

Function and distribution of protein mono- glycosylations in bacteria

DISSERTATION

Zur Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

der Fakultät für Biologie
der Ludwig-Maximilians-Universität München
vorgelegt von

Franziska Maria Teresa Koller

München, Februar 2021

Gutachter:

1. Prof. Dr. Jürgen Lassak
2. Prof. Dr. Christof Osman

Datum der Abgabe: 23.02.2021

Datum der mündlichen Prüfung: 28.04.2021

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbstständig und ohne unerlaubte Hilfe angefertigt wurde. Des Weiteren erkläre ich, dass ich nicht anderweitig ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen. Die folgende Dissertation liegt weder ganz, noch in wesentlichen Teilen einer anderen Prüfungskommission vor.

München, 23.02.2021

Franziska Koller

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources. As well I declare, that I have not submitted a dissertation without success and not passed the oral exam. The present dissertation (neither the entire dissertation nor parts) has not been presented to another examination board.

Munich, 23.02.2021

Franziska Koller

Contents

Eidesstattliche Erklärung	III
Statutory Declaration	III
Nomenclature	VI
Abbreviation	VI
Publications and manuscripts originating from this thesis	VIII
Contributions to publications and manuscripts presented in this thesis.....	IX
Zusammenfassung	XIII
Summary.....	XV
1 Introduction	1
1.1 Glycobiology.....	1
1.2 Post-translational protein glycosylation.....	2
1.2.1 Eukaryotic and bacterial protein glycosylation processes	3
1.2.2 Protein rhamnosylation.....	7
1.3 Glycation	10
1.3.1 Enzymatic deglycation of Amadori products	11
2 Two RmlC paralogs catalyse dTDP-4-keto-6-deoxy-D-glucose epimerization in <i>Pseudomonas putida</i> KT2440	14
3 A set of rhamnosylation-specific antibodies enables detection of novel protein glycosylations in bacteria.....	33
4 A β -hairpin epitope as novel structural requirement for protein arginine rhamnosylation.....	40
5 Analysis of substrate specificity of the rhamnosyltransferase EarP reveals an unprecedented mode of target recognition in bacterial N-glycosylation	49
6 Transcriptional regulation of the <i>N</i> -fructoselysine metabolism in <i>Escherichia coli</i> by global and substrate-specific cues	67
7 Concluding discussion and outlook	84
7.1 Alternative TDP-Rha biosynthesis gene	84
7.2 Distribution of rhamnosylation	85
7.3 Rhamnose containing glycans as target for therapeutics.....	86
7.4 Donor substrate promiscuity of EarP	88

7.5	Acceptor substrate promiscuity of EarP	89
7.6	EarP as synthetic glycosyltransferase	92
7.7	Co-translational glycosylation using ϵ -Frk as non-canonical amino acid.....	94
7.8	Outlook.....	96
	References for chapters 1 and 7	97
	Curriculum Vitae.....	106
	Danksagung	108

Nomenclature

Amino acids identifiers are given in the three-letter code followed by the position within the primary sequence (for example Arg32). The numbering of amino acids within a protein starts with “1” at the first methionine/valine of the wild-type protein.

Antibodies are abbreviated with “anti-“ followed by the antigen (for example anti-Arg^{Rha}).

Abbreviation

AAH	AIDA-associated heptosyltransferase
AGE	advanced glycated end products
CAP	catabolite activator protein
CMP	cytidine monophosphate
EarP	EF-P arginine 32 rhamnosyltransferase essential for post-translational activation
EF-P	elongation factor P
eIF5A	eukaryotic translation initiation factor 5A
EPEC	enteropathogenic <i>E. coli</i>
EPS	exopolysaccharides
ER	endoplasmic reticulum
F-Asn	fructoseasparagine
Frk	fructoselysine
Frk-6P	fructoselysine-6-phosphate
Gal	galactose
GalNAc	<i>N</i> -acetylgalactoseamine
GDP	guanosine diphosphate
Glu	glucose
GT	glycosyltransferase

LLO	lipid-linked oligosaccharide
LPS	lipopolysaccharide
Man ₅ GlcNAc ₂	mannopentaose-di-(<i>N</i> -acetyl-D-glucosamine)
ncAA	non-canonical amino acid
OST	oligosaccharyltransferase
PTM	post-translational modification
PyIRS	pyrrolysyl-tRNA synthetase
Rha	rhamnose
RBPs	rhamnose binding proteins
PyIRS/tRNA _{CUA}	pyrrolysine tRNA synthetase-tRNA pair
TDP	thymidine diphosphate
TMP	thymidine monophosphate
TTP	thymidine triphosphate
TNFR1	tumor necrosis factor receptor 1
UDP	uridine diphosphate
UTP	uridine triphosphate
UDP-DATDH	UDP-2,4-diacetamido-2,4,6-trideoxyhexose
UDP-GATDH	UDP-2-glyceramido 4-acetamido-2,4,6-trideoxyhexose
UndP	undecaprenyl pyrophosphate

Publications and manuscripts originating from this thesis

Chapter 2:

Manuscript in preparation

Franziska Koller and Jürgen Lassak. Two RmlC paralogs catalyse dTDP-4-keto-6-deoxy-D-glucose epimerization in *Pseudomonas putida* KT2440.

Chapter 3:

Daniel Gast*, Franziska Koller*, Ralph Krafczyk, Lukas Bauer, Svetlana Wunder, Jürgen Lassak and Anja Hoffmann-Röder (2020). A set of rhamnosylation-specific antibodies enables detection of novel protein glycosylations in bacteria. *Org Biomol Chem.* 2020; 18(35): 6823-6828.

Chapter 4:

Liubov Yakovlieva, Thomas M. Wood, Johan Kemmink, Ioli Kotsogianni, Franziska Koller, Jürgen Lassak, Nathaniel I. Martin and Marthe T. C. Walvoort (2020). A β -hairpin epitope as novel structural requirement for protein arginine rhamnosylation. *Chem Sci.* 2021, Advance Article

Chapter 5:

Manuscript in preparation

Ralph Krafczyk*, Franziska Koller*, Laura Lindenthal, Alina Sieber, Bernd Simon, Jannis Brehm, Daniel Gast, Amit Kumar Jha, Jürgen Rohr, Anja Hoffmann-Röder, Janosch Hennig, Kirsten Jung and Jürgen Lassak. Analysis of substrate specificity of the rhamnosyltransferase EarP reveals an unprecedented mode of target recognition in bacterial N-glycosylation.

Chapter 6:

Benedikt Graf von Armanberg*, Franziska Koller*, Nicola Gericke*, Michael Hellwig, Pravin Kumar Ankush Jagtap, Ralf Heermann, Janosch Hennig, Thomas Henle and Jürgen Lassak (2020). Transcriptional regulation of the *N*-fructoselysine metabolism in *Escherichia coli* by global and substrate-specific cues. *Mol Microbiol.* 2020; 00:1–16.

Further publications not included in the dissertation:

Review article:

Jürgen Lassak, Franziska Koller, Ralph Krafczyk and Wolfram Volkwein (2019). Exceptionally versatile – Arginine in bacterial post-translational protein modifications. *Biol Chem.* 2020; 400, 1397-1427.

*Authors contributed equally

Contributions to publications and manuscripts presented in this thesis

Chapter 2:

All experiments were performed by Franziska Koller. Franziska Koller and Jürgen Lassak designed the study. The manuscript was written by Franziska Koller and Jürgen Lassak.

Chapter 3:

Daniel Gast performed the organic synthesis and NMR studies. Lukas Bauer and Swetlana Wunder contributed to the organic synthesis. Franziska Koller performed the confirmation of antibody specificity and sensitivity and Western Blot analysis using the set of generated antibodies. Ralph Krafczyk assisted in the confirmation of antibody specificity and sensitivity. Jürgen Lassak and Anja Hoffmann-Röder designed the study. The manuscript was written by Daniel Gast, Franziska Koller, Jürgen Lassak, and Anja Hoffmann-Röder.

Chapter 4:

Liubov Yakovlieva, Thomas M. Wood, Johan Kemmink, and Ioli Kotsogianni performed the protein purification, peptide synthesis, *in vitro* rhamnosylation and NMR studies. Franziska Koller performed the dot blot affinity assay. Jürgen Lassak contributed to study design. Nathaniel I. Martin and Marthe T. C. Walvoort designed the study. The manuscript was written by all authors.

Chapter 5:

Ralph Krafczyk and Franziska Koller performed protein purification, *in vitro* rhamnosylation assay, and western blot analysis. Alina Sieber assisted in acceptor substrate specificity studies. Laura Lindenthal, Bernd Simon, Janosch Hennig, Jannis Brehm, Daniel Gast, and Hoffmann-Röder were involved in donor substrate specificity studies. Amit Kumar Jha and Jürgen Rohr synthesized TDP-Phenol. Kirsten Jung contributed to study design. Ralph Krafczyk and Jürgen Lassak designed the study. The manuscript was written by Ralph Krafczyk, Franziska Koller, and Jürgen Lassak.

Chapter 6:

Biochemical and genetic analyses on FrIR/FraR/NagR were conducted by Benedikt Graf von Armansperg, Franziska Koller, and Nicola Gericke. NMR studies were performed by Pravin Kumar, Ankush Jagtap, and Janosch Hennig. The corresponding proteins were produced and purified by Benedikt Graf von Armansperg, Franziska Koller, and Nicola Gericke. Michael Hellwig and Thomas Henle synthesized ARPs and analyzed ϵ/α -FrK turnover. Jürgen Lassak designed the study. The manuscript was written by Benedikt Graf von Armansperg, Franziska

Koller, Nicola Gericke, Pravin Kumar Ankush Jagtap, Janosch Hennig, Michael Hellwig, and Jürgen Lassak.

We hereby confirm the above statements:



Franziska Koller

Jürgen Lassak

Chapter 3:

Daniel Gast conducted the chemical synthesis and characterisation of rhamnosyl amino acid building blocks and glycopeptides and prepared the BSA-conjugates. These were further used for the antibody production and later for antibody characterization. Franziska Koller performed determination of BSA-Loading and characterized antibody specificity and sensitivity. In addition, Franziska Koller separated bacterial lysates after growth and performed detection of rhamnosylated proteins on Western Blots using the antibody set. Together, Daniel Gast and Franziska Koller generated sensitive and specific antibodies against mono-rhamnosylated proteins.

We hereby confirm the above statements:



Daniel Gast

Franziska Koller

Chapter 5:

Ralph Krafczyk constructed and purified EF-P variants. The functionality of these variants was tested by Ralph Krafczyk in β -galactosidase assays and Western Blot analysis. Further, Ralph Krafczyk performed the overexpression of EarP in *Escherichia coli* with subsequent Western Blot analysis. Franziska Koller performed the purification of wild-type EF-P and EarP proteins. Franziska Koller applied these proteins in *in vitro* rhamnosylation assays testing different donor substrates. Subsequent Western Blot analysis and calculation of K_i values revealed binding affinities of the tested substrates. Taken together, Ralph Krafczyk and Franziska Koller could

X

determine the minimal recognition motif of EarP on the acceptor substrate site. Further, they were able to show substrate promiscuity of EarP on the donor substrate site.

We hereby confirm the above statements:



Ralph Krafczyk

Franziska Koller

Chapter 6:

Benedikt Graf von Armansperg performed the *in vivo* analysis of the *fri*ABCD promoter region by construction of promoter truncations. Benedikt Graf von Armansperg generated the biotin-labeled DNA fragment which was used to analyse repressor (FrlR) binding to the *fri*ABCD operon by surface plasmon resonance spectroscopy. In addition, Benedikt Graf von Armansperg performed the size exclusion chromatography of purified FrlR UTR domain. Franziska Koller contributed to *fri*ABCD operon (*fri*O) characterization investigating the role of catabolite repressor and adenylate cyclase. Franziska Koller constructed the *fri*O-dependent sfGFP reporter and analysed the binding of FrlR to the operator using microscopy. The *in vivo* self-interaction analysis of FrlR was analysed by Franziska Koller in a bacterial two-hybrid assay. Nicola Gericke performed a multiple sequence alignment of FrlR/FraR/NagR proteins and investigated the role of FrlR on fructoselysine degradation *in vivo*. Nicola Gericke generated all FrlR mutants and tested their ability to bind to prevent fructoselysine metabolism using luciferase reporter assay.

We hereby confirm the above statements:



Benedikt Graf von Armansperg

Franziska Koller



Nicola Gericke

Zusammenfassung

Die Modifikation von Proteinen mit Kohlenhydraten ist in allen Domänen des Lebens verbreitet und an einer Vielzahl von zellulären Prozessen beteiligt. Die bakterielle Glykosylierung wird meist im Zusammenhang mit Pathogenität und Virulenz beschrieben. Kürzlich wurde gezeigt, dass eine N-glykosidische Verknüpfung einer einzelnen L-Rhamnose-Einheit mit Arginin die Funktion des zytosolischen Translationelongationsfaktors P (EF-P) kontrolliert. EF-P war das erste Beispiel für ein zytosolisches bakterielles N-Glykoprotein. Das wirft die Frage auf, wie verbreitet die Protein-Monoglykosylierung in Bakterien ist. Angesichts des großen Interesses an glykosylierten Wirkstoffen ist es auch wichtig, die Substratspezifität der EF-P-Rhamnosyltransferase EarP zu verstehen. Dadurch kann EarP als potenzielles Werkzeug für synthetische Proteinglykosylierung etablieren werden.

Um die Verbreitung der Protein-Monorhamnosylierung zu untersuchen, generierten wir neue Rhamnose-spezifische Antikörper für proteomische Studien. Diese Anti-Ser^{Rha}-, Anti-Thr^{Rha}- und Anti-Asn^{Rha}-Antikörper wurden unter Verwendung synthetischer Glykopeptide hergestellt und führten zum Nachweis von Protein-Monorhamnosylierung in verschiedenen Mikroorganismen, darunter sowohl membrangebundene als auch zytosolische Proteine. Parallel dazu analysierten wir den weit verbreiteten Biosyntheseweg des Nucleotidzuckers dTDP- β -L-Rhamnose (TDP-Rha), der das Donorsubstrat bei der Protein-Rhamnosylierung darstellt. Unter Verwendung einer Genbibliothek und eines bioinformatischen Ansatzes haben wir nach alternativen Biosynthesegenen gesucht und zwei Paraloge der dTDP-4-Keto-6-Desoxy-D-Glucose-Epimerase in *Pseudomonas putida* identifiziert, die beide den vorletzten von vier Schritten der TDP-Rha-Biosynthese katalysieren. Zusammengenommen sind diese Daten von großem Interesse für die Bekämpfung bakterieller Infektionskrankheiten, insbesondere im Hinblick auf ihren Einfluss auf die Pathogenität.

Wir untersuchten die Möglichkeit, synthetische Proteinglykosylierung mit Hilfe der Rhamnosyltransferase EarP durchzuführen, indem wir die Spezifität des Enzyms in Bezug auf seine Akzeptor- und Donorsubstrate testeten. Der Akzeptor wurde als strukturelles Motiv identifiziert, das eine Sequenzvariation erlaubt. Das Motiv wurde sowohl *in vitro*, bestehend aus einem zyklischen Peptid, als auch *in vivo*, wobei wir die Kernakzeptorstruktur künstlich an Modellproteine wie mCherry fusionierten, effizient modifiziert. Damit wurde die gerichtete Modifikation von nicht-kognitiven Zielen demonstriert. Untersuchung der Donorsubstratspezifität ergab, dass EarP Substrate binden kann, die strukturell TDP-Rha ähnlich sind, insbesondere mit Variationen im Nucleotidteil. Diesbezüglich deuteten Inhibitorstudien darauf hin, dass das Enzym in der Lage sein sollte, UDP-aktivierte Zucker für den Transfer zu nutzen. Zusammengenommen legen unsere Analysen die Grundlage, EarP

zu einer Glykosynthese für die synthetische Proteinglykosylierung im Zytosol prokaryotischer Zellen umzubauen.

Die synthetische Glykosylierung mit Hilfe von Glykosyltransferasen ist nur eine Möglichkeit, Glykoproteine in einem bakteriellen Expressionssystem zu erzeugen. Ein weiterer Ansatz ist die co-translationale Insertion durch Ambersuppression mit dem orthogonalen Paar von tRNA^{CUA} und tRNA^{CUA}-Synthetase. Dieser Ansatz wird jedoch durch die gehemmte Aufnahme von glykosylierten Aminosäuren behindert. Eine Ausnahme ist Fructose-Lysin (FrK), das von *Escherichia coli* katabolisiert wird. Da die molekularen Details nicht vollständig aufgeklärt sind, untersuchten wir den Abbau von und die Regulation des Metabolismus von FrK. In diesem Zusammenhang identifizierten wir einen FrK-P spezifischen Repressor FrIR, der im Zusammenspiel mit dem alternativen Sigmafaktor RpoH und dem Katabolit-Aktivatorprotein CRP die Expression des *frlABCD*-Operons vermittelt. Neben den FrK abbauenden Enzymen FrIBCD ist die Permease FrIA kodiert, die vermutlich für die Aufnahme des Aminozuckers verantwortlich ist. Die Charakterisierung der Transportfähigkeiten von FrIA könnte letztlich dazu beitragen die Barriere zu überwinden, die derzeit die Evolution der tRNA-Synthetasen in Richtung glykosylierter Aminosäuren verhindert.

Summary

Protein modification with carbohydrates is common in all domains of life and involved in a variety of cellular processes. Bacterial glycosylation is mostly described in the context of pathogenicity and virulence. Recently, a N-glycosidic linkage of a single L-rhamnose moiety to arginine was shown to control the function of the cytosolic translation elongation factor P (EF-P). It was the first example of a cytosolic bacterial N-linked glycoprotein raising the question how common protein mono-glycosylation in bacteria might be. Given the great interest in glycosylated biologicals, it is also important to understand the substrate specificity of the EF-P rhamnosyltransferase EarP as potential tool in glycoengineering.

To investigate the distribution of protein mono rhamnosylation we generated new rhamnose-specific antibodies for proteomic studies. These anti-Ser^{Rha}, anti-Thr^{Rha} and anti-Asn^{Rha} antibodies were produced using synthetic glycopeptides and led to the detection of protein monorhamnosylation in various microorganisms including both membrane-bound and cytosolic proteins. In parallel, we analysed the widely distributed biosynthesis pathway of the nucleotide sugar dTDP-β-L-rhamnose (TDP-Rha), which is the donor substrate in protein rhamnosylation. Using a genomic library and a bioinformatical approach, we screened for alternative biosynthesis genes and identified two paralogs of the dTDP-4-keto-6-deoxy-D-glucose epimerase in *Pseudomonas putida*, both catalysing the third of four steps in TDP-Rha biosynthesis. Together, these data are of great interest for combating bacterial infectious diseases especially when looking at their impact on pathogenicity.

The glycoengineering capabilities of the rhamnosyltransferase EarP were investigated by testing the enzyme's specificity with respect to its acceptor and donor substrates. The acceptor was identified as structural strand-loop-strand motif allowing for sequence variation. Efficient modification of this motif was shown *in vitro*, constituted by a cyclic peptide, and *in vivo*, where we artificially fused the core acceptor structure to model proteins such as mCherry, thus demonstrating directed modification of non-cognate targets. Investigation of the donor substrate specificity revealed that EarP can bind substrates structurally similar to TDP-Rha, especially with variations in the nucleotide moiety. In this regard, inhibitor studies indicated that the enzyme should be capable of utilizing UDP-activated sugars for transfer. Together, our analyses lay the basis to convert EarP into a glycosynthase for synthetic protein glycosylation in the cytosol of prokaryotic cells.

Glycoengineering with the help of glycosyltransferases is only one possibility to generate glycoproteins in a bacterial expression system. Another approach is the co-translational insertion by amber suppression with an orthogonal pair of tRNA^{CUA} and tRNA^{CUA}-synthetase.

However, this approach is hampered by impaired uptake of glycosylated amino acids. A literature screen revealed the exception of fructose lysine (FrK) which is catabolized by *Escherichia coli*. Since the molecular details are not fully understood, we assessed the degradation of and regulation towards FrK. In this regard, we identified a FrK-P specific repressor FrIR which mediates the expression of the *frIABCD* operon in interplay with the alternative sigma factor RpoH and the catabolite activator protein CRP. Beside the FrK degrading enzymes FrIABCD we recognized a permease FrIA which is presumably responsible for the uptake of the amino sugar. Characterizing the transport capabilities of FrIA will ultimately help to overcome the barrier that currently prevents the evolution of tRNA synthetases towards glycosylated amino acid.

1 Introduction

1.1 Glycobiology

Saccharides (also called carbohydrates, glycans or sugar) are widely distributed in nature and have numerous functions in living organism. The science of glycobiology investigates the structure, biosynthesis, biology, and evolution of saccharides (1). During the last decades, the field of glycobiology was growing rapidly (2, 3). Initially, saccharides were considered more as a source of energy or as structural materials (for example cellulose). Today, it is known that saccharides play an important role in diverse processes within the cell making them relevant to biomedicine, biotechnology, and basic research.

Monosaccharides are the simplest form of carbohydrates with a carbonyl group at the end of the carbon chain (aldehyde group) or at an inner carbon (ketone group) (1). Oligosaccharides are linear or branched chains of monosaccharides which are linked *via* glycosidic linkages. These linkages are formed *via* hydroxyl groups of the saccharide. When more than one hydroxyl group within one saccharide forms a glycosidic linkage, branched glycans arise (4). A glycan attached to a macromolecule is called a glycoconjugate (for example glycoproteins). The stereochemistry of a given glycosidic linkage can have a severe impact on structural properties and biological functions. However, glycans and glycoconjugates are not encoded by the genome as it is the case for protein sequences. They are secondary gene products and are typically assembled and attached to proteins and lipids by enzymes, a process called glycosylation. Nevertheless, non-enzymatic and hence spontaneous and uncontrolled reactions are known and described by glycation.

