbi.3>
- Weisskopf, L., Tomasi, N., Santelia, D., Martinoia, E., Langlade, N. B., Tabacchi, R., & Abou-Mansour, E. (2006). Isoflavonoid exudation from white lupin roots is influenced by phosphate supply, root type and cluster-root stage. *New Phytologist*, 171(3), 657-668. <https://doi.org/10.1111/j.1469-8137.2006.01776.x>
- Wierenga, R. K., De Jong, R. J., Kalk, K. H., Hol, W. G. J., & Drenth, J. (1979). Crystal structure of p-hydroxybenzoate hydroxylase. *Journal of Molecular Biology*, 131(1), 55-73. [https://doi.org/10.1016/0022-2836\(79\)90301-2](https://doi.org/10.1016/0022-2836(79)90301-2)
- Wierenga, R. K., Drenth, J., Schulz, G. E., & Huber, R. (1983). Comparison of the three-dimensional protein and nucleotide structure of the FAD-binding domain of p-hydroxybenzoate hydroxylase with the FAD-as well as NADPH-binding domains of glutathione reductase. *Journal of Molecular Biology*, 167(3), 725-739. [https://doi.org/10.1016/S0022-2836\(83\)80106-5](https://doi.org/10.1016/S0022-2836(83)80106-5)
- Wierenga, R. K., Terpstra, P., & Hol, W. G. (1986). Prediction of the occurrence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint. *Journal of Molecular Biology*, 187(1), 101-107. [https://doi.org/10.1016/0022-2836\(86\)90409-2](https://doi.org/10.1016/0022-2836(86)90409-2)
- Wilson, M. M., & Bernstein, H. D. (2016). Surface-exposed lipoproteins: an emerging secretion phenomenon in Gram-negative bacteria. *Trends in Microbiology*, 24(3), 198-208. <https://doi.org/10.1016/j.tim.2015.11.006>
- Wongnate, T., Sucharitakul, J., & Chaiyen, P. (2011). Identification of a catalytic base for sugar oxidation in the pyranose 2-oxidase reaction. *ChemBioChem*, 12(17), 2577-2586. <https://doi.org/10.1002/cbic.201100564>
- Wongnate, T., Surawatanawong, P., Visitsathawong, S., Sucharitakul, J., Scrutton, N. S., & Chaiyen, P. (2014). Proton-coupled electron transfer and adduct configuration are important for C4a-hydroperoxyflavin formation and stabilization in a flavoenzyme. *Journal of the American Chemical Society*, 136(1), 241-253. <https://doi.org/10.1021/ja4088055>

- Wongnate, T., Surawatanawong, P., Chuaboon, L., Lawan, N., & Chaiyen, P. (2019). The mechanism of sugar C–H bond oxidation by a flavoprotein oxidase occurs by a hydride transfer before proton abstraction. *Chemistry–A European Journal*, 25(17), 4460–4471. <https://doi.org/10.1002/chem.201806078>
- Xia, P., Wu, Y., Lian, S., Yan, L., Meng, X., Duan, Q., & Zhu, G. (2021). Research progress on Toll-like receptor signal transduction and its roles in antimicrobial immune responses. *Applied Microbiology and Biotechnology*, 105, 5341–5355. <https://doi.org/10.1007/s00253-021-11406-8>
- Xu, R. Q., Blanvillain, S., Feng, J. X., Jiang, B. L., Li, X. Z., Wei, H. Y., ... & Tang, J. L. (2008). AvrAC_{Xcc8004}, a type III effector with a leucine-rich repeat domain from *Xanthomonas campestris* pathovar *campestris* confers avirulence in vascular tissues of *Arabidopsis thaliana* ecotype Col-0. *Journal of Bacteriology*, 190(1), 343–355. <https://doi.org/10.1128/jb.00978-07>
- Yakhnina, A. A., & Bernhardt, T. G. (2020). The Tol-Pal system is required for peptidoglycan-cleaving enzymes to complete bacterial cell division. *Proceedings of the National Academy of Sciences USA*, 117(12), 6777–6783. <https://doi.org/10.1073/pnas.1919267117>
- Yakushi, T., Masuda, K., Narita, S. I., Matsuyama, S. I., & Tokuda, H. (2000). A new ABC transporter mediating the detachment of lipid-modified proteins from membranes. *Nature Cell Biology*, 2(4), 212–218. <https://doi.org/10.1038/35008635>
- Yamazaki, A., Hirata, H., & Tsuyumu, S. (2008). HrpG regulates type II secretory proteins in *Xanthomonas axonopodis* pv. *citri*. *Journal of General Plant Pathology*, 74, 138–150. <https://doi.org/10.1007/s10327-008-0075-7>
- Yang, Y., Itoh, T., Yokobori, S. I., Itahashi, S., Shimada, H., Satoh, K., ... & Yamagishi, A. (2009). *Deinococcus aerius* sp. nov., isolated from the high atmosphere. *International Journal of Systematic and Evolutionary Microbiology*, 59(8), 1862–1866. <https://doi.org/10.1099/ijs.0.007963-0>
- Yonekura-Sakakibara, K., Higashi, Y., & Nakabayashi, R. (2019). The origin and evolution of plant flavonoid metabolism. *Frontiers in Plant Science*, 943. <https://doi.org/10.3389/fpls.2019.00943>
- Yu, O., & Jez, J. M. (2008). Nature's assembly line: biosynthesis of simple phenylpropanoids and polyketides. *The Plant Journal*, 54(4), 750–762. <https://doi.org/10.1111/j.1365-3113X.2008.03436.x>
- Yu, P., He, X., Baer, M., Beirinckx, S., Tian, T., Moya, Y. A., ... & Hochholdinger, F. (2021). Plant flavones enrich rhizosphere *Oxalobacteraceae* to improve maize performance under nitrogen deprivation. *Nature Plants*, 7(4), 481–499. <https://doi.org/10.1038/s41477-021-00897-y>
- Zhang, J., & Zhou, J. M. (2010). Plant immunity triggered by microbial molecular signatures. *Molecular Plant*, 3(5), 783–793. <https://doi.org/10.1093/mp/ssq035>
- Zhang, Y. L., Li, S., Jiang, D. H., Kong, L. C., Zhang, P. H., & Xu, J. D. (2013). Antifungal activities of metabolites produced by a termite-associated *Streptomyces canus* BYBo2. *Journal of Agricultural and Food Chemistry*, 61(7), 1521–1524. <https://doi.org/10.1021/jf305210u>

96 REFERENCES

- Zimaro, T., Thomas, L., Marondedze, C., Garavaglia, B. S., Gehring, C., Ottado, J., & Gottig, N. (2013). Insights into *Xanthomonas axonopodis* pv. *citri* biofilm through proteomics. *BMC Microbiology*, 13, 1-14. <https://doi.org/10.1186/1471-2180-13-186>