Together with the great diversity of monosaccharides in nature, this results in a huge complexity of possible glycan structures (5). By this, glycans contribute to the enormous biological complexities with simultaneously relatively small numbers of genes in genomes (1, 3). In line with this, there is a huge amount of genes that encode for proteins that are involved in glycan biosynthesis, degradation, or transport (for instance around two percent of the human genome) (6). Therefore, glycans have been postulated to be one of the fundamental macromolecular components defined so far as nucleic acids (including DNA and RNA), proteins, and lipids (7) as they seem to be essential to the existence of all known living organisms (8).

To this day, great progress has been made in elucidating the evolution, formation, and function of different glycans in a variety of organisms. Particularly noteworthy are the latest developments in which glycans, glycoconjugates, and their corresponding enzymes become targets for next-generation therapeutics, vaccines, and diagnostics (9). In addition to that, the

field of synthetic glycobiology expands where glycosylation systems are redesigned to obtain defined glycosylation structures on proteins for diverse applications in medicine, materials, and diagnostics (10). Interestingly, bacterial and archaeal glycosylation seems to be far more diverse than eukaryotic (11). Hence, it is postulated that microorganisms play a key role in glycobiology (12).

In this thesis, different aspects of bacterial protein glycobiology will be investigated. We will examine the distribution, the substrate specificity, and biosynthesis of the sugar donor substrate of a specific glycosylation process called protein rhamnosylation. Further, we will investigate the catabolism of glycosylated amino acids arising from glycation.

1.2 Post-translational protein glycosylation

Post-translational modifications (PTMs) are alterations of proteins occurring after protein biosynthesis at the ribosome. It is a ubiquitous paradigm for dynamic cellular response and signalling found in all domains of life (13). 15 out of the 20 common proteinogenic amino acids can be modified by the transfer of a chemical group (14). This can influence the maturation or enhance the capabilities of proteins (15). The most commonly found PTMs are methylation, acylation, phosphorylation as well as glycosylation.

Protein glycosylation is the process whereby a glycan moiety is enzymatically transferred from an activated donor to an acceptor substrate (amino acid side chain within a polypeptide). In most cases, the glycan is covalently attached to the amid nitrogen atom of asparagine (Asn) residues (N-glycosylation) or the hydroxyl oxygen of serine (Ser) or threonine (Thr) residues (O-glycosylation) (16). The transfer is catalysed by glycosyltransferases (GTs). Based on the protein fold, these enzymes can be classified into four structural families: GT-A, GT-B, GT-C, and lysozyme-type (17). Most of the structurally analysed GTs exhibit either GT-A or GT-B fold (Fig. 1). These two folds can also be attributed to Leloir-type GT as they use nucleotide sugars as donor substrates. Both folds consist of two associated $\beta/\alpha/\beta$ domains. In GT-A fold enzymes (Fig. 1A), these two domains are closely associated and form a central open twisted β -sheet. The GT-B fold (Fig. 1B) is characterized by Rossmann-like $\beta/\alpha/\beta$ domains, but these are associated less tightly (18). The glycosyltransferase reaction can have two different outcomes: Inverted or retained conformation of the anomeric carbon atom of the saccharide. This stereochemical outcome is not determined by the overall GT fold as both structural families, GT-A and GT-B, comprise inverting and retaining members (18).

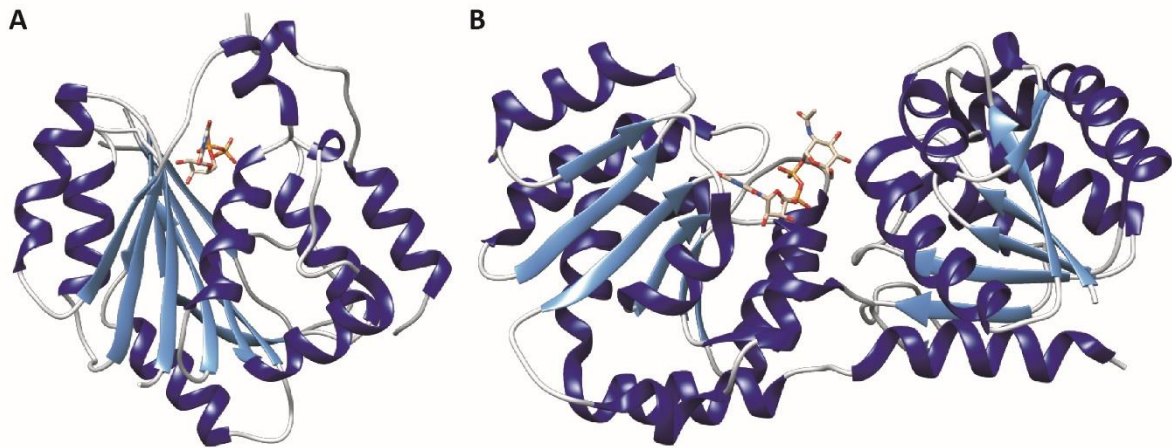


Figure 1: **Structural overview of GT-A and GT-B glycosyltransferase folds.** Ribbon diagrams of representative **A**) GT-A (SpsA from *Bacillus subtilis*, PDB: 1qgq) and **B**) GT-B fold (MurG from *Escherichia coli*, PDB: 1nlm) glycosyltransferases (helix: dark blue, Strand: light blue, coil: grey). The bound ligands are depicted as sticks (carbon: grey, phosphate: orange, oxygen: red, nitrogen: blue). Illustrations were generated with UCSF Chimera (19).

According to the Carbohydrate Active Enzymes database (20) (CAZy, <http://www.cazy.org/>, specialist database providing genomic, structural and biochemical information on carbohydrate-active enzymes), 1-5 % of the predicted coding sequences of free-living organisms correspond to carbohydrate-active enzymes. The genomic distribution of GTs seems to be widespread in nature as archaeal, bacterial, and eukaryotic genomes show this ratio.

1.2.1 Eukaryotic and bacterial protein glycosylation processes

In eukaryotes, protein glycosylation was reported for the first time in 1938 (21, 22). Today we know that more than half of the eukaryotic proteins are glycosylated and that 90 % of these are N-glycosylated (23). Concomitantly with their folding, proteins are N-glycosylated in the endoplasmic reticulum (ER) (Fig. 2A). A generic polysaccharide (Mannopentaose-di-(N-acetyl-D-glucosamine), $\text{Man}_5\text{GlcNAc}_2$) is transferred *en bloc*, mediated from a lipid carrier called dolichol-pyrophosphate-oligosaccharide which is located in the membranes of the ER. An oligosaccharyltransferase (OST) complex catalyses the transfer to an Asn residue that is part of a consensus peptide recognition sequence (sequon) Asn-X-Ser/Thr (where X represents any amino acid except proline) (24, 25). STT3 is the catalytic component of this nine-member OST complex (26). Several glycosidases and glycosyltransferases process the N-glycan, determining whether the polypeptide will be retained in the ER, transported along the secretory pathway, or dislocated across the ER membrane for degradation (27). N-glycans assist the protein folding process and serve as a bar code that provides information on age and folding

status of the glycoproteins. This quality check prevents exit to the Golgi-apparatus of folding intermediates or misfolded glycoproteins (28).

Eukaryotic O-glycosylation is less conserved but generally initiated in the Golgi-apparatus (partially in the ER) and starts with the attachment of a linking monosaccharide to a Ser or Thr residue (Fig. 2D) (29). The linking sugar is typically α -N-acetyl-D-galactosamine (GalNAc) (mucin-type) but others are described as well (30, 31). It is a sequential transfer, where the glycan is extended to form one of four common core structures (29). These core structures can subsequently be extended and result in diverse mature linear or branched O-glycans (32). The glycan is attached to tandem repeat domains of mucins without a specific recognition sequon by GalNAc transferases (29).

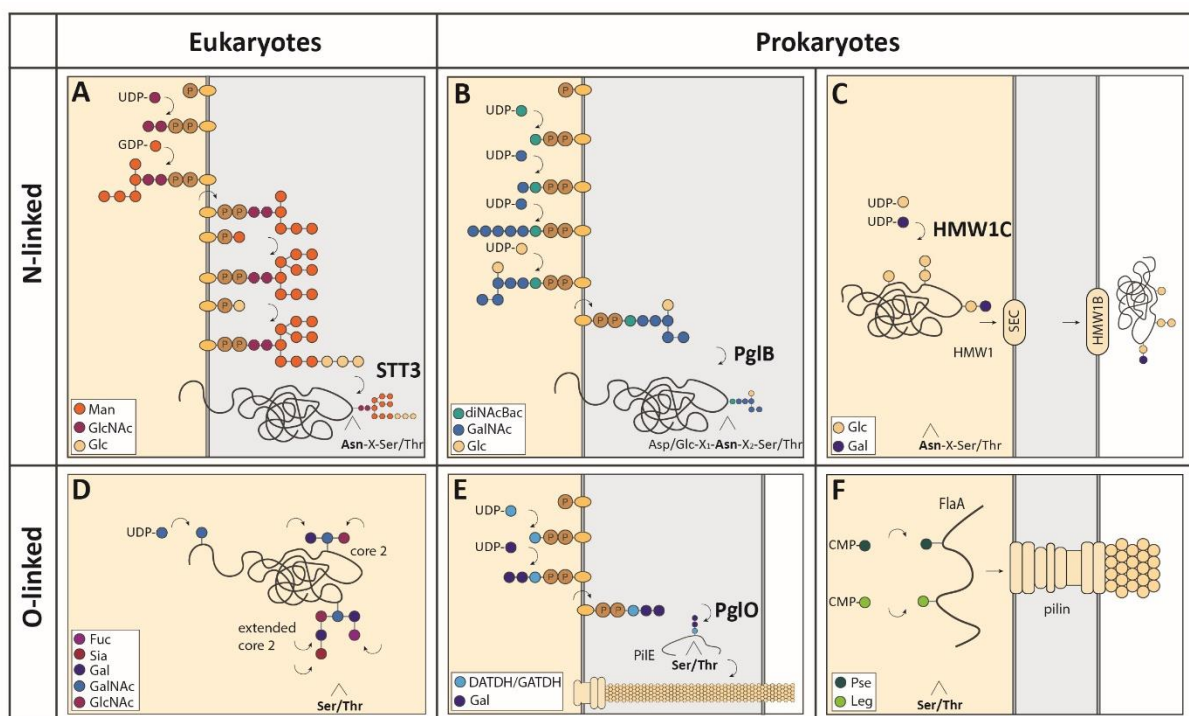


Figure 2: **Protein glycosylation processes in eukaryotes and bacteria.** Overview of N-linked (A/B/C) and O-linked glycosylation (D/E/F) strategies. **A)** In eukaryotic **N-linked** glycosylation, a generic polysaccharide (Man₅GlcNAc₂) is generated through sequential addition of nucleotide-activated sugars onto a lipid carrier in the endoplasmic reticulum. The glycan is flipped across the membrane, further processed to a 14-mer, and transferred **en bloc** to the growing polypeptide by an OST complex including the catalytic component STT3. **B)** Bacterial OST-mediated **N-linked** glycosylation **en bloc** in *Campylobacter jejuni* is similar to the eukaryotic apart from further processing of the glycan after flipping. It takes place at the periplasm and the transfer is mediated by PglB. **C)** *Haemophilus influenzae* **sequential N-linked** glycosylation takes place in the cytosol. HMW1C transfers sugar moieties from nucleotide-activated sequentially to distinct sites of HMW1. Glycosylated HMW1 is shuttled into the periplasm and further to the outer membrane *via* the SEC pathway an HMW1B, respectively. **D)** Eukaryotic **O-linked glycosylation** starts with the attachment of a linking GalNAc and is extended **sequentially** to common core structures (here core 2). Core structures are further processed, the whole process takes place in the Golgi-apparatus. **E)** Bacterial OST-mediated **O-linked** glycosylation starts with the sequential addition of nucleotide-activated sugars onto a lipid carrier. In *Neisseria gonorrhoeae*, a trisaccharide is assembled, flipped into the periplasm and transferred **en bloc** by PglO to the pilin polypeptide PilE. **F)** Bacterial **sequential O-linked** glycosylation takes presumably place at the cytoplasm–inner membrane interface. In *C. jejuni*, CMP-activated sugar moieties are transferred to flagellin subunit FlaA. Target amino acids and consensus sequences are depicted beneath the target site. Man: mannose, GlcNAc: N-acetylglucosamine, Glc: glucose, diNAcBac: UDP-N,N-diacetylbaucillosamine, GalNAc: N-acetylgalactosamine, Gal: galactose, Fuc: fucose, Sia: sialic acid, DATDH/GATDH: diacetamido-2,4,6-trideoxyhexose/ glyceramido 4-acetamido-2,4,6-trideoxyhexose, Pse: 5,7-

diamino-3,5,7,9-tetradexynon-2-ulosonic acid, Leg: legionaminic acid. This figure was adapted and extended from Szymanski *et al.*(33) and Nothaft *et al.* (16).

It has long been assumed that protein glycosylation is restricted to eukaryotes (16). But in 1975, the first O-glycosylation in bacteria was described: The cell surface S-layers (34, 35). Until today, different bacterial O-glycans were found and O-glycosylation in bacteria can be divided into two processes: Sequential (Fig. 2F) and block transfer (Fig. 2E) (16).

The sequential transfer is similar to the eukaryotic O-glycosylation, where nucleotide-activated sugars are added individually to surface exposed Ser or Thr residues (Fig. 2F). The process is assumed to occur at the cytoplasm–inner membrane interface (16). A well-studied example for this mechanism is the flagellar protein of *Campylobacter jejuni*. Here, CMP-activated sugars are attached to the flagellin subunit FlaA by the glycosyltransferase GTase. The glycosylation of the flagellar proteins is important for their assembly (16, 33).

Studies on *Neisseria* and *Pseudomonas* species have revealed a new mechanism of O-glycosylation: The block transfer (Fig. 2F) (36-39). Here, preassembly of sugars onto lipid carriers was observed which are then transferred *en bloc* to Ser or Thr residues by an OST. This mechanism seems to occur especially in gram-negative bacteria. In *N. gonorrhoeae* for example, UDP-GlcNAc is converted into uridine diphosphate (UDP)-2,4-diacetamido-2,4,6-trideoxyhexose (UDP-DATDH) or UDP-2-glyceramido 4-acetamido-2,4,6-trideoxyhexose (UDP-GATDH) and transferred onto the lipid carrier undecaprenyl pyrophosphate (UndP) (involved enzymes: pilin glycosylation D (PglD), PglC and PglB/PglB2, respectively). Further GTs (PglA and PglE) add Gal residues to complete the UndPP-linked trisaccharide. Afterwards, the lipid-linked oligosaccharide (LLO) is flipped into the periplasm (PglF) and attached to a Ser residue of the pilin polypeptide PilE by the OST PglO (11, 40, 41).

Bacterial N-glycosylation was first described in 2002 in *C. jejuni* (42). OST-mediated glycosylation in *C. jejuni* is considered as a prototype for bacterial *en bloc* N-glycosylation (Fig. 2B). Starting from UDP-GlcNAc, UDP-diNAcBac is formed by the enzymes PglF, PglE, and PglD. The sugar residue is transferred to an UndP carrier by PglC. The LLO is subsequently elongated by the action of PglA, PglJ, and PglH and flipped into the periplasm by PglK. In the last step, the glycan is transferred *en bloc* by the OST PglB to an Asn residue present in the protein consensus sequence Asp/Glc-X1-Asn-X2-Ser/Thr (X1 and X2 represent any amino acid except proline) (16, 40, 43).

Besides the transfer *en bloc*, another type of N-glycosylation was discovered in bacteria: the sequential transfer in *Haemophilus influenzae* (Fig. 2C) (44, 45). The high-molecular-weight adhesin 1 (HMW1) is glycosylated at multiple Asn residues embedded in the eukaryotic-like sequon Asn-X-Ser/Thr (except one case) (44). The glycosylation takes place in the cytosol, where the GT HMW1C sequentially N-glycosylates HMW1 at 31 distinct sites (46). HMW1C

transfers Gal and Glc from nucleotide-activated donors and is able to mono- and diglycosylate the protein. Glycosylated HMW1 is shuttled into the periplasm *via* the Sec pathway and then tethered to the outer-membrane channel-forming β -barrel translocator protein, HMW1B (45).

Taken together, there are four major glycosylation pathways (Fig. 2). The OST-mediated N-glycosylation *en bloc* (eukaryotes and bacteria), the sequential cytoplasmic N-glycosylation (bacteria), the OST-mediated O-glycosylation *en bloc* (bacteria), and the sequential O-glycosylation (eukaryotes and bacteria) (30).

A major difference between eukaryotic and prokaryotic glycosylation is striking when comparing the sugar moieties found in the glycan structures. In eukaryotes, glycans used in glycosylation are built from nine nucleotide sugar donors: UDP-D-Glu, UDP-D-Gal, UDP-2-*N*-acetyl-D-glucosamine, UDP-2-*N*-acetyl-D-galactosamine, GDP-D-mannose, GDP-L-fucose, UDP-D-glucuronic acid, UDP-dxylose, CMP-*N*-acetyl-neuraminic acid (sialic acid). Eukaryotes also have conserved oligosaccharides which are transferred from LLO to proteins (Man5GlcNAc₂). Diversity in eukaryotic glycan structures mainly derives from variation in the linkages between glycan components, number and type of the saccharide (47). Compared with prokaryotes, this is a surprisingly low number: Prokaryotic glycans contain more than one hundred different saccharides (47-50). Due to this, a vast variation is possible. Here, the oligosaccharide portion is not conserved but in some cases a similar structure can be found (51). In fact, archaea are thereby thought to have the greatest structural variance (51-53).

In general, glycosylation in eukaryotes is involved in many important processes in the cell: proper protein folding, cell adhesion, molecular trafficking and clearance, receptor activation, signal transduction, and endocytosis (54). Bacterial glycosylation is mostly described in the context of pathogenicity and virulence (33, 55). In some bacteria, it has been shown that colonization and virulence decrease significantly when glycosylation null mutants are created (52, 56, 57). Since surface-exposed proteins, which are important for bacterial motility and for colonization such as flagella (58), are very often glycosylated in bacteria, it is obvious that they are a target for therapeutics (11). However, since glycosylated proteins are also described in some non-pathogenic strains, glycosylation in bacteria seems not only to be associated with pathogenicity. It can therefore be assumed that glycosylation plays a far greater role in bacteria than previously supposed.

1.2.2 Protein rhamnosylation

Rhamnose is a deoxyhexose saccharide and belongs to the unusual saccharides found in glycan structures (49, 59). The saccharide is incorporated in different bacterial cell compartments: Lipopolysaccharide (LPS) and exopolysaccharides (EPS) (60, 61), cell walls of gram-positive bacteria (62), and flagella (63, 64). Rhamnose is found in plants as well (65, 66) but is missing in higher eukaryotes (67). The saccharide occurs in two isomers: D and L-rhamnose. The latter is much more common and widely found in bacteria (49).

The activated form of L-rhamnose in bacteria, the nucleotide sugar dTDP- β -L-Rhamnose (TDP-Rha), is used as sugar donor in glycosylation reactions. The biosynthesis pathway of TDP-Rha (RmlBDAC) is ubiquitous and highly conserved in bacteria (Fig. 3A). The transferase RmlA (glucose-1-phosphate thymidyltransferase), the first enzyme of the pathway, transfers a thymidylmonophosphate nucleotide to glucose-1-phosphate which is further oxidized by the dehydrase RmlB (dTDP-D-glucose 4,6-dehydratase) at the C4 hydroxyl group of the saccharide. The next step is an unusual double epimerisation reaction at positions C3 and C5, catalysed by the epimerase RmlC (dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase). The final product TDP-Rha is achieved by reduction of the C4 keto group by the reductase RmlD (dTDP-4-keto-6-deoxy-L-mannose reductase) (67).

Rhamnose moieties are also reported to occur in many pathogenic bacteria. In TDP-Rha deletions strains, it has been shown that the saccharide plays a major role in pathogenicity (68-72). Deletion of *rmlB* or *rmlD* in *Vibrio cholerae* for example led to defects in colonization (73). In *Mycobacteria*, it was even shown that TDP-Rha biosynthesis is essential, as *rmlD* could only be inactivated in *M. smegmatis* in the presence of a rescue plasmid carrying functional *rmlD* (74). Since TDP-Rha or the biosynthesis genes do not occur in humans, it makes it an interesting therapeutic target for targeting the tuberculosis pathogen *M. tuberculosis* (75).

A very unusual post-translational glycosylation using TDP-Rha as substrate was discovered in 2015 at the translation elongation factor P (EF-P) in *Shewanella oneidensis* (76) and subsequently confirmed in *P. aeruginosa* (77). EF-P, which is the ortholog of eukaryotic and archaeal initiation factor 5 (e/alf5A), alleviates ribosomal pausing at polyproline motives (78-82). The glycosylation of EF-P with rhamnose was found in about 10 % of the sequenced bacteria. They all have in common that they code for an arginine (Arg) residue at the tip of the protein (76). In another group (about 25 % of sequenced bacteria) including for example *Escherichia coli*, the post-translational activation of EF-P is ensured by transfer of a (*R*)- β -lysine to a conserved lysine (Lys) residue at the tip of the protein (79, 83). The so called

rhamnosylation of EF-P was an unexpected finding as this was the first description of N-linked protein glycosylation on Arg in bacteria (76).

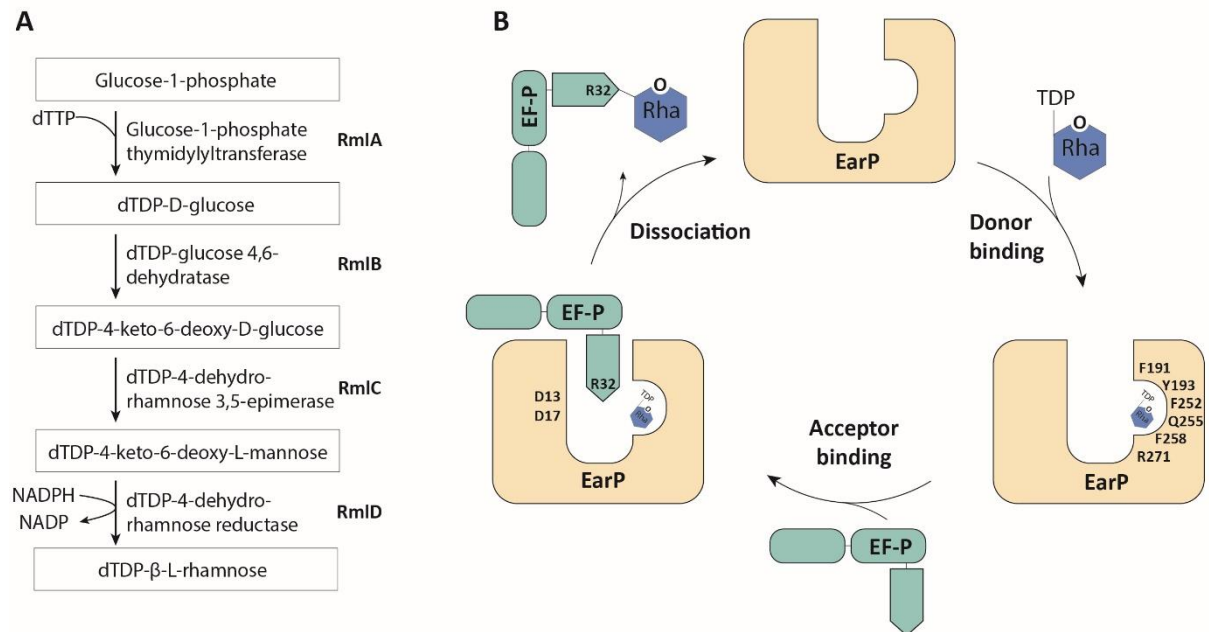


Figure 3: dTDP-β-L-rhamnose biosynthesis and glycosylation mechanism of the rhamnosyltransferase EarP. **A)** TDP-Rha biosynthesis by RmlBDAC. The transferase RmlA transfers a thymidylmonophosphate nucleotide to glucose-1-phosphate, which is further oxidated by the dehydrase RmlB. The next step is a double epimerisation reaction catalysed by the epimerase RmlC. The final product TDP-Rha is achieved by a reduction catalysed by the reductase RmlD. **B)** Proposed glycosylation mechanism of the rhamnosyltransferase EarP from *Pseudomonas putida*. TDP-Rha donor (blue) binding occurs in the C-terminal domain of EarP (yellow) and is mediated by several amino acid residues in the binding pocket. The acceptor KOW-like N-domain of EF-P (green) is bound in the N-terminal domain of EarP. The nucleophilic attack onto the anomeric centre of TDP-Rha is facilitated by two catalytic amino acids by activating the Arg32 (R32) guanidinium of EF-P. Rhamnosylated EF-P is released and EarP is available for another cycle of rhamnosylation. This figure was adapted and extended from Lassak *et al.* (84). D: aspartic acid, F: phenylalanine, Y: tyrosine, Q: glutamine, R: arginine, EF-P: elongation factor P, EarP: EF-P arginine 32 rhamnosyltransferase essential for post-translational activation

N-glycosylation on Arg were first described for sweetcorn and rice, and are not restricted to rhamnose. The results of the corresponding studies indicated auto-glycosylation of amylogenin and UDP-arabinopyranose mutase, respectively (85, 86). However, both datasets lack further studies on mechanism of transfer and activation (84). Finally, Arg glycosylation of tumor necrosis factor receptor 1 (TNFR1) and its downstream adaptors was uncovered in eukaryotes in 2013 (87, 88). The effector glycosyltransferase NleB1 from enteropathogenic *E. coli* (EPEC) is secreted via type III secretion system and transfers a GlcNAc moiety to a conserved Arg on the death domains of TNFR1 (TRADD). This blocks the death domain interactions and the assembly of the oligomeric TNFR1 complex, thereby disrupting TNF signalling in EPEC-infected cells. TNF signalling is crucial for immune homeostasis, cell death, and inflammation and plays a major role in antimicrobial host response. By blocking these mechanisms *via* Arg glycosylation, the pathogen is able to evade the host immune response. Orthologues of NleB1 and the paralog NleB2 (89) were found in *Salmonella enterica*: SseK1, SseK2, and SseK3 (90).

EF-P rhamnosylation is unusual both in the targeted amino acid and in the cellular localization of the resulting glycoprotein. EF-P is modified by the GT EF-P arginine 32 rhamnosyltransferase essential for post-translational activation (EarP) catalysing the transfer of a single sugar moiety from the activated nucleotide sugar TDP-Rha to the highly conserved Arg (Arg32 in *P. putida*) (76, 77, 91). The glycosylation takes place in the cytosol. The thereby activated EF-P protein facilitates the translation of polyproline motifs in the arrested ribosome. The location and type of glycosylation is similar to that of HMW1, with the difference that the glycoprotein is not translocated (76). Hence, EF-P is the only known example of a bacterial cytoplasmic N-linked glycoprotein.

Further crystallographic and biochemical studies revealed details of the glycosylation mechanism (Fig. 3B) (91-94). At first, the donor substrate TDP-Rha is bound and oriented in the C-terminal domain of EarP which consists of two opposing Rossmann-like domains and adopts a GT-B fold (91). The acceptor EF-P is recognized and bound by the N-terminal EarP domain. Two aspartate (Asp) residues of EarP play an important role in facilitating the nucleophilic attack onto the anomeric center of TDP-Rha by activating the relatively inert Arg guanidinium of EF-P (91, 94). The glycosylation reaction results in the formation of α -rhamnosylarginine characterizing EarP as an inverting GT (92). Finally, the glycosylated substrate is released from the active site making EarP available for another cycle of rhamnosylation (91, 93).

1.3 Glycation

Contrary to glycosylation, glycation is a non-enzymatic process. Here, reducing sugars such as glucose and fructose are covalently attached to a free amino or guanidinium group of proteins, lipids and nucleic acids (95-101). The glycosylation process, known as Maillard reaction, can be divided into early and late stages (Fig. 4) (102-104). In the early phase, amino groups react with reducing carbohydrates, yielding Amadori and Heyns products (105, 106). In glycation of proteins, the reaction starts with a nucleophilic attack of the carbonyl group of the sugar by an ϵ -amino group of lysine or a guanidine moiety of arginine. The fast and highly reversible reaction results in a Schiff base which can undergo isomerization (Amadori rearrangement) and results in a ketoamine (Amadori product). Here, the reverse reaction is much slower than the formation leading to an accumulation of Amadori products (107). Further reaction, such as oxidation, fragmentation, condensation, reduction, dehydration, isomerization, and cyclization of the Amadori products is described in the late phase (84, 101, 108). It results in stable, advanced glycation end-products (AGEs). Formation of the highly heterogeneous group of AGEs takes place in weeks or months and is irreversible. Three different categories of AGEs are defined according to their ability either to create cross-links on proteins or/and to show fluorescence: A) fluorescent cross-linking, B) non-fluorescent cross-linking and C) non-cross-linking AGEs (109).

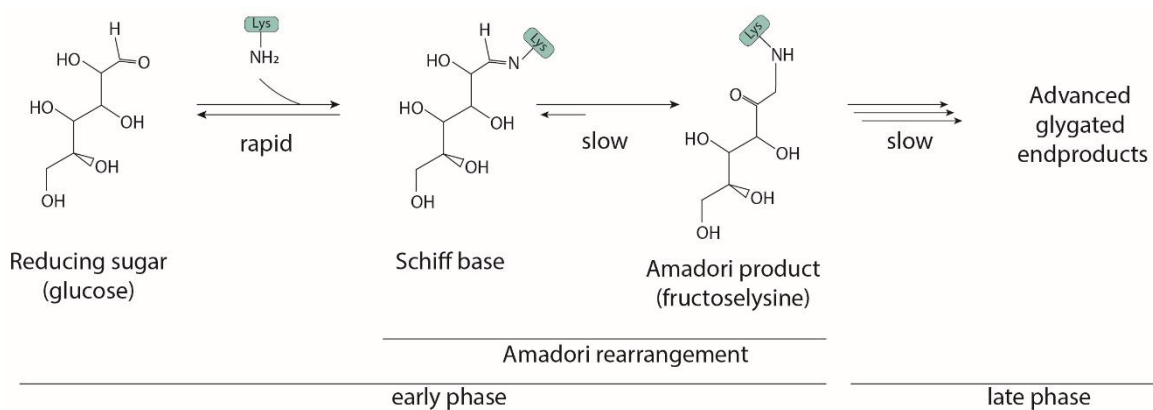


Figure 4: **Maillard reaction: Formation of Amadori products and advanced glycated endproducts.** The Maillard reaction is separated in early and late phase. In the early phase, a reducing sugar such as glucose is covalently attached to free amino or guanidinium group (for example lysine (Lys, green)). This reaction is fast and highly reversible and results in a Schiff base. Amadori rearrangement of the product results in Amadori products such as fructoselysine. In the late phase, further reactions and modifications can lead to the formation of Advanced glycated endproducts. This figure was adapted and extended from Ulrich *et al.* (107).

Diverse sugars differ in their ability to react with amino groups. In general, the reactivity depends on the extent to which the sugar exists in the open (carbonyl), more reactive or in the ring (hemiacetal or hemiketal) structure (110, 111). Fructose for example shows a higher steady state concentration of the reactive open-chain than glucose and is 10-fold more reactive than glucose (112). Additionally, sugars with less carbon atoms are more reactive; pentoses

are more reactive than hexoses, which are more reactive than disaccharides (113). On the acceptor side, a free amino group is necessary for the reaction. Amino groups with low pK_a values are in general more reactive toward glycation because of their high reactivity and nucleophilicity (114). In proteins, only the α -amino group of N-terminal amino acids and those with an amino acid side chain are accessible (Lys, Arg, and histidine (His)) (100). Furthermore, it was shown for Lys that the properties of amino acids in the proximity, for example positive charge (115, 116) and acidity (116), can promote glycation (117, 118).

1.3.1 Enzymatic deglycation of Amadori products

For more than 100 years, the Maillard reaction is known as food browning which typically proceeds rapidly at high temperatures (≥ 140 °C). It is an important source of aroma and color in cooked and processed foods. In the western diet, depending on the nutrient composition, humans consume several milligrams of Maillard products per day (119, 120). During the 1970s and 1980s, it was discovered that a rather slow Maillard reaction also occurs *in vivo* (95, 121-124).

To cope with the amount of Amadori products, many organisms have developed different ways to degrade them. The deglycating enzymes have different substrate specificities and physiological function like intracellular repair functions of DNA as well as proteins or degradation and utilization of Amadori products (125). So far, three different mechanisms for the deglycation of ketoamines are described (Fig. 5): Oxidative degradation by fructosyl amino acid oxidases (Amadoriases) (Fig. 5A) (126), phosphorylation by fructosamines 3-kinases (FN3K) (Fig. 5B) (127), and phosphorylation by fructosamine 6-kinases (FN6K) and subsequent cleavage by deglycase (Fig. 5C) (125).

Amadoriases have been found in fungi, yeast, and bacteria and catalyze oxygen-mediated oxidation of Maillard products (126). In general, they cleave small glycated substrates such as fructosyl amino acids at the ketoamine bond between C1 of the sugar moiety and the nitrogen atom and release amine, glucosone, and H_2O_2 (Fig. 5A) (128). Fructosamine 3-kinases FN3K and related proteins, are found in all taxa and deglycate fructosamines in a two-step reaction (Fig. 5B) (127). By phosphorylation of the third carbon of the sugar moiety, the ketoamine is destabilized. In the following, this leads to the spontaneous formation of 2-keto-3-deoxyaldoses by releasing the amine (129, 130).

The third deglycation mechanism, which is also described as reverse Amadori reaction, is only found in bacteria. Fructosamines are also degraded in a two-step reaction including two enzymes (Fig. 5C) (125). The fructosamine-6-kinases FN6K phosphorylates the sixth carbon

atom of the sugar moiety leading to fructosamine 6-phosphate. Further deglycation leads to the formation of an aldose like glucose 6-phosphate and release of a free amine (131-133). In this way, bacteria like *E. coli* (131), *Salmonella enterica* (134), and *Bacillus subtilis* (132) are able to use fructosamines as sole carbon source. Interestingly, the Amadori metabolism is widely distributed among gut-colonizing bacteria (135). This is in line with the estimation that gut microbiota are exposed to more than 500 mg of Amadori products daily (120, 136).

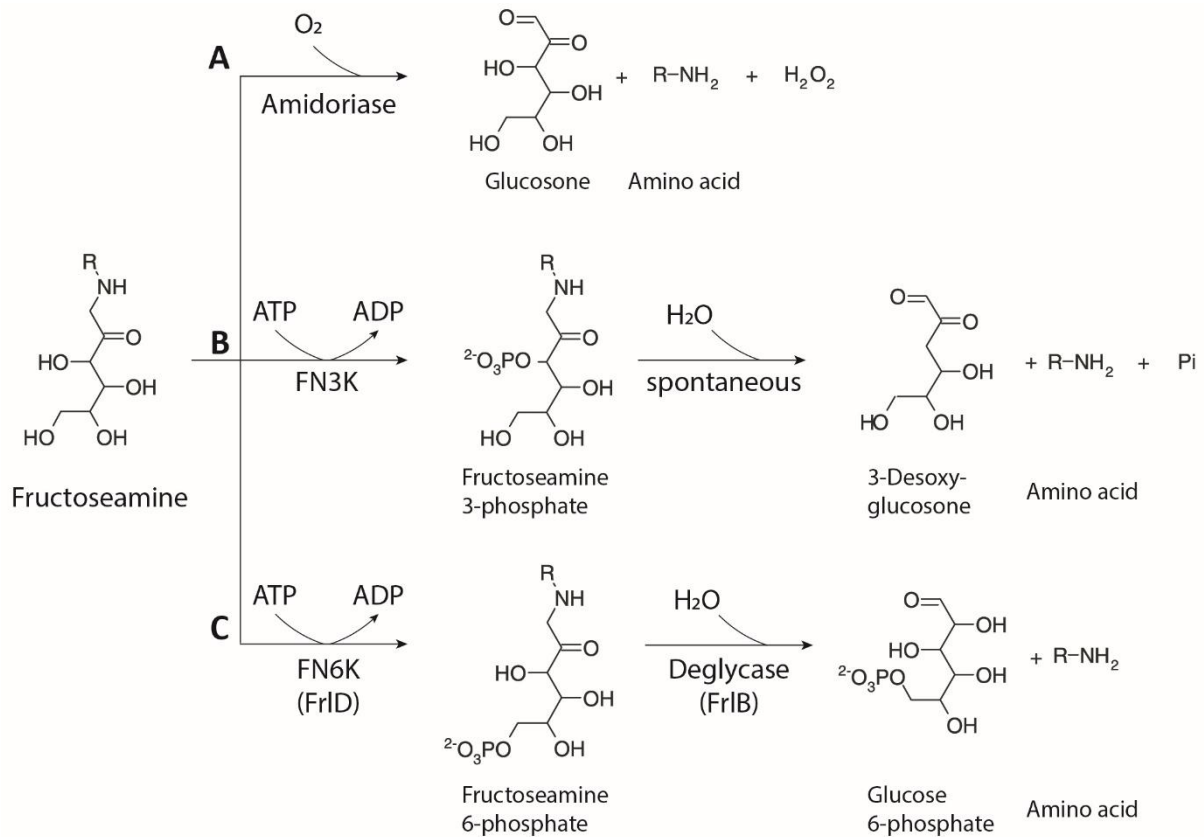


Figure 5: **Fructoseamine degradation.** **A)** Oxidative degradation. Fructosyl amino acid oxidases (Amadoriases) cleave fructoseamines between C1 of the sugar moiety and the nitrogen atom and release amine, glucosone and H_2O_2 **B)** Phosphorylation. Fructoseamine 3-kinases (FN3K) phosphorylate the third carbon of the sugar moiety leading to fructoseamine 3-phosphate. This destabilization leads to the spontaneous formation of 2-keto-3-deoxyaldoses by releasing the amine. **C)** Phosphorylation. Fructoseamine 6-kinases (FN6K) phosphorylate the sixth carbon atom of the sugar moiety leading to fructoseamine 6-phosphate. Deglycation by a deglycase leads to the formation of an aldose like glucose 6-phosphate and release of a free amine. Related *E. coli* proteins are depicted below (FrID and FrIB). This figure was adapted and extended from Deppe *et al.* (125).

The enzymes needed for the Amadori metabolism are encoded by the operon *frlABCD* (*E. coli*) or *frlBONMDR* (*B. subtilis*) or *fraRBDAE* (*S. enterica*). The substrates which are degraded by the different pathways differ: *N*- ϵ -fructoselysine (ϵ -FrK) in *E. coli* (131), *N*- α -fructoselysine (α -FrK) and other glycated amino acids such as *N*- α -fructoseglycine or *N*- α -fructosevaline in *B. subtilis* (132), and *N*- α -fructoseasparagine (F-Asn) in *S. enterica* (134). However, the degradation of these Amadori products follows a conserved route in bacteria and can be depicted by the *E. coli* ϵ -FrK metabolism (131). The uptake of ϵ -FrK is presumably ensured by the putative permease FrIA. The kinase FrID phosphorylates the sixth carbon atom resulting

in fructoselysine-6-phosphate (FrK-6P) which is hydrolysed by the deglycase FrIB into glucose-6-phosphate and lysine. Glucose-6-phosphate is further metabolised in the glycolysis whereas lysine is processed in the amino acid metabolism. The fructose-lysine 3-epimerase FrIC catalyses the reversible conversion of fructose-lysine and psicose-lysine further allowing for the metabolism of this compound in *E. coli* (137). So far, nothing is known about the regulation of the *E. coli* ϵ -FrK metabolism. In the human gut environment, utilization of alternative carbon sources is beneficial as distinct niches are created and the competition is minimized. A tightly regulated uptake of ϵ -FrK and the further utilization as carbon source can be therefore beneficial to compete with other bacteria of the gut microbiota.

2 Two RmlC paralogs catalyse dTDP-4-keto-6-deoxy-D-glucose epimerization in *Pseudomonas putida* KT2440

Manuscript

Franziska Koller¹ and Jürgen Lassak^{1*}

¹Department of Biology I, Microbiology, Ludwig-Maximilians-Universität München, Planegg/Martinsried, Germany

*corresponding author, email: juergen.lassak@lmu.de

Abstract

L-rhamnose is an important monosaccharide both as nutrient source as well as building block in prokaryotic glycoproteins and glycolipids. Generation of those composite molecules demands for activated precursors being provided e.g. in form of nucleotide sugars such as dTDP- β -L-rhamnose (TDP-Rha). TDP-Rha is synthesized in a conserved 4-step reaction which is canonically catalysed by the enzymes RmlABCD. An intact pathway is especially important for the fitness of pseudomonads, as TDP-Rha is essential for the activation the polyproline specific translation elongation factor EF-P in these bacteria. Within the scope of this study, we investigated the TDP-Rha-biosynthesis route of *Pseudomonas putida* KT2440 with a focus on the last two steps. Bioinformatic analysis in combination with a screening approach revealed that epimerization of dTDP-4-keto-6-deoxy-D-glucose to dTDP-4-keto-6-deoxy-L-mannose is catalysed by the two paralogous proteins PP_1782 (RmlC1) and PP_0265 (RmlC2), whereas the reduction to the final product is solely mediated by PP_1784 (RmlD). Thus, we also exclude the distinct RmlD homolog PP_0500 and the genetically linked nucleoside diphosphate-sugar epimerase PP_0501 to be involved in TDP-Rha formation, other than suggested by certain databases. Together our analysis contributes to the molecular understanding how this important nucleotide-sugar is synthesized in pseudomonads.

Key-words: RfbBDAC, glycosylation, PA4068, PA4069, *Pseudomonas aeruginosa*

Introduction

Rhamnose (Rha) is a naturally occurring sugar being widely distributed among bacteria and plants (1). As such it can serve – on the one hand – as sole carbon source for many microorganisms (2). In this regard, regulation towards Rha has turned out as powerful molecular tool to drive gene expression in bacteria (3, 4). On the other hand, Rha is utilized in antibiotic synthesis (5) or forms an integral part of saponins (6), certain bacterial surface glycans such as rhamnolipids (7) or mycolic acids (8), extracellular polysaccharides (9) and even cytosolic proteins (10) (Fig 1A). Incorporation of rhamnose into these requires an activated precursor which is provided as a nucleotide sugar. To date, two forms of activated Rha are known to be produced by bacteria: Guanosine diphosphate- α -D-rhamnose (GDP-Rha) (11) and deoxythymidine- β -L-rhamnose (TDP-Rha) (12). While GDP-Rha is synthesized from mannose-1-phosphate (11), the pathway for TDP-Rha starts with glucose-1-phosphate (Glc-1P) (Fig. 1B).

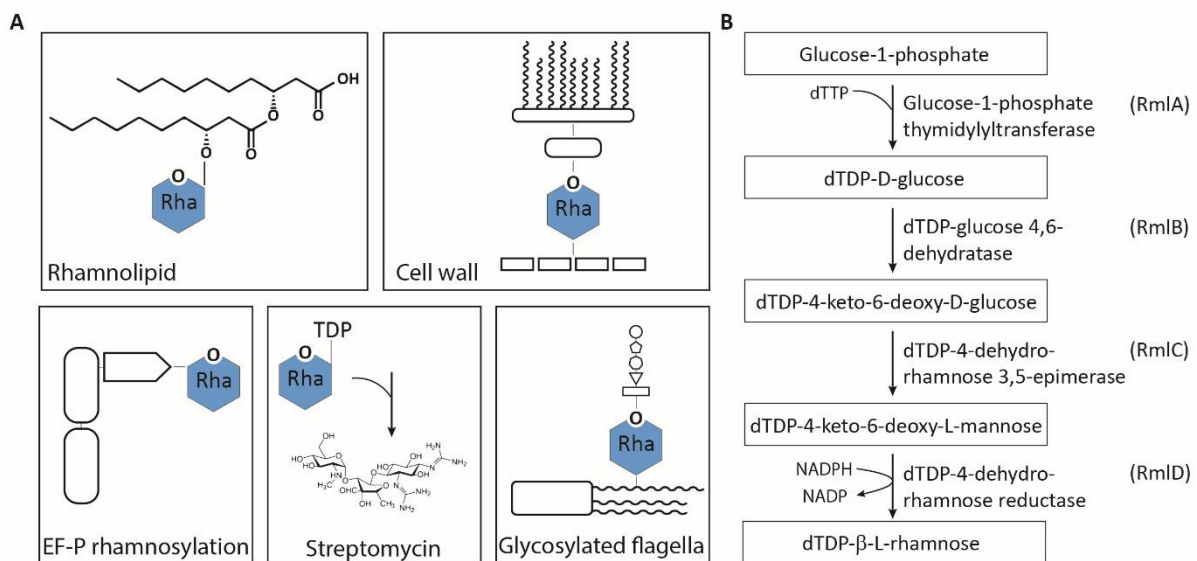


Fig. 1: Rhamnose as versatile building block in composite biomolecules **A)** L-Rha in bacterial biomolecules. Top left: Rhamnolipids consisting of a rhamnose moiety and a fatty acid tail in *P. aeruginosa* (7). Top right: Mycobacterial cell wall containing L-Rha as linking sugar between arabinogalactan and peptidoglycan (7). Bottom left: Rhamnosylation of translation elongation factor EF-P in about 10 % of all bacteria (10). Bottom middle: Biosynthesis of Streptomycin inter alia originating from TDP-Rha (5). Bottom right: Glycosylated flagella with a linking L-Rha moiety in certain *Pseudomonads* (13). **B)** dTDP-β-L-rhamnose biosynthesis pathway. Glucose-1-phosphate thymidyltransferase, the first enzyme of the pathway, transfers a thymidylmonophosphate nucleotide to glucose-1-phosphate, which is further oxidated by dTDP-D-glucose 4,6-dehydratase at the C4 hydroxyl group of the saccharide. The double epimerization reaction at positions C3 and C5 is catalyzed by the dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase. Finally, the reduction of the C4 keto group by the dTDP-4-keto-6-deoxy-L-mannose reductase leads to dTDP-β-L-rhamnose.

Homologs for the synthesis genes of TDP-Rha, *rmIBDAC*, can be identified in gram-positive and gram-negative bacteria (1) and according to their number the pathway consists of four steps (Fig. 1B) (1). First, a nucleotide transferase RmlA (RfbA) (14) or RffG (15) transfers a deoxythymidine monophosphate moiety from deoxythymidine triphosphate to Glc-1P accompanied by the release of pyrophosphate. In the second step, a dehydratase RmlB (RfbB) (16) or RffH (15) catalyzes the conversion of TDP-glucose into TDP-4-keto-6-deoxy-D-glucose. The third enzyme – an isomerase RmlC (RfbC) (17) – mediates a double epimerization reaction leading to the formation of TDP-4-keto-6-deoxy-L-mannose. Fourth, RmlD (RfbD) (17) reduces the C4 keto group of the 4-keto-6-deoxy-L-mannose and with this TDP-Rha synthesis is completed. Notably, the pathway was shown to be critical or even essential for viability in the human pathogens *Streptococcus pyogenes*, *S. mutans* (18) and *Mycobacterium tuberculosis* (19). In the clinically relevant *Pseudomonas aeruginosa* (20), TDP-Rha is important for the synthesis of rhamnolipids (21). These are bacterial surfactants with a rhamnose moiety as head group and act as a key virulence determinant (22). Moreover, in about 10 % of all bacteria including pseudomonads, a protein monorhamnosylation was described in 2015 which is essential for activation of the polyproline specific translation elongation factor EF-P (10). Specifically, the glycosyl transferase EarP transfers a rhamnose moiety onto a conserved EF-P arginine residue R32 thereby utilizing TDP-Rha as a precursor (10, 23-25). Loss of either *efp* or *earP* reduces bacterial fitness dramatically (10, 26, 27) while

deletion of *rmIC* surprisingly leads to a milder mutant phenotype in *Shewanella oneidensis* (10) and *P. aeruginosa* (25). This in turn suggests for a compensatory mechanism. One plausible explanation is the presence of a yet unidentified gene substituting for *rmIC* and being widely distributed in the γ -proteobacterial orders of Alteromonadales and Pseudomonadales. We chose *P. putida* KT2440 to test our hypothesis as it is probably the best-characterized non-pathogenic laboratory pseudomonad (28). *P. putida* strains in general are fast-growing and genetically easily accessible (29). They are a paradigm of metabolically versatile microorganisms being able to recycle organic wastes and are key players in the maintenance of environmental quality (29).

Following an unbiased approach and utilizing a restriction based genomic library, we identified the two paralogous proteins PP_1782 (now termed RmIC1) and PP_0265 (now termed RmIC2) as dTDP-4-dehydrorhamnose 3,5-epimerases while the last step namely the reduction to TDP-Rha seems to be solely catalyzed by PP_1784 (RmID). By contrast, two further candidate genes that were identified by database mining and homology analyses – PP_0500 and PP_0501 – are not involved in TDP-Rha biosynthesis. Taken together, our findings contribute to the molecular understanding how TDP-Rha is synthesized in pseudomonads. This is important as the RmlABCD has been postulated as target for new antibacterial therapeutics (8, 18, 30)

Results

A screening system that allows for the discovery of TDP-Rha synthesis genes

To identify genes involved in TDP-Rha biosynthesis, we took advantage of the pathway's dependence on activation of pseudomonal EF-P and its cross functionality in *Escherichia coli*. This cannot necessarily be expected, as the *E. coli* endogenous EF-P significantly differs from its pseudomonal counterpart (31): Although both proteins alleviate ribosome stalling at polyproline stretches (10, 32), their modes of activation are phylogenetically unrelated (10, 33). While *E. coli* EF-P (EF-P_{Eco}) strictly depends on (R)- β -lysylation (24, 34-36) and hydroxylation (37) of a conserved lysine *Pseudomonas* EF-P (EF-P_{Ppu}) is rhamnosylated at arginine by the glycosyltransferase EarP (EarP_{Ppu}) at the structurally equivalent position (10, 23). Despite these apparent distinct post-translational modifications, a combination of *efp*_{Ppu} and *earP*_{Ppu} from *P. putida* KT2440 can compensate for a lack of *efp* in *E. coli* (Δ *efp*) as long as the endogenous TDP-Rha pathway remains intact (Fig. 2A, C) (10). Interestingly, loss of any synthesis gene – here exemplified with a Δ *rmIC* strain – does not simply phenocopy Δ *efp* but even results in more severe growth defects, as can be concluded from the corresponding doubling times (Fig. 2B). In fact, *E. coli* Δ *efp* cross complemented with *efp/earP*_{Ppu} grows twice

as fast as the same strain additionally lacking *rmIC* ($\Delta efp \Delta rmIC$). These differences in growth rates provide us with a selection regime to identify dTDP-4-dehydrohamnose-3,5-epimerase genes from a *P. putida* genomic library.

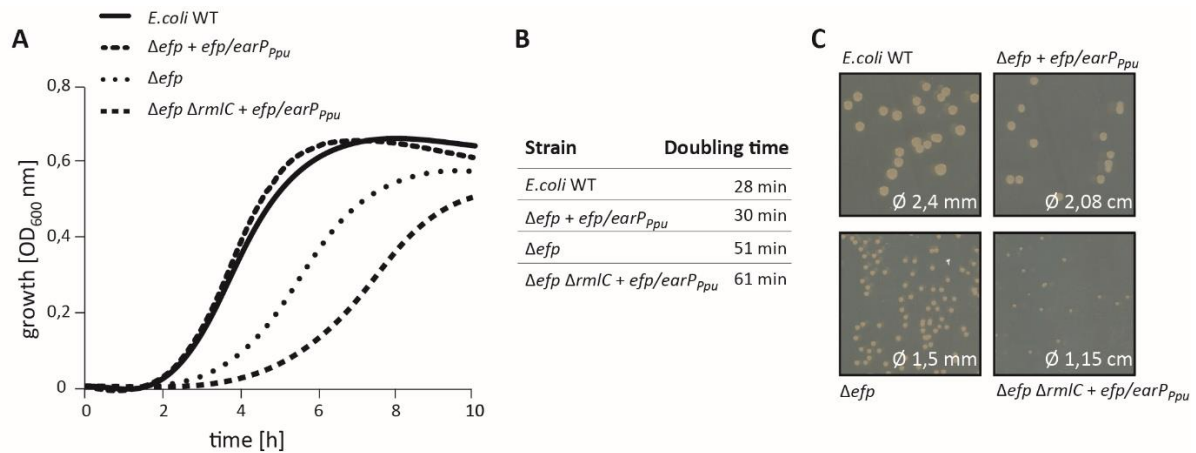


Fig. 2: **Growth analysis of cross complemented *E. coli* Δefp mutants in dependence of the TDP-Rha pathway.** **A:** *E. coli* MG1655 (*E. coli* WT), *E. coli* MG1655 Δefp expressing pBBR MCS2 *efp/earP_{ppu}* ($\Delta efp efp/earP_{ppu}$), *E. coli* MG1655 Δefp (Δefp) and *E. coli* MG1655 $\Delta efp \Delta rmIC$ expressing pBBR MCS2 *efp/earP_{ppu}* ($\Delta efp \Delta rmIC efp/earP_{ppu}$) were grown in LB at 37 °C. Shown is a the mean curve from three independent biological replicates. **B:** Doubling times of *E. coli* strains listed in A. Doubling times were calculated from growth analysis from three independent biological replicates. **C:** Comparison of colony size. The same strains listed in A were plated on LB Agar (1.5 %). Pictures were taken after o/n growth at 37 °C.

The library was constructed by partial restriction digestion of the *P. putida* genome with the *dam* and CpG methylation insensitive enzyme *StuI* (NEB) (Fig. 3). The average fragment size was set to 5 kb to ensure that at least one gene was completely covered (average gene size: 1.132 kbp). These were cloned into *SmaI* linearized pBAD33, which allows for high-level expression by induction of the *P_{BAD}* promoter with L-arabinose (38). Transformation of *E. coli* DH10B cells with the library revealed ~430.000 clones indicating a for an around 350fold coverage of the *P. putida* KT2440 genome (total length 6.18187 Mbp).

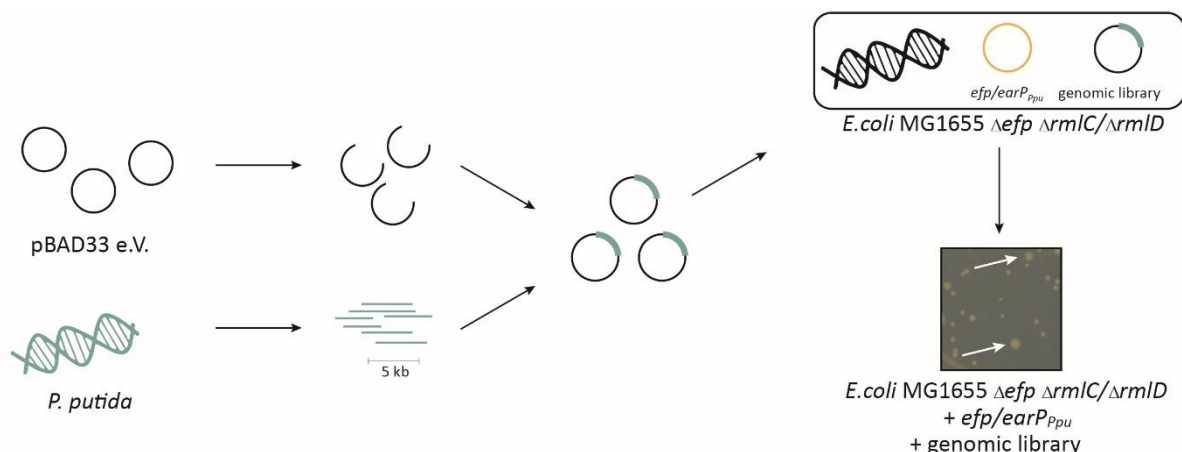


Fig.3: **Screening strategy for the identification TDP-Rha biosynthesis genes in *P. putida*.** Chromosomal DNA of *P. putida* (green) was fragmented by restriction digestion. Fragments with an average size of 5 kb were then ligated into the arabinose

inducible vector (pBAD33). The resulting library was transformed into an *E. coli* Δefp $P_{cadBA}::lacZ$ reporter strain that concomitantly lacks either *rmIC* or *rmID* ($\Delta rmIC/\Delta rmID$) and additionally encodes *earP_{Ppu}* and *efp_{Ppu}* (orange) of *P. putida* in trans. Genes cross-complementing $\Delta rmIC$ and $\Delta rmID$ recover the impaired growth phenotype and can be selected by size (white arrows).

Next, we transformed *E. coli* Δefp $\Delta rmIC$ + *efp/earP_{Ppu}* with the library and cultivated the cells in LB containing 0.2 % L-arabinose. Considering duplication times (Fig. 2B) and genome coverage, we expect *rmIC* copies to accumulate already to a single-digit percentage of the total population within latest two days (= ~16 generations with mutant growth phenotype and ~32 for wild-type phenotype), even under unfavorable circumstances. Indeed, when plating the second overnight culture on LB agar we obtained colonies of two different sizes. Consequently, we isolated plasmids from 16 large clones and sequencing identified 12 times PP_1782 and four times PP_0265 as the insert. PP_0265 (from now on *rmIC2/RmIC2*) resides next to genes encoding a putative two component signal-transduction system (Fig. 4A). PP_1782 (from now on *rmIC1/RmIC1*) on the other hand is the last of four genes in a putative TDP-rhamnose biosynthesis operon PP_1785-PP_1782 (Fig. 4B). To substantiate our hypothesis on the TDP-Rha biosynthetic operon, we conducted a second library screen with *E. coli* cells now lacking *rmID* in addition to *efp* (Δefp $\Delta rmID$ + *efp/earP_{Ppu}*) instead of *rmIC*. With this strain we exclusively enriched clones harboring a copy of PP_1784 (from now on *rmID/RmID*), a homolog of *E. coli* RmID. Thus, we provide experimental evidence that PP_1785-PP_1782 form a *rmIBDAC1* operon in *P. putida* KT2440 and further identified a second gene encoding for an dTDP-4-dehydrorhamnose 3,5-epimerases – RmIC2.

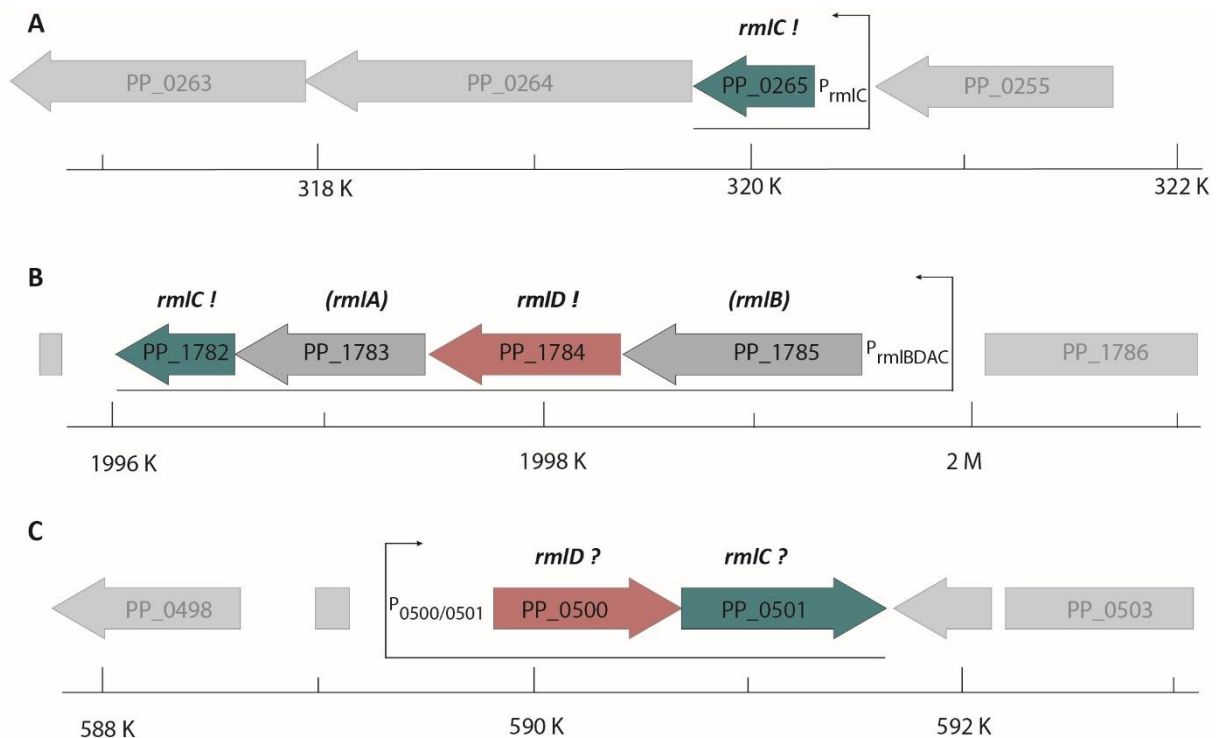


Fig. 4: **Genomic organization of *rmIC* and *rmID* candidate genes in *P. putida*.** **A)** PP_0265 gene region **B)** PP_1782-PP_1784 gene region. **C)** PP_0500 and PP_0501 gene region. Putative (?) or validated (!) homologs/analogs of *rmIC* and *rmID* are shown

in green and red respectively. Bottom: Position within *P. putida* genome. Arrows indicate monocistrons. The scale indicates the position within the *P. putida* genome.

PP_0265/PP_1782 and PP_1784 are dTDP-4-dehydrorhamnose 3,5-epimerases and dTDP-4-dehydrorhamnose reductase, respectively

Our library screen was complemented by database mining and a homology search. In addition to *rmIC1* and *rmIC2* we found PP_0501 being annotated as nucleoside diphosphate- sugar epimerase of unknown specificity and as such might function as further dTDP-4-dehydrorhamnose 3,5-epimerase (String (39, 40), Pfam (41), Uniprot (42), Metacyc (43) database). However, while RmIC1 & RmIC2 are highly homologous to each other (64 % identity), PP_0501 shares no similarities at the sequence level (Fig. S1). Nonetheless and in addition to its annotated function PP_0501 forms an operon with a putative dTDP-4-dehydrorhamnose reductase gene, PP_0500 (39, 40, 44) (Fig. 4C). This protein, on the contrary, shares similarities with RmID both at the sequence level (29 % identity) as well as structurally (Fig. S2).

To test the putative role of PP_0500 and PP_0501 in TDP-Rha biosynthesis we made again benefit of EarP mediated activation of *P. putida* EF-P and its functionality in *E. coli*. Hence, we cloned the two genes into pBAD33 simultaneously adding a His₆-tag coding sequence for immunodetection in order to ensure proper protein production (Fig. 5). *rmIC1*, *rmIC2* and *rmID* were also included in the study. The resulting plasmids pBAD33-*rmIC1*, pBAD33-*rmIC2*, pBAD33-PP_0501 as well as pBAD33-*rmID* and pBAD33-PP_0500 were introduced into *E. coli* $\Delta efp \Delta rmIC + efp/earP_{Ppu}$ and $\Delta efp \Delta rmID + efp/earP_{Ppu}$, respectively. Of note, these are reporter strains in which EF-P functionality is coupled to LacZ expression (Fig. 5A). Whereas β -galactosidase activity is low in cell with an incomplete TDP-Rha biosynthesis pathway, introduction of either *rmIC1*, *rmIC2* (Fig. 5B) or *rmID* (Fig. 5C) into the respective mutant strains led to a significant increase. By contrast, neither PP_0500 nor PP_0501 were able to rescue the Δefp_{Eco} mutant phenotype.

In parallel we analyzed the rhamnosylation status of EF-P_{Ppu} utilizing anti-rhamnosylarginine specific antibodies (anti-Arg^{Rha}) (23, 45, 46). Immunodetection of EF-P_{Ppu} rhamnosylation matched with the reporter expression levels on the one hand confirming the enzymatic activities of RmIC1, RmIC2 and RmID as dTDP-4-dehydrorhamnose 3,5-epimerase and dTDP-4-dehydrorhamnose reductase, respectively. On the other hand, they falsify speculation and database annotations that attribute PP_0500 and PP_0501 a function in TDP-Rha biosynthesis (String (39, 40), Pfam (41), Uniprot (42), Metacyc (43) database).

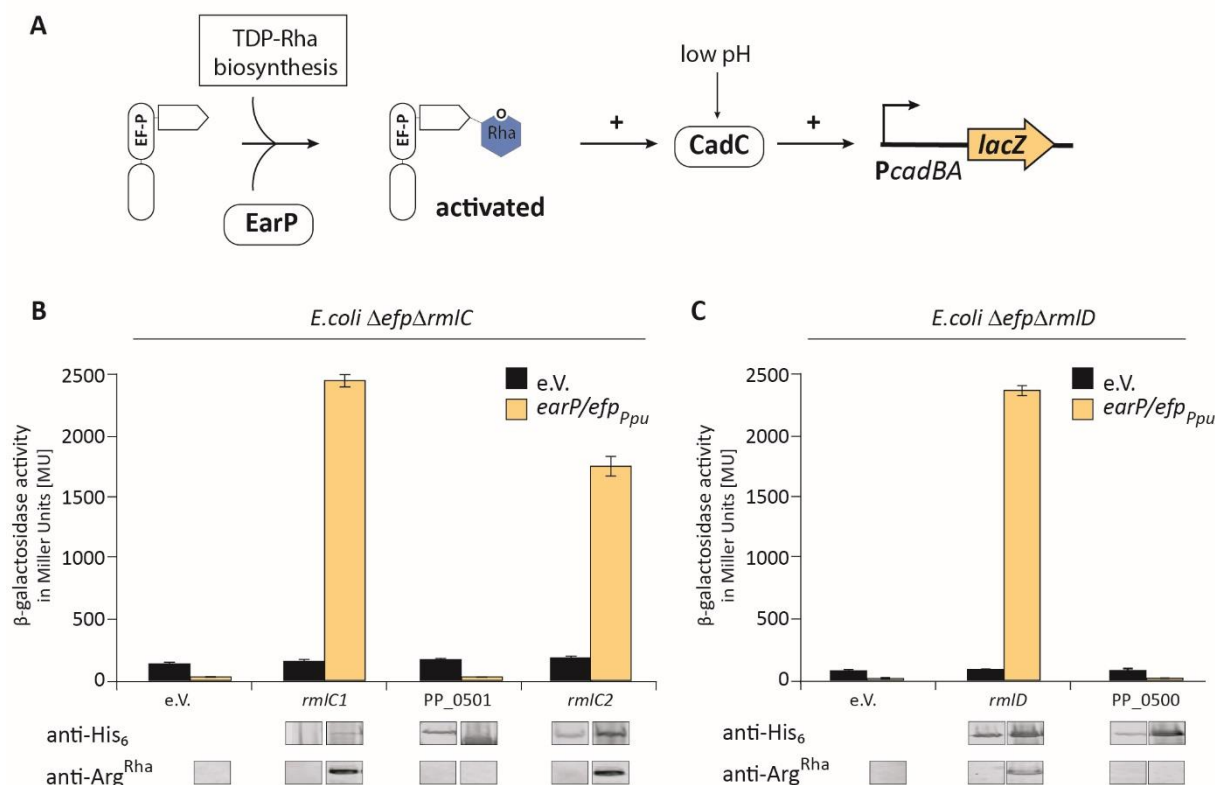


Fig. 5: Analysis of *in vivo* activities of activated EF-P in TDP-Rha biosynthesis deletion strains. **A:** β -Galactosidase reporter assay. The assay is based on the lysine decarboxylase acid stress response of *E. coli*, the CadABC module (47). At low pH, the transcriptional activator CadC activates the promoter of its two downstream genes (P_{cadBA}) thereby inducing the expression of *lacZ* in an *E. coli* MG1655 $P_{cadBA}::lacZ$ strain. Proper translation of CadC is dependent on the presence of EF-P which is activated by mono-rhamnosylation, a reaction catalysed by the glycosyltransferase EarP using dTDP- β -L-rhamnose (TDP-Rha, blue) as substrate. Thus β -galactosidase activity can be taken as an indirect readout for functional TDP-Rha biosynthesis. **B and C:** Functionalities of RmlC1, RmlC2, RmlD, PP_0500 and PP_0501 were determined by measuring the β -galactosidase activities of *E. coli* MG1655 $P_{cadBA}::lacZ \Delta$ efp Δ rmIC (**B**)/ Δ rmID (**C**) with heterologous expression of a candidate gene from the pBAD33 vector. The empty vector (e.V.) was included as negative control. Additionally, all strains encoded the *earP/efp_{Ppu}* operon in trans, being encoded from pBBR MCS2 vectors (grey bars) expressed from the native promoter. Again, the corresponding empty vector served as control (black bars). All strains were grown o/n in LB pH 5.8 and activity is given in Miller Units (MU). Means of three independent measurements are shown. Standard deviations from three independent experiments were determined. Bottom: Western blot analysis of o/n cultures *E. coli* depicted in B and C. Rhamnosylated EF- P_{Ppu} (EF- P^{Rha}) was detected using 0.25 μ g/ml anti-Arg^{Rha}. Expression of candidate genes was verified using 0.1 μ g/ml anti-His₆. Full-length Western Blots and corresponding SDS-gels are depicted in Fig.S3.

Discussion

In the scope of this study, we have investigated the TDP-rhamnose pathway of *P. putida* KT2440 with a focus on the isomerization of dTDP-4-keto-6-deoxy-D-glucose. Combining an unbiased approach and utilizing a genomic library, we identified two paralogous proteins RmlC1 and RmlC2. Duplication of *rmlC* is not restricted to *P. putida* but certain other pseudomonads such as *P. monteilii*, *P. fulva*, *P. plecoglossicida* or *P. asiatica* harbor also two gene copies. In fact, functional redundancy in the TDP-Rha biosynthesis pathway is nothing

unusual. As an example, the two enzymes RffG and RffH of *E. coli* catalyze the same two steps as RmlA and RmlB, respectively (15). Such duplications may be useful, e.g., to compensate for bottleneck reactions in the TDP-Rha biosynthesis (48). Such bottlenecks can occur at different stages as the pathway is not only utilized to ultimately generate TDP-Rha. Specifically, dTDP-4-keto-6-deoxy-D-glucose is also a precursor of dTDP-3-acetamido- α -D-fucose (49) and TDP-D-viosamine (50) which are found as part of the glycan pattern in *P. syringae* (51). Similarly, the two paralogs RmlC1 and RmlC2 in *P. putida* KT2440 might serve as starting point of similar but so far unknown reactions. Moreover, gene duplications open the gate for regulated expression in turn allowing the precise adjustment of the desired ratio of distinct NDP-sugars depending on parts of the TDP-Rha biosynthesis pathway. It would also allow for the accumulation of educts or products of the preceding reactions such as dTDP-glucose and Glc-1P. Notably, whereas *rmlC1* is part of an operon in which presumably the full TDP-Rha pathway is encoded, the *rmlC2* resides in the vicinity of two genes encoding a two-component system (TCS) of thus far unknown function. Based on the predicated domain composition, this specific TCS presumably transduces external signals into gene transcription. One might therefore speculate on regulated expression of *rmlC2* according to the environmental conditions.

While our study revealed two RmlC paralogs in *P. putida* database mining in the genome of both *P. aeruginosa* and *S. oneidensis* did not lead to the identification of gene duplicates of *rmlC*. In the latter organism, however, there might be two enzymes catalyzing the reduction to TDP-Rha (SO_1653 und SO_4174). Similarly, *P. aeruginosa* might encode two *rmlD* copies – PA5162 and PA4069. Whereas PA5162 shares 60 % identity with PP_1784 (RmlD), PA4069 is rather an ortholog of PP_0501 (81 % identity). As we have excluded PP_0501 as dTDP-4-dehydrorhamnose reductase it is also unlikely, that PA4069 is involved in TDP-Rha biosynthesis. Similarly, the adjacent PA4068, which is orthologous to PP_0501 is presumably not a player in the pathway. Notably, our structural model (Fig. S1) revealed the UDP-N-acetylglucosamine C4-epimerase PelX from *P. protegens* Pf-5 as close homolog (identity 67 %) (52). PelX is involved in the biosynthesis of the GalNAc-rich bacterial polysaccharidepolysaccharide Pel, that is essential for pellicle biofilm formation (52, 53). One can hence hypothesize, that PP_0500 and PP_0501 might be involved in that pathway, instead.

One question remains: How can the discrepancy in the growth phenotypes between the $\Delta earP$ strains in *S. oneidensis* and *P. aeruginosa* and their corresponding deletions of *rmlC* be explained (10, 25). One possibility might be that there is a yet undiscovered dTDP-4-dehydrorhamnose 3,5-epimerase, which is either not as widely distributed as initially assumed or with low enzymatic activity and thus hindering its identification with our experimental setup.

Alternatively, one could hypothesize on a “non TDP-Rha” NDP-Deoxyhexose pathway: Beside variations in the nucleotide moiety as in the case of GDP-Rha (11), TDP-linked epimeric sugars such as pneumose have been described (54). If these exist in pseudomonads and at the same time the glycosyltransferase EarP recognizes such NDP-sugar as donor substrate it might lead to a modified and/or partially activated EF-P. Promiscuity of EarP has yet been described with respect only to the protein acceptor (55, 56). Nevertheless, based on the available structural data it is plausible, that EarP’s specificity is also not restricted TDP-Rha. In fact, the architecture of its catalytic pocket might tolerate off-substrate binding especially when they deviate in the sugar moiety (23, 57, 58). While both the GDP-moiety and the distinct sugar configuration are good arguments to exclude GDP- α -D-rhamnose, other NDP-sugars such as TDP- β -L-pneumose might be suitable substitutes for TDP-Rha. Due to the close relatedness, this could even work at the ribosome and assist in polyproline induced translational arrest. Whether or not a pathway for the synthesis of TDP- β -L-pneumose does exist remains, however, elusive and further studies are needed to test this hypothesis.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft research grant LA 3658/1-1 and Research Training Group GRK2062/1 (Molecular Principles of Synthetic Biology). We thank Ralph Krafczyk and Kirsten Jung for fruitful discussions. Further we thank Lis Winter, Maximilian Dorok and Martin Erhard for their contributing results.

Supplement Figures

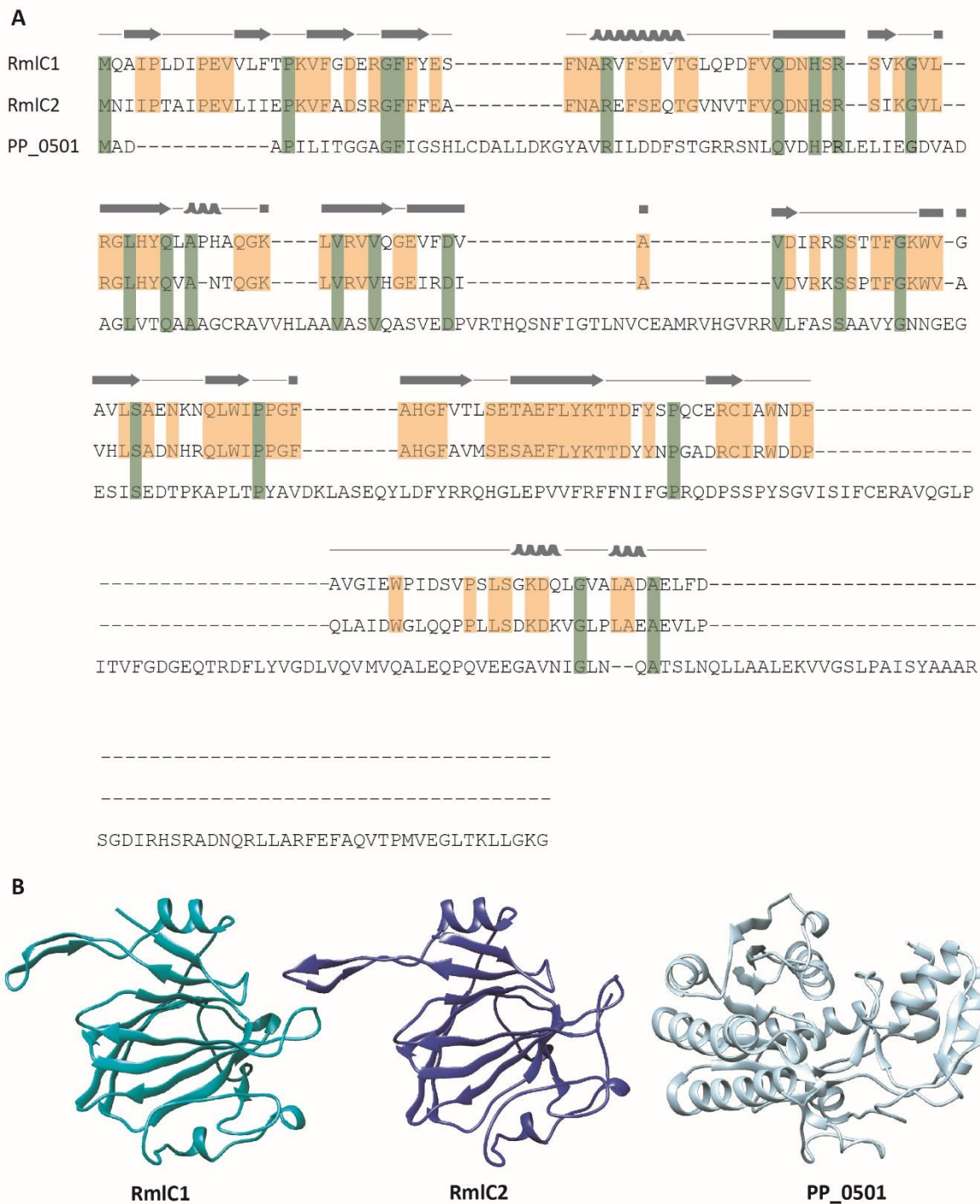


Fig.S1: **Sequence alignment and structure prediction of RmlC1, RmlC2 and PP_0501.** **A:** Multiple sequence alignment of RmlC1 (PP_1782), RmlC2 (PP_0265) and PP_0501 proteins from *Pseudomonas putida*. The multiple sequence alignment was generated using Clustal Omega (59). Secondary-structure elements (α -Helices and β -strands) are depicted at the top of the sequence and based on the *Salmonella typhimurium* RmlC crystal structure (1DZT). Conserved residues are colored in yellow (RmlC1 and RmlC2) and green (RmlC1, RmlC2 and PP_0501). **B:** Predicted protein folds of RmlC1 (PP_1782), RmlC2 (PP_0265) and PP_0500 proteins from *Pseudomonas putida*. Protein structures were predicted using Phyre2 (60). Illustrations were generated with UCSF Chimera (61).

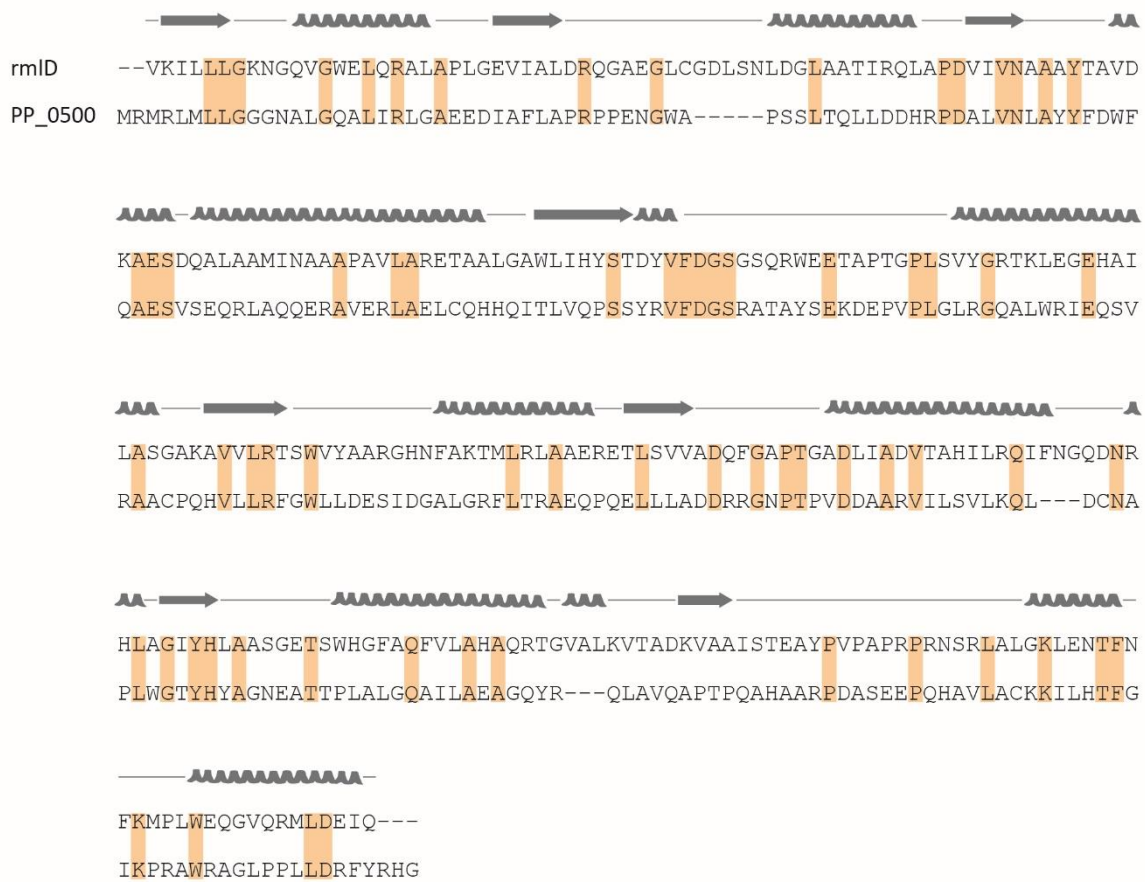
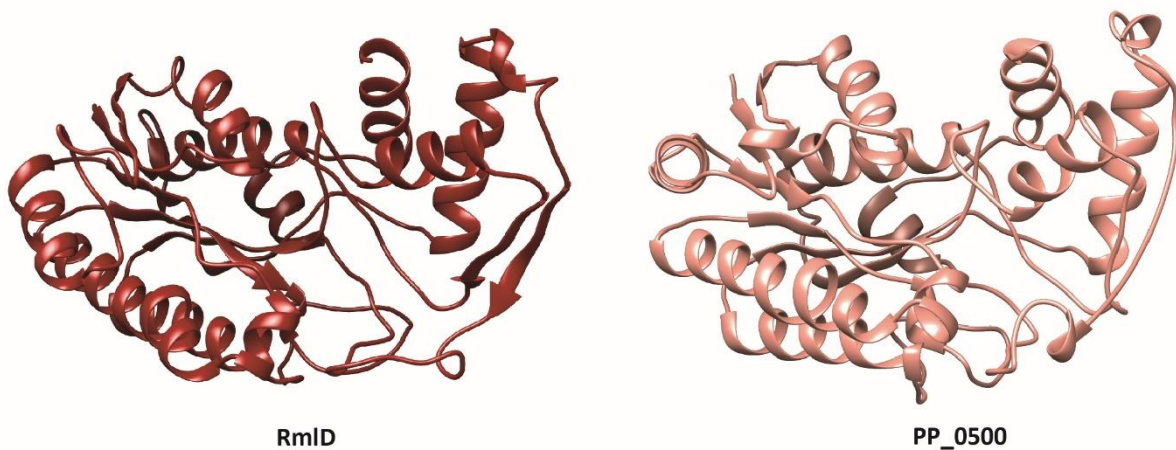
A**B**

Fig.S2: **Sequence alignment and structure prediction of RmlD and PP_0500.** **A:** Multiple sequence alignment of RmlD (PP_1784) and PP_0500 proteins from *Pseudomonas putida*. The multiple sequence alignment was generated using Clustal Omega (59). Secondary-structure elements (α -Helices and β -strands) are depicted at the top of the sequence and based on the *Salmonella typhimurium* RmlD crystal structure (1KBZ). Conserved residues are colored in yellow. **B:** Predicted protein folds of RmlD (PP_1784) and PP_0500 proteins from *Pseudomonas putida*. Protein structures were predicted using Phyre2 (60). Illustrations were generated with UCSF Chimera (61).

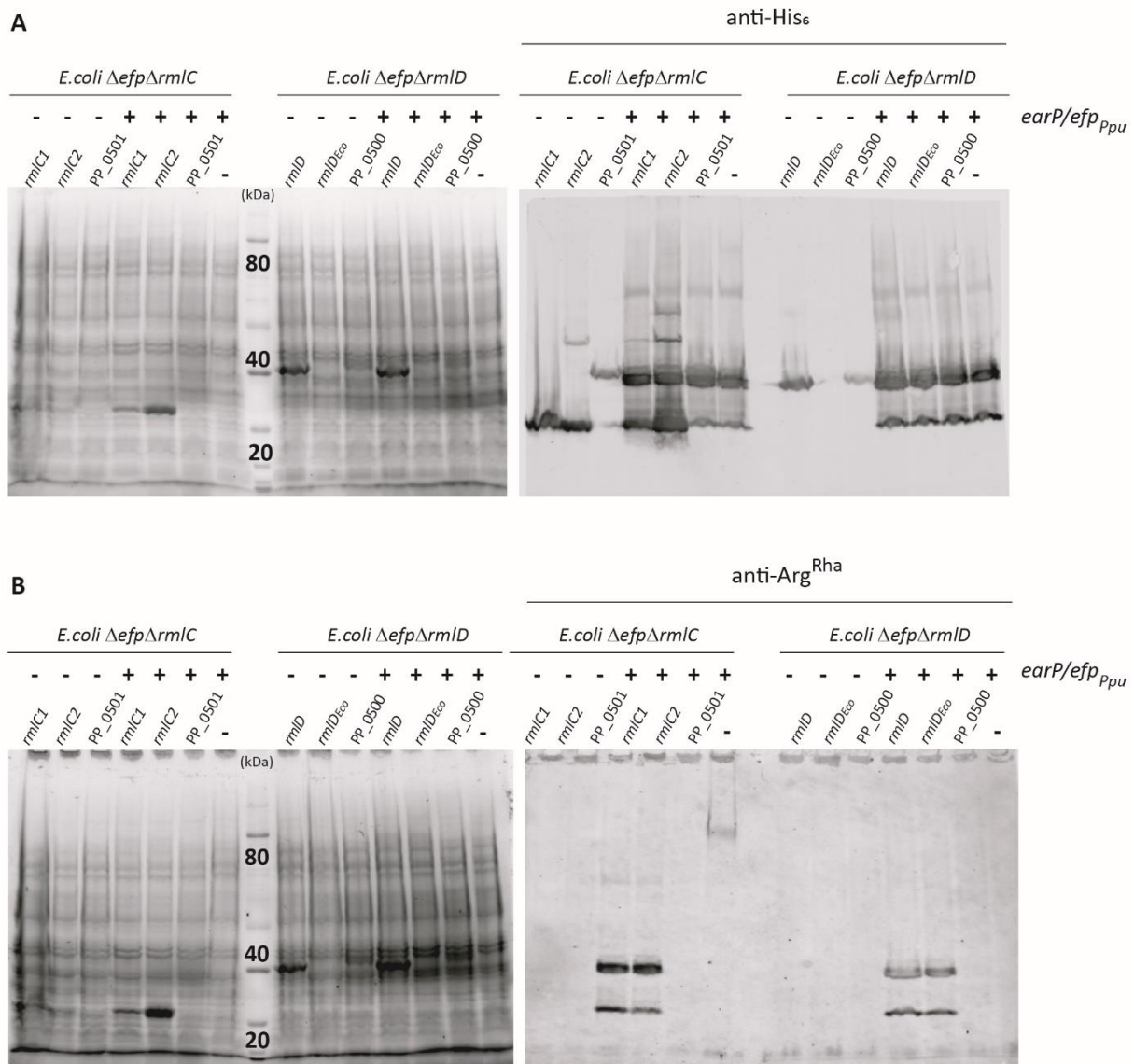


Fig. S3. **Analysis of activated EF-P in TDP-Rha biosynthesis deletion strains via western blot.** Functionalities of RmlC1, RmlC2, RmlD, PP_0500 and PP_0501 were determined by western blot in *E. coli* MG1655 $P_{cadBA}::lacZ \Delta$ efp Δ rmlC/ Δ rmlD with heterologous expression of a candidate gene from the pBAD33 vector. The empty vector (-) was included as negative control. Additionally, all strains encoded the *earP/efp_{Ppu}* operon in trans, being encoded from pBBR MCS2 vectors expressed from the native promoter (+). Again, the corresponding empty vector served as control (-). All strains were grown o/n in LB pH 5.8. **A**) Expression of candidate genes was verified using 0.1 μ g/ml anti-His₆. **B**) Rhamnosylated *EF-P_{Ppu}* (*EF-P^{Rha}*) was detected using 0.25 μ g/ml anti-Arg^{Rha}.

Material and methods

Bacterial strains and growth condition

All strains and plasmids used in this study are listed and described in table 1. *E. coli* cells were grown in Miller modified Lysogeny Broth (LB) (62, 63) at 37 °C aerobically under agitation, if not indicated otherwise. LB agar plates contained 1.5 % agar. The medium was supplemented with antibiotics at the following concentrations: 50 µg/ml kanamycin sulfate and 30 µg/ml chloramphenicol. Plasmids carrying the P_{BAD} promoter (38) were induced with L-arabinose at a final concentration of 0.2 % (w/v).

Table 1: Plasmids and strains used in this study.

Plasmid	Feature/ Genotype	Reference
pBAD33	CamR-cassette, p15A origin, araC coding sequence, ara operator	(38)
pBAD33_ <i>rmIC1</i>	CamR-cassette, arabinose inducible expression of RmIC1	this study
pBAD33_ <i>rmID</i>	CamR-cassette, arabinose inducible expression of RmID	this study
pBAD33_PP_0265	CamR-cassette, arabinose inducible expression of PP_0265	this study
pBAD33_PP_0500	CamR-cassette, arabinose inducible expression of PP_0500	this study
pBAD33_PP_0501	CamR-cassette, arabinose inducible expression of PP_0501	this study
Strain		
<i>E. coli</i> DH5αpir	F- λ- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF) U169, hsdR17(rK- mK+)	(64)
<i>E. coli</i> DH10B	F- <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) φ80 <i>lacZ</i> ΔM15 Δ <i>lacX74</i> <i>recA1</i> <i>endA1</i> <i>araD139</i> Δ (<i>ara-leu</i>)7697 <i>galU galK</i> λ- <i>rpsL</i> (Str ^R) <i>nupG</i>	(65)
<i>E.coli</i> MG1655	K-12 F- λ- <i>ilvG</i> - <i>rfb</i> -50 <i>rph</i> -1	(66)
<i>E.coli</i> P <i>cadBA::lacZ</i> Δ <i>efp</i>	MG1655 P <i>cadBA::lacZ</i> Δ(<i>cadBA</i>) Δ <i>efp</i>	(32)
<i>E.coli</i> P <i>cadBA::lacZ</i> Δ <i>efp</i> Δ <i>rmIC</i>	MG1655 P <i>cadBA::lacZ</i> Δ(<i>cadBA</i>) Δ <i>efp</i> Δ <i>rmIC</i>	(32)

<i>E. coli</i> P _{cadBA::lacZ} Δ <i>efp</i> Δ <i>rmID</i>	MG1655 P _{cadBA::lacZ} Δ (<i>cadBA</i>) Δ <i>efp</i> Δ <i>rmID</i>	(32)
---	--	------

Molecular biology methods

Oligonucleotides used in this study are listed and described in the supplementary table S1. Plasmid DNA was isolated using the Hi Yield® Plasmid Mini Kit from Süd Laborbedarf according to manufacturer's instructions. DNA fragments were purified from agarose gels using the Hi Yield® Gel/PCR DNA fragment extraction kit from Süd Laborbedarf. All restriction enzymes, DNA modifying enzymes and the Q5® high fidelity DNA polymerase for PCR amplification were purchased from New England BioLabs and used according to manufacturer's instructions.

Genomic library

The genomic DNA (gDNA) was isolated from 50 ml o/n culture of *P. putida* KT2440 according to the protocol described in reference (67). Further purification was achieved using Phase Lock Gel™ (QuantaBio) with Phenol-Chloroform. After the centrifugation, isopropanol precipitation was repeated. The pellet was resuspended in water, the final amount was 60 µg DNA.

Plasmid DNA was purified as described in **Molecular biology methods** from 12 ml *E. coli* DH5α cells. The plasmid DNA was diluted in water, the final amount was 10 µg DNA.

The library was constructed using *Sma*I (pBAD33 vector) and *Stu*I (gDNA) for digestion resulting in an average size of 5 kb per insert (Bionexus, Inc.). After ligation, the plasmids were transformed into *E. coli* DH10 B (Lucigen). Quality control was done by restriction digest of library clones with *Bam*HI. All restriction enzymes were produced by New England Biolabs, Frankfurt.

Bioinformatic tools

The multiple sequence alignment was generated using NCBI BLAST (68, 69) and Clustal Omega (59). Candidate homologues were identified and analysed using String (39, 40), Pfam (41), Uniprot (42), Metacyc (43) databases. Protein structures were predicted using Phyre2 (60). Illustrations were generated with UCSF Chimera (61).

β-galactosidase assay

E. coli MG1655 P_{cadBA::lacZ} Δ *efp* Δ *rmIC*/ Δ *rmID* expressing *lacZ* under the control of the *cadBA* promoter were grown in buffered LB (pH 5.8) overnight (o/n) and harvested by centrifugation. β-Galactosidase activities were determined as described in reference in biological triplicates

and are given in Miller units (MU) (70). Standard deviations from three independent experiments were determined.

SDS-PAGE and Western blotting

Electrophoretic separation of proteins was carried out using 12.5 % SDS-PAGE as described by Laemmli (71). Separated proteins were visualized in gel using 0.5% (vol/vol) 2-2,2-trichloroethanol (72) and detected within a Gel Doc™ EZ gel documentation system (Bio-Rad). The proteins were transferred onto a nitrocellulose membrane by vertical Western blotting (4 °C). Antigens were detected using 0.1 g/ml anti-His₆ tag (Abcam, Inc.) or 0.25 g/ml of anti-Arg^{Rha} (73). Primary antibodies (rabbit) were targeted using 0.1 µg/ml anti-rabbit IgG (IRDye® 680RD) (donkey) antibodies (Abcam). Target proteins were visualized via Odyssey® CLx Imaging System (LI-COR, Inc).

References

1. M. F. Giraud, J. H. Naismith, The rhamnose pathway. *Curr. Opin. Struct. Biol.* **10**, 687-696 (2000).
2. R. G. Eagon, Bacterial dissimilation of L-fucose and L-rhamnose. *J. Bacteriol.* **82**, 548-550 (1961).
3. A. Wegerer, T. Sun, J. Altenbuchner, Optimization of an *E. coli* L-rhamnose-inducible expression vector: test of various genetic module combinations. *BMC Biotechnol.* **8**, 2 (2008).
4. C. L. Kelly, G. M. Taylor, A. Hitchcock, A. Torres-Méndez, J. T. Heap, A Rhamnose-Inducible System for Precise and Temporal Control of Gene Expression in Cyanobacteria. *ACS Synth. Biol.* **7**, 1056-1066 (2018).
5. H. P. Wahl, H. Grisebach, Biosynthesis of streptomycin. dTDP-dihydrostreptose synthase from *Streptomyces griseus* and dTDP-4-keto-L-rhamnose 3,5-epimerase from *S. griseus* and *Escherichia coli* Y10. *Biochim. Biophys. Acta.* **568**, 243-252 (1979).
6. T. Moses, K. K. Papadopoulou, A. Osbourn, Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. *Crit. Rev. Biochem. Mol. Biol.* **49**, 439-462 (2014).
7. M. Wild, A. D. Caro, A. L. Hernández, R. M. Miller, G. Soberón-Chávez, Selection and partial characterization of a *Pseudomonas aeruginosa* mono-rhamnolipid deficient mutant. *FEMS Microbiol. Lett.* **153**, 279-285 (1997).
8. Y. Ma *et al.*, Drug targeting *Mycobacterium tuberculosis* cell wall synthesis: genetics of dTDP-rhamnose synthetic enzymes and development of a microtiter plate-based screen for inhibitors of conversion of dTDP-glucose to dTDP-rhamnose. *Antimicrob. Agents Chemother.* **45**, 1407-1416 (2001).
9. M. Gao, W. D'Haese, R. De Rycke, B. Wolucka, M. Holsters, Knockout of an azorhizobial dTDP-L-rhamnose synthase affects lipopolysaccharide and extracellular polysaccharide production and disables symbiosis with *Sesbania rostrata*. *Mol. Plant Microbe Interact.* **14**, 857-866 (2001).
10. J. Lassak *et al.*, Arginine-rhamnosylation as new strategy to activate translation elongation factor P. *Nat. Chem. Biol.* **11**, 266-270 (2015).

11. B. Kneidinger *et al.*, Identification of two GDP-6-deoxy-D-lyxo-4-hexulose reductases synthesizing GDP-D-rhamnose in *Aneurinibacillus thermoaerophilus* L420-91T. *J. Biol. Chem.* **276**, 5577-5583 (2001).
12. V. N. Shibaev, Biosynthesis of bacterial polysaccharide chains composed of repeating units. *Adv. Carbohydr. Chem. Biochem.* **44**, 277-339 (1986).
13. M. Schirm *et al.*, Structural and genetic characterization of glycosylation of type a flagellin in *Pseudomonas aeruginosa*. *J. Bacteriol.* **186**, 2523-2531 (2004).
14. W. Blankenfeldt, M. Asuncion, J. S. Lam, J. H. Naismith, The structural basis of the catalytic mechanism and regulation of glucose-1-phosphate thymidyltransferase (RmlA). *Embo J.* **19**, 6652-6663 (2000).
15. C. L. Marolda, M. A. Valvano, Genetic analysis of the dTDP-rhamnose biosynthesis region of the *Escherichia coli* VW187 (O7:K1) rfb gene cluster: identification of functional homologs of rfbB and rfbA in the rff cluster and correct location of the rffE gene. *J. Bacteriol.* **177**, 5539-5546 (1995).
16. S. T. Allard, M. F. Giraud, C. Whitfield, P. Messner, J. H. Naismith, The purification, crystallization and structural elucidation of dTDP-D-glucose 4,6-dehydratase (RmlB), the second enzyme of the dTDP-L-rhamnose synthesis pathway from *Salmonella enterica* serovar typhimurium. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 222-225 (2000).
17. M. Graninger, B. Nidetzky, D. E. Heinrichs, C. Whitfield, P. Messner, Characterization of dTDP-4-dehydrorhamnose 3,5-epimerase and dTDP-4-dehydrorhamnose reductase, required for dTDP-L-rhamnose biosynthesis in *Salmonella enterica* serovar Typhimurium LT2. *J. Biol. Chem.* **274**, 25069-25077 (1999).
18. S. L. van der Beek *et al.*, Streptococcal dTDP-L-rhamnose biosynthesis enzymes: functional characterization and lead compound identification. *Mol. Microbiol.* **111**, 951-964 (2019).
19. Y. Ma, F. Pan, M. McNeil, Formation of dTDP-rhamnose is essential for growth of mycobacteria. *J. Bacteriol.* **184**, 3392-3395 (2002).
20. M. Alhede, T. Bjarnsholt, M. Givskov, M. Alhede, *Pseudomonas aeruginosa* biofilms: mechanisms of immune evasion. *Adv. Appl. Microbiol.* **86**, 1-40 (2014).
21. N. B. Rendell *et al.*, Characterisation of *Pseudomonas* rhamnolipids. *BBA* **1045**, 189-193 (1990).
22. L. Zulianello *et al.*, Rhamnolipids are virulence factors that promote early infiltration of primary human airway epithelia by *Pseudomonas aeruginosa*. *Infect. Immun.* **74**, 3134-3147 (2006).
23. R. Krafczyk *et al.*, Structural Basis for EarP-Mediated Arginine Glycosylation of Translation Elongation Factor EF-P. *mBio* **8**, (2017).
24. T. Yanagisawa, T. Sumida, R. Ishii, C. Takemoto, S. Yokoyama, A paralog of lysyl-tRNA synthetase aminoacylates a conserved lysine residue in translation elongation factor P. *Nat. Struct. Mol. Biol.* **17**, 1136-1143 (2010).
25. A. Rajkovic *et al.*, Cyclic Rhamnosylated Elongation Factor P Establishes Antibiotic Resistance in *Pseudomonas aeruginosa*. *mBio* **6**, e00823 (2015).
26. S. L. Chiang, J. J. Mekalanos, rfb mutations in *Vibrio cholerae* do not affect surface production of toxin-coregulated pili but still inhibit intestinal colonization. *Infection and immunity* **67**, 976-980 (1999).
27. F. E. Aas, A. Vik, J. Vedde, M. Koomey, W. Egge-Jacobsen, *Neisseria gonorrhoeae* O-linked pilin glycosylation: functional analyses define both the biosynthetic pathway and glycan structure. *Molecular microbiology* **65**, 607-624 (2007).
28. K. E. Nelson *et al.*, Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environ. Microbiol.* **4**, 799-808 (2002).
29. V. A. P. Martins dos Santos, K. N. Timmis, B. Tümmler, C. Weinel, in *Pseudomonas: Volume 1 Genomics, Life Style and Molecular Architecture*, J.-L. Ramos, Ed. (Springer US, Boston, MA, 2004), pp. 77-112.
30. S. Sivendran *et al.*, Identification of triazinoindol-benzimidazolones as nanomolar inhibitors of the *Mycobacterium tuberculosis* enzyme TDP-6-deoxy-d-xylo-4-

- hexopyranosid-4-ulose 3,5-epimerase (RmIC). *Bioorg. Med. Chem.* **18**, 896-908 (2010).
31. S. Choi, J. Choe, Crystal structure of elongation factor P from *Pseudomonas aeruginosa* at 1.75 Å resolution. *Proteins* **79**, 1688-1693 (2011).
 32. S. Ude *et al.*, Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science (New York, N. Y.)* **339**, 82-85 (2013).
 33. J. Lassak, D. N. Wilson, K. Jung, Stall no more at polyproline stretches with the translation elongation factors EF-P and IF-5A. *Mol. Microbiol.* **99**, 219-235 (2016).
 34. W. W. Navarre *et al.*, PoxA, yjeK, and elongation factor P coordinately modulate virulence and drug resistance in *Salmonella enterica*. *Mol. Cell.* **39**, 209-221 (2010).
 35. M. S. Gilreath *et al.*, β-Lysine discrimination by lysyl-tRNA synthetase. *FEBS Lett.* **585**, 3284-3288 (2011).
 36. M. Pfab *et al.*, Synthetic post-translational modifications of elongation factor P using the ligase EpmA. *Febs j.* **288**, 663-677 (2021).
 37. L. Peil *et al.*, Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15265-15270 (2013).
 38. L. M. Guzman, D. Belin, M. J. Carson, J. Beckwith, Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121-4130 (1995).
 39. B. Snel, G. Lehmann, P. Bork, M. A. Huynen, STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* **28**, 3442-3444 (2000).
 40. D. Szklarczyk *et al.*, STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607-d613 (2019).
 41. S. El-Gebali *et al.*, The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427-d432 (2019).
 42. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506-d515 (2019).
 43. R. Caspi *et al.*, The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res.* **48**, D445-d453 (2020).
 44. J. Seo, A. Brencic, A. J. Darwin, Analysis of secretin-induced stress in *Pseudomonas aeruginosa* suggests prevention rather than response and identifies a novel protein involved in secretin function. *J. Bacteriol.* **191**, 898-908 (2009).
 45. X. Li *et al.*, Resolving the α-glycosidic linkage of arginine-rhamnosylated translation elongation factor P triggers generation of the first Arg(Rha) specific antibody. *Chem. Sci.* **7**, 6995-7001 (2016).
 46. D. Gast *et al.*, A set of rhamnosylation-specific antibodies enables detection of novel protein glycosylations in bacteria. *Org. Biomol. Chem.* **18**, 6823-6828 (2020).
 47. L. Tetsch, C. Koller, I. Haneburger, K. Jung, The membrane-integrated transcriptional activator CadC of *Escherichia coli* senses lysine indirectly via the interaction with the lysine permease LysP. *Molecular microbiology* **67**, 570-583 (2008).
 48. G. Brandis, D. Hughes, The SNAP hypothesis: Chromosomal rearrangements could emerge from positive selection during niche adaptation. *PLoS Genet.* **16**, e1008615 (2020).
 49. A. Pföstl *et al.*, Biosynthesis of dTDP-3-acetamido-3,6-dideoxy-α-D-glucose. *Biochem. J.* **410**, 187-194 (2008).
 50. R. P. Pandey, P. Parajuli, R. B. Gurung, J. K. Sohng, Donor specificity of YjiC glycosyltransferase determines the conjugation of cytosolic NDP-sugar in *in vivo* glycosylation reactions. *Enzyme Microb. Technol.* **91**, 26-33 (2016).
 51. M. Yamamoto *et al.*, Identification of Genes Involved in the Glycosylation of Modified Viosamine of Flagellins in *Pseudomonas syringae* by Mass Spectrometry. *Genes* **2**, 788-803 (2011).

52. L. S. Marmont *et al.*, PelX is a UDP-N-acetylglucosamine C4-epimerase involved in Pel polysaccharide-dependent biofilm formation. *J. Biol. Chem.* **295**, 11949-11962 (2020).
53. L. Friedman, R. Kolter, Genes involved in matrix formation in *Pseudomonas aeruginosa* PA14 biofilms. *Mol. Microbiol.* **51**, 675-690 (2004).
54. Y. Nakano *et al.*, Thymidine diphosphate-6-deoxy-L-lyxo-4-hexulose reductase synthesizing dTDP-6-deoxy-L-talose from *Actinobacillus actinomycetemcomitans*. *J. Biol. Chem.* **275**, 6806-6812 (2000).
55. W. Volkwein *et al.*, Switching the Post-translational Modification of Translation Elongation Factor EF-P. *Front. Microbiol.* **10**, 1148 (2019).
56. L. Yakovlieva *et al.*, A β -hairpin epitope as novel structural requirement for protein arginine rhamnosylation. *Chem. Science*, (2021).
57. T. Sengoku *et al.*, Structural basis of protein arginine rhamnosylation by glycosyltransferase EarP. *Nat. Chem. Biol.* **14**, 368-374 (2018).
58. C. He *et al.*, Complex Structure of *Pseudomonas aeruginosa* Arginine Rhamnosyltransferase EarP with Its Acceptor Elongation Factor P. *J. Bacteriol.* **201**, (2019).
59. F. Sievers, D. G. Higgins, Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol. Biol.* **1079**, 105-116 (2014).
60. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10**, 845-858 (2015).
61. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605-1612 (2004).
62. G. Bertani, Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* **62**, 293-300 (1951).
63. G. Bertani, Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J. Bacteriol.* **186**, 595-600 (2004).
64. D. R. Macinga, M. M. Parojic, P. N. Rather, Identification and analysis of aarP, a transcriptional activator of the 2'-N-acetyltransferase in *Providencia stuartii*. *J. Bacteriol.* **177**, 3407-3413 (1995).
65. T. Durfee *et al.*, The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. *J. Bacteriol.* **190**, 2597-2606 (2008).
66. M. S. Guyer, R. R. Reed, J. A. Steitz, K. B. Low, Identification of a sex-factor-affinity site in *E. coli* as gamma delta. *Cold Spring Harb. Symp. Quant. Biol.* **45 Pt 1**, 135-140 (1981).
67. A. Pospiech, B. Neumann, A versatile quick-prep of genomic DNA from gram-positive bacteria. *Trends Genet.* **11**, 217-218 (1995).
68. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **44**, D7-19 (2016).
69. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410 (1990).
70. J. H. Miller, *A short course in bacterial genetics – A laboratory manual and handbook for Escherichia coli and related bacteria.*, Cold Spring Harbor Laboratory Press (Cold Spring Harbor, 1992).
71. U. K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685 (1970).
72. C. L. Ladner, J. Yang, R. J. Turner, R. A. Edwards, Visible fluorescent detection of proteins in polyacrylamide gels without staining. *Anal. Biochem.* **326**, 13-20 (2004).
73. X. Li *et al.*, Resolving the alpha-glycosidic linkage of arginine-rhamnosylated translation elongation factor P triggers generation of the first Arg(Rha) specific antibody. *Chem. Sci.* **7**, 6995-7001 (2016).

3 A set of rhamnosylation-specific antibodies enables detection of novel protein glycosylations in bacteria

Daniel Gast*, Franziska Koller*, Ralph Krafczyk, Lukas Bauer, Swetlana Wunder, Jürgen Lassak and Anja Hoffmann-Röder (2020). A set of rhamnosylation-specific antibodies enables detection of novel protein glycosylations in bacteria. *Org Biomol Chem.* 2020; 18(35): 6823-6828.



Cite this: DOI: 10.1039/d0ob01289k

Received 23rd June 2020,
Accepted 14th August 2020

DOI: 10.1039/d0ob01289k

rsc.li/obc

A set of rhamnosylation-specific antibodies enables detection of novel protein glycosylations in bacteria†

Daniel Gast,^{‡a} Franziska Koller,^{‡b} Ralph Krafczyk,^{‡b} Lukas Bauer,^a Swetlana Wunder,^a Jürgen Lassak^{‡b} and Anja Hoffmann-Röder^{‡*a}

Despite its potential importance for bacterial virulence, protein rhamnosylation has not yet been sufficiently studied. Specific *anti-Ser^{Rha}*, *anti-Thr^{Rha}* and *anti-Asn^{Rha}* antibodies allowed the identification of previously unknown monorhamnosylated proteins in cytosol and membrane fractions of bacterial cell lysates. Mapping of the complete rhamnoprotoeome in pathogens should facilitate development of targeted therapies against bacterial infections.

Introduction

To perform their specific biological function, peptides and proteins are usually altered post-translationally by attachment of specific groups. Glycosylation, *i.e.* the enzymatic attachment of sugar residues to the side chains of amino acids, is among the most common and diverse post-translational modifications (PTMs). Such protein glycosylations occur in all domains of life and regulate important cellular processes, including differentiation, signal transduction and the immune response.¹ In addition, bacterial protein glycosylation often plays an important role in the pathogenicity of bacteria.² Typical prokaryotic glycosylation patterns include *O*-linked glycans, preferably bound to serine (Ser), threonine (Thr), and less frequently to tyrosine (Tyr), as well as *N*-bound sugars.^{3,4} The latter includes the extensively studied glycosylation pattern of *Campylobacter jejuni*,⁵ in which a pre-assembled heptasaccharide is enzymatically transferred *en-bloc* onto asparagine (Asn), as well as the attachment of monosaccharides.^{6–8} For instance, the glycosyltransferase HMW1C from *Haemophilus influenzae* attaches various single glucose and galactose units to its target protein,

the adhesin HMW1, with the help of the sugar nucleotides UDP-glucose and UDP-galactose as activated glycosyl donors.^{9,10} Similarly NleB1, initially reported to attach *N*-acetylglucosamine to arginine (Arg) side chains of host proteins, also regulates bacterial physiology by promoting glycosylation of a specific Arg residue of the bacterial glutathione synthetase (GshB) in *Citrobacter rodentium*.^{11–13} Another example for this rather unusual Arg glycosylation was recently described for translation elongation factor P (EF-P), which is estimated to be glycosylated in approximately 10% of all bacteria,¹⁴ including clinically important germs like *Pseudomonas aeruginosa*^{14,15} and *Neisseria meningitidis*.¹⁶ Here, a single rhamnose (Rha) moiety is linked to a key Arg unit of EF-P by the glycosyltransferase EarP.^{14,17} Rhamnosylation of EF-P is essential for preventing ribosome stalling when translating polyproline motifs. It also represents the first example for a protein regulation by *N*-monoglycosylation in bacteria.^{14,18} However, the widespread existence of synthesis pathways for activated rhamnose dinucleotide precursors (dT/UDP-β-L-rhamnose, GDP-α-D-rhamnose) in bacteria, archaea and eukaryotes raises the question whether such monorhamnosylations are indeed only curiosities or whether they rather represent a general mechanism for functional protein regulation in prokaryotes.⁶ More importantly, and in view of the close link between protein glycosylation and bacterial pathogenicity, research in this direction might also offer new therapeutic avenues to combat bacterial infectious diseases,¹⁹ which has become even more urgent with the spread of infections caused by multiresistant germs. An important prerequisite for the development of new anti-infectives or therapeutic antibacterial approaches is the identification of (new) glycoprotein targets and their specific biological functions. Despite recent technological advances, profiling of bacterial glycosylation sites is still a difficult task and usually requires specific enrichment strategies due to the dynamics of glycosylation and its stoichiometry.²⁰ Here, immunoblotting using modification-specific antibodies is particularly useful, although multiple analyses with different (complementary) antibodies are often

^aDepartment of Chemistry, Ludwig-Maximilians-Universität München, Munich, Germany. E-mail: anja.hoffmann-roeder@cup.lmu.de

^bDepartment of Biology I, Microbiology, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany. E-mail: juergen.lassak@lmu.de

† Electronic supplementary information (ESI) available: Details on chemical procedures and characterizations of new compounds; details on antibody evaluation and detection of rhamnoprotoeins. See DOI: 10.1039/d0ob01289k

‡ These authors contributed equally to this work.

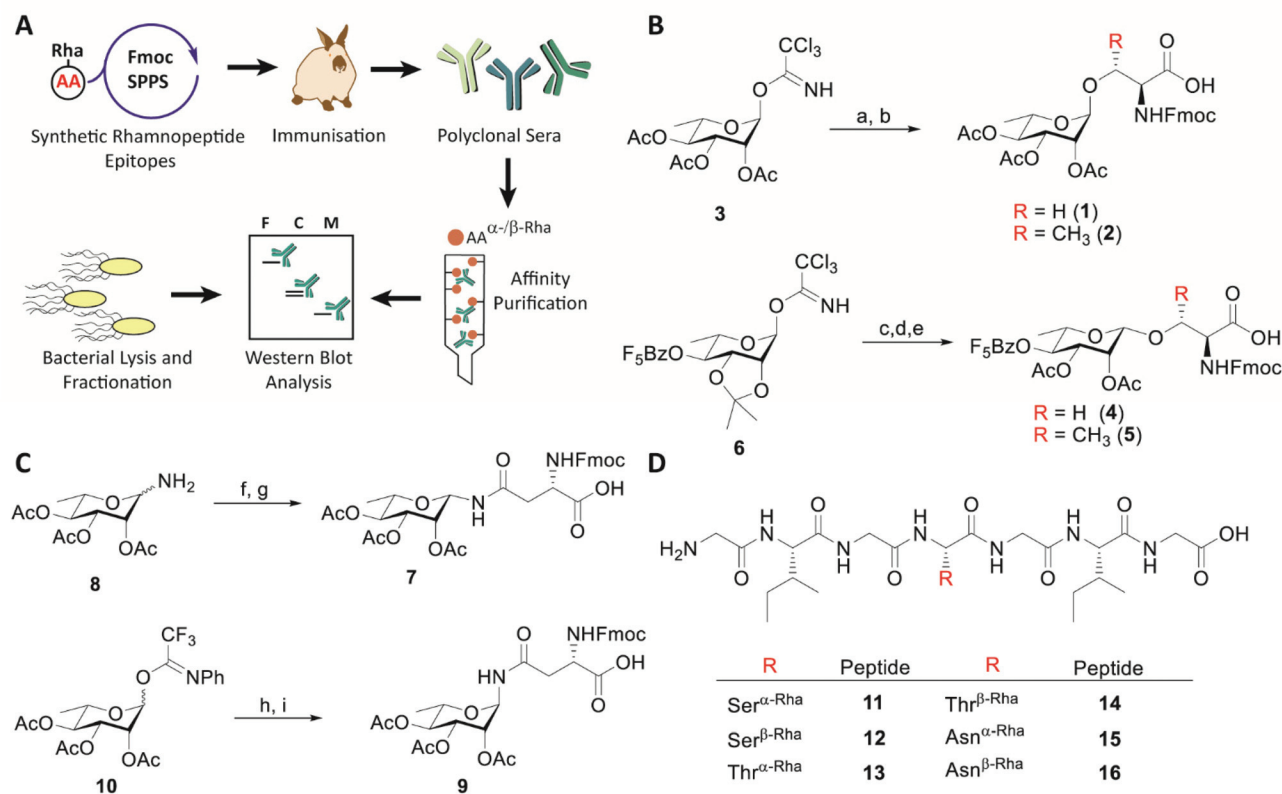


Fig. 1 (A) Generation of polyclonal *anti-Ser^{Rha}*, *anti-Thr^{Rha}* and *anti-Asn^{Rha}* from synthetic rhamnopeptide epitopes and their use in profiling of the bacterial rhamnosyl proteome (F = flagella, C = cytosol fraction, M = membrane fraction). (B) Synthesis of α -O-rhamnosyl serine and threonine building blocks: reagents and conditions: (a) serine: Fmoc-L-Ser-OAll, TMSOTf, CH₂Cl₂, 74%; threonine: Fmoc-L-Thr-OAll, TMSOTf, CH₂Cl₂, 75%; (b) PhSiH₃, [Pd(PPh₃)₄], CH₂Cl₂, 97% (1), 85% (2). Synthesis of β -O-rhamnosyl serine and threonine building blocks: reagents and conditions: (c) serine: Fmoc-L-Ser-OAll, HNTf₂, Et₂O/CH₂Cl₂ 76%; threonine: Fmoc-L-Thr-OAll, HNTf₂, Et₂O/CH₂Cl₂ 85%; (d) (1) TFA, H₂O, (2) Ac₂O, pyridine, serine: 87% threonine: 74% over 2 steps; (e) [Pd(PPh₃)₄], *N*-methyl aniline, THF, 98% (4), 94% (5). (C) Synthesis of β -N-rhamnosyl asparagine building block: reagents and conditions: (f) Fmoc-L-Asp-OAll, PyBOP, DIPEA, DMF, 79%; (g) PhSiH₃, [Pd(PPh₃)₄], CH₂Cl₂, 81%. Synthesis of α -N-rhamnosyl asparagine building block: reagents and conditions: (h) Fmoc-L-Asn-OAll, TMSOTf, MeNO₂/CH₂Cl₂, 63%; (i) PhSiH₃, [Pd(PPh₃)₄], CH₂Cl₂, 76%. (D) Schematic depiction of O- and N-rhamnosylated glycopeptides 11–16 for antibody generation.

needed.²¹ So far, only *anti-Arg^{Rha}* antibodies have been described for the detection of protein monorhamnosylation in cell lysates.^{17,22} However, to be able to detect further potentially important rhamnose *N*- and *O*-linked glycosylation patterns, we are now expanding the existing biochemical toolbox with this study. Novel, and specific *anti-Asn^{Rha}*, *anti-Ser^{Rha}* and *anti-Thr^{Rha}* polyclonal antibodies are presented, which should allow for the detection of previously unknown monorhamnosylated bacterial proteins (Fig. 1A).

Results and discussion

Based on previous work of Hu and co-workers, who generated first *anti-Arg^{Rha}* antibodies with the aid of synthetic glycopeptides,^{22,23} we also used an artificial glycopeptide scaffold of the sequence Gly-Ile-Gly-AA^{Rha}-Gly-Ile-Gly for production of our extended set of specific *anti-AA^{Rha}* antibodies. Here, AA^{Rha} stands either for Asn^{α/β-Rha}, Ser^{α/β-Rha} or Thr^{α/β-Rha}, which represent chemically pre-formed Rha-building blocks of

Asn, Ser, and Thr. In general, these pre-assembled rhamnosyl amino acids can be incorporated into any given peptide sequence by solid-phase peptide synthesis (SPPS) according to the Fmoc-protocol. Chemical approaches towards such building blocks have been reported earlier,^{24,25} with exception of *N*-linked Asn^{Rha} and the synthetically challenging β -O-rhamnosyl Ser/Thr building blocks. Detailed information on the preparation of all of these (novel) building blocks are provided in the ESI.† In brief, *O*-linked α -rhamnosylated Ser and Thr SPPS building blocks 1 and 2 were assembled from Fmoc-L-Ser-OAll²⁶ and Fmoc-L-Thr-OAll²⁶ via TMSOTf-catalysed glycosylation with trichloroacetimidate donor 3²⁷ (Fig. 1B). Subsequent reductive deallylation furnished the desired products stereochemically pure and in 72% and 64% overall yield, respectively. Similarly, the corresponding β -rhamnosylated Ser-*O*- and Thr-*O*-building blocks 4 and 5 were obtained from Fmoc-L-Ser-OAll²⁶ and Fmoc-L-Thr-OAll²⁶ using rhamnosyl donor 6²⁸ and Tf₂NH as synthesis promotor. Acidic cleavage of the isopropylidene acetal protecting group, followed by reacetylation and deallylation, then provided the corresponding β -O-rhamnosy-

lated Ser and Thr derivatives **4** and **5**, in 65% and 59% overall yield, respectively. *N*-Linked β -rhamnosylated Asn SPPS building block **7** was obtained in 64% yield by PyBOP-mediated coupling of rhamnosylamine precursor **8**²⁹ to Fmoc-L-Asp-OAll³⁰ and Pd-catalysed deallylation (Fig. 1C). In contrast, α -*N*-linked rhamnosyl Asn derivative **9** was not accessible *via* this route due to the high racemization tendency of the intermediate rhamnosylamine **8**.²⁹ However, TMSOTf-catalyzed glycosylation of Fmoc-L-Asn-OAll with *N*-phenyl trifluoroacetimidate donor **10**^{31,32} and subsequent ester cleavage furnished the desired product in 48% yield. The stereochemical assignments of all six *N*- and *O*-rhamnosyl amino acid building blocks were proven by ¹H-¹³C-coupled HSQC NMR experiments (see ESI†).^{29,33} With the pre-formed amino acid building blocks **1**, **2**, **4**, **5**, **7** and **9** at hands, a set of *O*- and *N*-rhamnosylated glycopeptides (**11–16**) comprising either α - or β -rhamnosylated Ser, Thr or Asn residues was synthesized on 2-chlorotrityl resin using standard Fmoc-SPPS procedures (see ESI†). Additional glycine units were introduced into the sequence to act as separators, while the corresponding isoleucine residue should facilitate HPLC purification (Fig. 1D). Polyclonal *anti*-Asn^{Rha}, *anti*-Ser^{Rha}, and *anti*-Thr^{Rha} antibodies were raised commercially by Eurogentec GmbH, Germany, according to their Rabbit Speedy 28-day (AS superantigen) program. For more details please refer to the ESI.† To ensure cross reactivity of the antibodies to both anomeric rhamnosyl linkages, equimolar mixtures of α - and β -configured glycopeptides were used for immunization of rabbits. The specificities of the obtained *anti*-Ser^{Rha}, *anti*-Thr^{Rha} and *anti*-Asn^{Rha} polyclonal antibodies were

investigated with semi-synthetic glycopeptide conjugates (BSA-Ser ^{α -/ β -Rha}, BSA-Thr ^{α -/ β -Rha}, BSA-Asn ^{α -/ β -Rha}) derived from bovine serum albumin (BSA) and the corresponding α - or β -rhamnosylated glycopeptides (Fig. 2). In addition, the BSA-conjugates of the naked, non-glycosylated peptide sequences were prepared as negative controls (BSA-Ser^{NP}, BSA-Thr^{NP}, BSA-Asn^{NP}). All three polyclonal antibodies strongly bound their cognate glycopeptide hapten up to a concentration range of only 1–2 ng of BSA conjugate (Fig. 2D and ESI Fig. 6B and 7B†). By contrast, the antisera neither recognized the corresponding naked BSA conjugates lacking the sugar moiety, nor synthetic EF-P variants (EF-P^{Ser-NP}, EF-P^{Thr-NP}, EF-P^{Asn-NP}) in which the internal loop region carrying the natural rhamnosylation site was substituted by the corresponding naked peptide (Fig. 2A, ESI Fig. 6A and 7A†).¹⁴ This demonstrates that the rhamnosylated glycosylamino acids are key structural elements of the antibodies' cognate haptens. This finding was confirmed by competition assays, in which the antibodies were pre-incubated with either L-rhamnose (up to 15 mM) or the respective amino acid (L-Ser, L-Thr, L-Asn, up to 15 mM) prior to immunodetection. Here, no impairment in recognition of the corresponding rhamnosylated BSA conjugates was observed (Fig. 2C, ESI Fig. 6A and 7A†). Moreover, cross-reactivity of the antibodies to non-cognate glycopeptide antigens was tested. While *anti*-Thr^{Rha} could detect β -*O*-rhamnosylated Ser (Fig. 2B) and *vice versa* (ESI, Fig. 6C†), almost no binding to *N*-rhamnosylated Asn was observed. *Anti*-Asn^{Rha}, on the other hand, did not recognize *N*-Arg rhamnosylation but showed some cross-reactivity with *O*-rhamnosylated Ser and

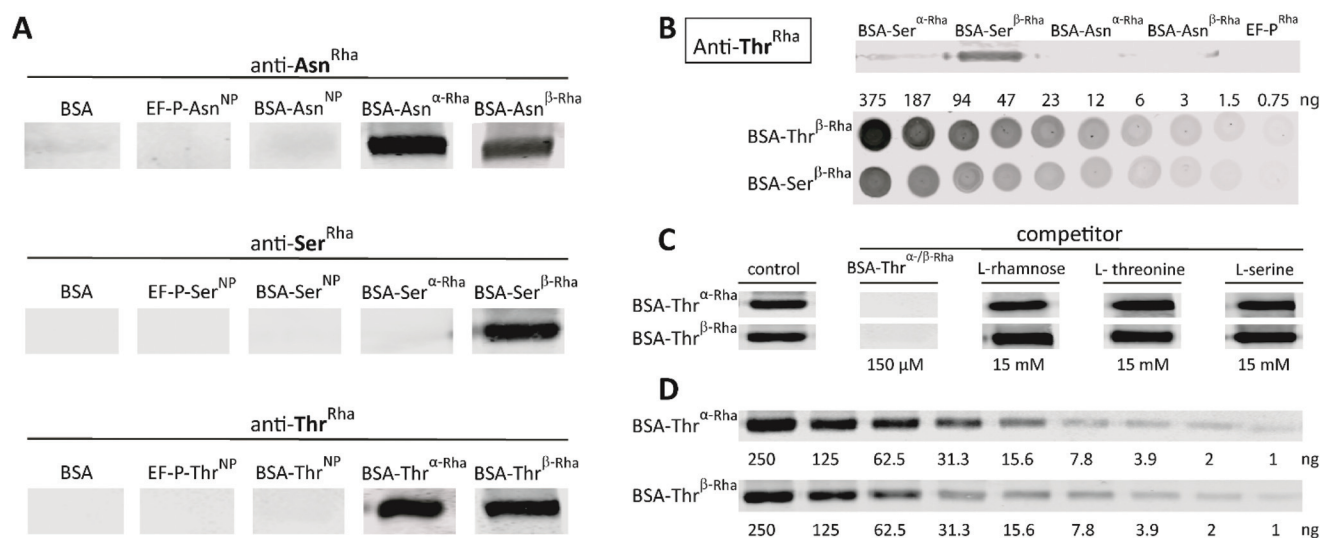


Fig. 2 Sensitivity and specificity analysis of *anti*-aa^{Rha} specific antibodies. (A) Specificity analysis: 0.5 μ g of the respective sample were subjected to SDS-PAGE and subsequent immunodetection analysis with 0.2 mg ml⁻¹ of *anti*-aa^{Rha}. BSA, synthetic EF-P variants comprising the naked peptide sequence in their loop region (EF-P-aa^{NP}) and BSA coupled to the naked peptide (BSA-aa^{NP}) served as negative controls. α - and β -rhamnosylated peptides coupled to BSA (BSA-aa ^{α -/ β -Rha}) served as positive controls. (B) *Anti*-Thr^{Rha} cross-reactivity analysis: 0.5 μ g of the respective sample were subjected to SDS-PAGE and subsequent Western Blot analysis with 0.2 mg ml⁻¹ of *anti*-Thr^{Rha}. Upper part: Immunodetection of BSA-Ser ^{α -/ β -Rha}, BSA-Asn ^{α -/ β -Rha} and EF-P^{Rha}. Lower part: Immunodetection of varying concentrations of BSA-Thr ^{β -Rha} and BSA-Ser ^{β -Rha}. (C) Cross-reactivity analysis of *anti*-Thr^{Rha} against L-rhamnose, L-Thr and L-Ser. *Anti*-Thr^{Rha} was preincubated prior to immunodetection with BSA-Thr ^{α -/ β -Rha} (150 μ M), L-rhamnose, L-Thr and L-Ser (15 mM). (D) *Anti*-Thr^{Rha} sensitivity analysis: immunodetection of varying concentrations of BSA-Thr ^{α -/ β -Rha}. Antibody concentrations were kept constant at 0.2 mg ml⁻¹.

Thr (ESI Fig. 7C†). It should be noted however, that the cognate antigen/antibody pairing was always superior (Fig. 2B lower part, and ESI Fig. 6 and 7†). Interestingly, whereas the *anti-Ser^{Rha}* antibody recognized almost exclusively the β -O-rhamnosylated glycopeptide BSA conjugate (Fig. 2A and ESI Fig. 6A†), no such anomeric preference was observed for the other antibodies. Finally, none of the antibodies was found cross-reactive with naturally occurring oligosaccharides carrying L-rhamnose at the reducing end linked to Asn or Ser/Thr. Thus, immunoblotting experiments with the S-layer protein from *Geobacillus stearothermophilus*, characterized by a N-rhamnosylated trisaccharide repeating unit,³⁴ as well as with the O-rhamnosylated flagellar filament protein FliC from *P. aeruginosa*³⁵ turned out negative and clearly demonstrate the desired specificity of the antibodies for monorhamnosylation (ESI Fig. 9†). With the rhamnosylation-specific antibodies *anti-Ser^{Rha}*, *anti-Thr^{Rha}* and *anti-Asn^{Rha}* at hands, we aimed at detection of rhamnosylated endogenous proteins from distinct

phyla (Fig. 3A). Since thus far monorhamnosylation of proteins has only been reported for bacteria,^{8,14–16} we have restricted our selection to prokaryotes and in particular to those microorganisms known to synthesize nucleoside diphosphate (NDP)-rhamnose (e.g. as donor substrate for EF-P-Arg rhamnosylation¹⁴). The only exceptions in this regard are *Micrococcus luteus* and *Staphylococcus aureus* which lack the corresponding NDP-rhamnose pathways and served as negative controls. To cover all potential rhamnosylation motifs, the already known *anti-Arg^{Rha}* antibody¹⁷ was also included in the study. Sensitivity analysis based on 10^9 cells per western blot allowed estimation of the lower detection limit to be in the range of 60–150 rhamnosylation events per cell. This in turn should reveal most of the putative rhamnoprotoe. To enhance sensitivity, the flagella of swimming bacteria were isolated, and cells were further separated into cytosol and membrane fractions prior to immunoblotting. Cellular fractionation does not only allow for conclusions on the localization of detected gly-

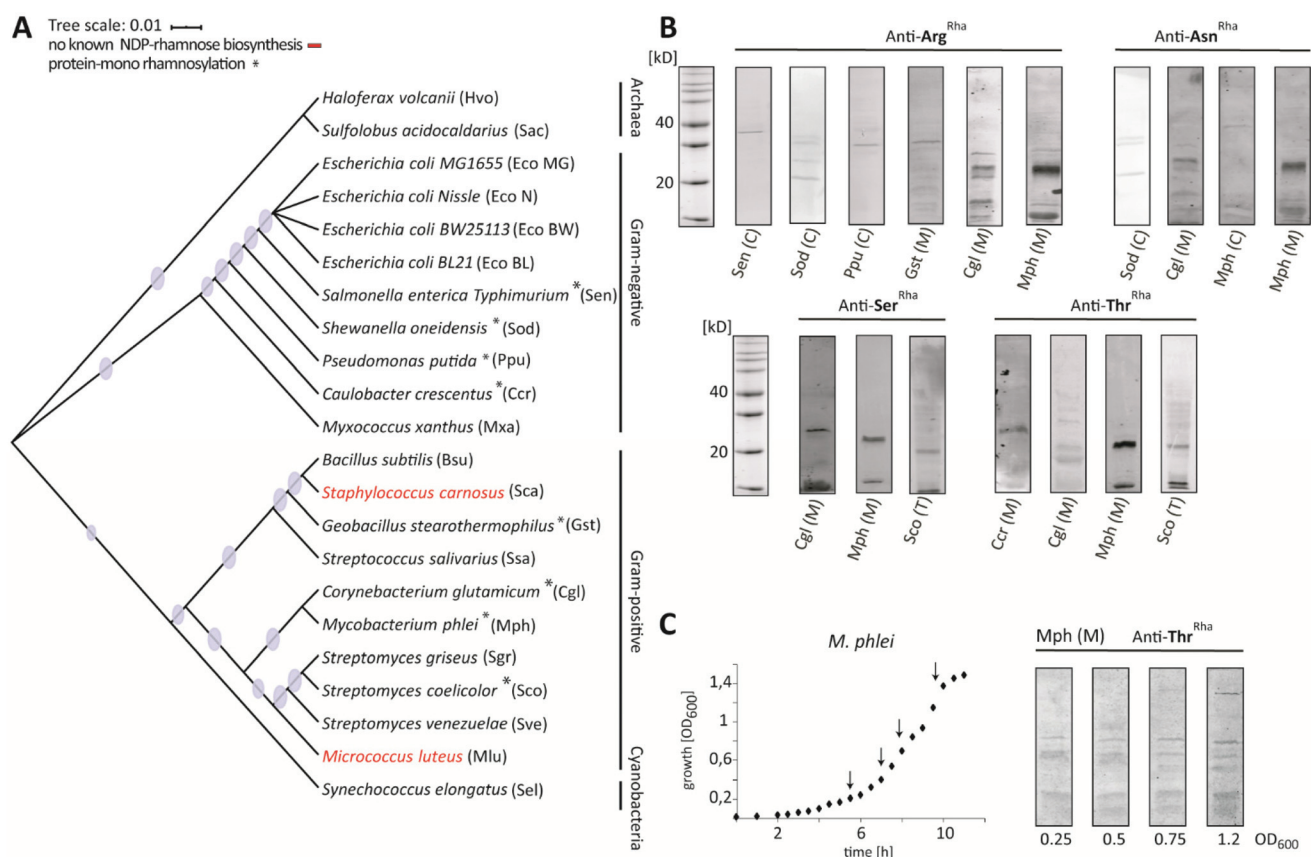


Fig. 3 (A) Phylogenetic distribution of microorganisms selected for rhamnoprotoe analysis. Depicted is a rooted phylogenetic tree based on 16S rDNA sequences. The tree was calculated using ClustalX and visualized using iTOL. Bootstrap values (0–1000) are given as transparent circles. The circle size indicates the probability of correct branching with the largest circles having a bootstrap value of 1000 and the smallest circle with a bootstrap value of 533. A star marks those bacteria in which monorhamnosylation was detected. Black font indicates for a known NDP-rhamnose biosynthesis pathway, whereas a red font marks those bacteria where the respective genes are absent. (B) Immunodetection of rhamnosylated proteins using *anti-Arg^{Rha}*, *anti-Asn^{Rha}*, *anti-Ser^{Rha}* and *anti-Thr^{Rha}*. Prokaryotic samples were separated into cytosolic (C) and membrane (M) fractions and subjected to SDS-PAGE and subsequent Western Blot analysis with 0.2 mg ml⁻¹ of the corresponding antibody. (C) Immunodetection of *M. phlei* rhamnoprotoe from different growth phases. Arrows indicate the different time points of sample collection. Immunodetection is exemplified for *anti-Thr^{Rha}* (0.2 mg ml⁻¹).

coproteins but also accumulates those rhamnopeptides present in the membrane or being part of the flagellum. This enrichment is important since many prokaryotic protein glycosylation events occur on cellular appendages or membrane proteins.^{8,36–39} While protein monorhamnosylation was not detected in the selected archaea, bands indicative of rhamnoproteins were readily identified in immunoblotting experiments with different bacterial phyla (Fig. 3B). Specifically, *N*-rhamnosylation was detected in both Gram-negative and Gram-positive species, whereas *O*-rhamnosylation was mainly observed in the latter. Rhamnoproteins appeared to be both, membrane-bound and cytosolic, while flagellin rhamnosylation was not detected (Fig. 3B and ESI Fig. 11†). In addition, significant differences in the relative abundance of monorhamnosylated proteins were revealed for various bacteria. While EF-P seems to be the only rhamnoprotein in *P. putida*, multiple rhamnoproteins were detected in *Corynebacterium glutamicum* and *Mycobacterium phlei*. It is noteworthy, that for the latter the dTDP- β -L-rhamnose biosynthesis pathway is essential.⁴⁰ Interestingly, the number of detected monorhamnosylated proteins varied over the time of growth with an additionally band appearing exclusively in the stationary phase of *M. phlei* (Fig. 3C and ESI Fig. 10†).

Conclusion

In summary, new polyclonal antibodies for the specific recognition of protein monorhamnosylation in bacteria were developed from novel synthetic designer *N*- and *O*-glycopeptide antigens. These *anti*-Ser^{Rha}, *anti*-Thr^{Rha} and *anti*-Asn^{Rha} antisera can be applied together with already known *anti*-Arg^{Rha} antibodies as useful tools for biochemical and proteomic studies, *e.g.* identifying potential *N*- and *O*-rhamnosylation sites. Thus, we have successfully used the complete set of antibodies in immunoblotting experiments to detect novel monorhamnosylated proteins in various bacteria. The existence of rhamnose-modified proteins in several bacterial species (8 out of 22 species tested) supports the assumption that the already known regulatory *N*-monorhamnosylation of translation elongation factor EF-P is not an exception. Rather, protein monorhamnosylation seems to be a more common PTM occurring across bacterial family boundaries. Ongoing research on the rhamnoproteome, for which the present study is the starting point, will require determination and (functional) characterization of the specific protein glycosylation sites. This will not only expand our knowledge of the bacterial glycoproteome but might potentially open a new way for combating infectious and antibiotic-resistant bacterial diseases.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Research Training Group GRK2062 (Molecular Principles of Synthetic Biology), research grant LA 3658/1-1 and SFB 1309–325871075. We thank Prof. Dr. Kirsten Jung for fruitful discussions, Prof. Dr. Sonja Verena Albers and Prof. Dr. Dario Leister for providing *Haloferax volcanii*/*Sulfolobus acidocaldarius* and *Synechococcus elongatus*, respectively. Further we thank Lis Winter for her contributing results.

Notes and references

- 1 C. T. Walsh, S. Garneau-Tsodikova and G. J. Gatto Jr., *Angew. Chem., Int. Ed.*, 2005, **44**, 7342–7372.
- 2 D. H. Dube, K. Champasa and B. Wang, *Chem. Commun.*, 2011, **47**, 87–101.
- 3 H. Nothaft and C. M. Szymanski, *Nat. Rev. Microbiol.*, 2010, **8**, 765–778.
- 4 K. F. Jarrell, Y. Ding, B. H. Meyer, S. V. Albers, L. Kaminski and J. Eichler, *Microbiol. Mol. Biol. Rev.*, 2014, **78**, 304–341.
- 5 P. Guerry and C. M. Szymanski, *Trends Microbiol.*, 2008, **16**, 428–435.
- 6 L. L. Lairson, B. Henrissat, G. J. Davies and S. G. Withers, *Annu. Rev. Biochem.*, 2008, **77**, 521–555.
- 7 P. Lafite and R. Daniellou, *Nat. Prod. Rep.*, 2012, **29**, 729–738.
- 8 J. Lassak, F. Koller, R. Krafczyk and W. Volkwein, *Biol. Chem.*, 2019, **400**, 1397–1427.
- 9 S. Grass, A. Z. Buscher, W. E. Swords, M. A. Apicella, S. J. Barenkamp, N. Ozchlewski and J. W. St Geme III, *Mol. Microbiol.*, 2003, **48**, 737–751.
- 10 S. Grass, C. F. Lichti, R. R. Townsend, J. Gross and J. W. St. Geme III, *PLoS Pathog.*, 2010, **6**, e1000919.
- 11 S. Li, L. Zhang, Q. Yao, L. Li, N. Dong, J. Rong, W. Gao, X. Ding, L. Sun, X. Chen, S. Chen and F. Shao, *Nature*, 2013, **501**, 242–246.
- 12 J. S. Pearson, C. Giogha, S. Y. Ong, C. L. Kennedy, M. Kelly, K. S. Robinson, T. W. F. Lung, A. Mansell, P. Riedmaier, C. V. L. Oates, A. Zaid, S. Mühlen, V. F. Crepin, O. Marches, C.-S. Ang, N. A. Williamson, L. A. O'Reilly, A. Bankovacki, U. Nachbur, G. Infusini, A. I. Webb, J. Silke, A. Strasser, G. Frankel and E. L. Hartland, *Nature*, 2013, **501**, 247–251.
- 13 S. El Qaidi, N. E. Scott, M. P. Hays, B. V. Geisbrecht, S. Watkins and P. R. Hardwidge, *Sci. Rep.*, 2020, **10**, 1073.
- 14 J. Lassak, E. C. Keilhauer, M. Fürst, K. Wuichet, J. Gödeke, A. L. Starosta, J.-M. Chen, L. Søgaard-Andersen, J. Rohr, D. N. Wilson, S. Häussler, M. Mann and K. Jung, *Nat. Chem. Biol.*, 2015, **11**, 266–270.
- 15 A. Rajkovic, S. Erickson, A. Witzky, O. E. Branson, J. Seo, P. R. Gafken, M. A. Frietas, J. P. Whitelegge, K. F. Faull, W. Navarre, A. J. Darwin and M. Ibba, *mBio*, 2015, **6**, e00823.

- 16 T. Yanagisawa, H. Takahashi, T. Suzuki, A. Masuda, N. Dohmae and S. Yokoyama, *PLoS One*, 2016, **11**, e0147907.
- 17 R. Krafczyk, J. Macošek, P. K. A. Jagtap, D. Gast, S. Wunder, P. Mitra, A. K. Jha, J. Rohr, A. Hoffmann-Röder, K. Jung, J. Hennig, J. Lassak and R. G. Brennan, *mBio*, 2017, **8**, e01412.
- 18 S. Ude, J. Lassak, A. L. Starosta, T. Kraxenberger, D. N. Wilson and K. Jung, *Science*, 2013, **339**, 82–85.
- 19 X. Long Sun, *Med. Chem.*, 2013, **03**, 1000e106.
- 20 B. Macek, K. Forchhammer, J. Hardouin, E. Weber-Ban, C. Grangeasse and I. Mijakovic, *Nat. Rev. Microbiol.*, 2019, **17**, 651–664.
- 21 J. Ma and G. W. Hart, *Clin. Proteomics*, 2014, **11**, 8.
- 22 X. Li, R. Krafczyk, J. Macošek, Y.-L. Li, Y. Zou, B. Simon, X. Pan, Q.-Y. Wu, F. Yan, S. Li, J. Hennig, K. Jung, J. Lassak and H.-G. Hu, *Chem. Sci.*, 2016, **7**, 6995–7001.
- 23 M. Pan, S. Li, X. Li, F. Shao, L. Liu and H. G. Hu, *Angew. Chem., Int. Ed.*, 2014, **53**, 14517–14521.
- 24 H. Vegad, C. J. Gray, P. J. Somers and A. S. Dutta, *J. Chem. Soc., Perkin Trans. 1*, 1997, 1429–1442.
- 25 A. R. Vartak, S. J. Sucheck, K. A. Wall, A. Quinn and M. F. Mcinerney, *US Patent* 20190231860A1, 2019, The University of Toledo, Toledo, OH (US).
- 26 S. Ficht, R. J. Payne, R. T. Guy and C.-H. Wong, *Chem. – Eur. J.*, 2008, **14**, 3620–3629.
- 27 A. M. P. van Steijn, J. P. Kamerling and J. F. G. Vliegthart, *Carbohydr. Res.*, 1991, **211**, 261–277.
- 28 H. Elferink and C. M. Pedersen, *Eur. J. Org. Chem.*, 2017, 53–59.
- 29 S. Wang, L. Corcilius, P. P. Sharp, A. Rajkovic, M. Ibba, B. L. Parker and R. J. Payne, *Chem. Sci.*, 2017, **8**, 2296–2302.
- 30 H. Herzner and H. Kunz, *Carbohydr. Res.*, 2007, **342**, 541–557.
- 31 H. Tanaka, Y. Iwata, D. Takahashi, M. Adachi and T. Takahashi, *J. Am. Chem. Soc.*, 2005, **127**, 1630–1631.
- 32 B. Yu and H. Tao, *Tetrahedron Lett.*, 2001, **42**, 2405–2407.
- 33 I. Tvaroska and F. R. Taravel, *Adv. Carbohydr. Chem. Biochem.*, 1995, **51**, 15–61.
- 34 P. Messner and U. B. Sleytr, *FEBS Lett.*, 1988, **228**, 317–320.
- 35 A. Verma, M. Schirm, S. K. Arora, P. Thibault, S. M. Logan and R. Ramphal, *J. Bacteriol.*, 2006, **188**, 4395–4403.
- 36 H. Nothaft and C. M. Szymanski, *J. Biol. Chem.*, 2013, **288**, 6912–6920.
- 37 R. K. Dubb, H. Nothaft, B. Beadle, M. R. Richards and C. M. Szymanski, *Glycobiology*, 2019, **30**, 105–119.
- 38 S. Merino and J. M. Tomas, *Int. J. Mol. Sci.*, 2014, **15**, 2840–2857.
- 39 M. Schirm, S. K. Arora, A. Verma, E. Vinogradov, P. Thibault, R. Ramphal and S. M. Logan, *J. Bacteriol.*, 2004, **186**, 2523–2531.
- 40 C. M. Sassetti, D. H. Boyd and E. J. Rubin, *Mol. Microbiol.*, 2003, **48**, 77–84.

4 A β -hairpin epitope as novel structural requirement for protein arginine rhamnosylation

Liubov Yakovlieva, Thomas M. Wood, Johan Kemmink, Ioli Kotsogianni, Franziska Koller, Jürgen Lassak, Nathaniel I. Martin and Marthe T. C. Walvoort (2020). A β -hairpin epitope as novel structural requirement for protein arginine rhamnosylation. Chem Sci. 2021, Advance Article



Cite this: DOI: 10.1039/d0sc05823h

All publication charges for this article have been paid for by the Royal Society of Chemistry

A β -hairpin epitope as novel structural requirement for protein arginine rhamnosylation†

Liubov Yakovlieva,^{†a} Thomas M. Wood,^{†bc} Johan Kemmink,^a Ioli Kotsogianni,^b Franziska Koller,^d Jürgen Lassak,^d Nathaniel I. Martin^{*b} and Marthe T. C. Walvoort^{†*a}

For canonical asparagine glycosylation, the primary amino acid sequence that directs glycosylation at specific asparagine residues is well-established. Here we reveal that a recently discovered bacterial enzyme EarP, that transfers rhamnose to a specific arginine residue in its acceptor protein EF-P, specifically recognizes a β -hairpin loop. Notably, while the *in vitro* rhamnosyltransferase activity of EarP is abolished when presented with linear substrate peptide sequences derived from EF-P, the enzyme readily glycosylates the same sequence in a cyclized β -hairpin mimic. Additional studies with other substrate-mimicking cyclic peptides revealed that EarP activity is sensitive to the method used to induce cyclization and in some cases is tolerant to amino acid sequence variation. Using detailed NMR approaches, we established that the active peptide substrates all share some degree of β -hairpin formation, and therefore conclude that the β -hairpin epitope is the major determinant of arginine-rhamnosylation by EarP. Our findings add a novel recognition motif to the existing knowledge on substrate specificity of protein glycosylation, and are expected to guide future identifications of rhamnosylation sites in other protein substrates.

Received 22nd October 2020
Accepted 4th December 2020

DOI: 10.1039/d0sc05823h

rsc.li/chemical-science

Introduction

Protein glycosylation, an enzymatic process in which a carbohydrate or glycan is covalently added to a specific amino acid residue, is one of the most ubiquitous post-translational modifications in nature.¹ Glycosylation confers specific properties on the acceptor protein, such as increased solubility, protection from degradation, tagging for transport or destruction, interaction with receptors, or functional activation. As a result, protein glycosylation influences a myriad of biological processes in all kingdoms of life.

Protein asparagine *N*-glycosylation is universally present and fairly conserved across species, and it involves the *en bloc* transfer of an oligosaccharide from a lipid-linked carrier to an acceptor protein, catalyzed by a membrane-associated glycosyltransferase, such as eukaryotic OST and prokaryotic PglB.^{2,3}

The requirements for the primary sequence and structural folds are well-established for canonical *N*-linked glycosylation. In general, asparagine residues are modified in so-called “sequons” – recognition sequences of N-X-S/T, with X being any amino acid except proline.⁴ In addition, PglB of *C. jejuni* recognizes an extended sequon of D/E-Z-N-X-S/T (Z and X \neq Pro).³ The X residue (+1) and Ser/Thr residue (+2) have been shown to play an important role in acceptor recognition by glycosyltransferases.⁵ In the co-crystal structure of the bacterial oligosaccharyltransferase PglB, the acceptor peptide was shown to adopt a distinct bound conformation, featuring the recognition sequon in a 180 degree loop.⁶ This structure would be impossible to adopt with proline in the +1 position, explaining the negative selection for Pro in the glycosylation sequon.⁵ The +2 hydroxy amino acid has been shown to contribute to recognition by interacting with the conserved WWD motif in PglB. Interestingly, when bound in the active site, Thr in position +2 has been shown to be engaged in more stabilizing interactions than Ser in position +2 which is reflected in faster glycosylation rates of Thr-containing sequons.^{6,7} Another important aspect of *N*-linked glycosylation is the mechanism of asparagine activation for nucleophilic attack. One of the early explanations implied the importance of the local secondary structure, namely the Asx turn (Fig. 1A).⁸ Within this structure, protonation of the Asn-amide carbonyl by the hydroxyl of the +2 Ser/Thr residue, in combination with deprotonation of nitrogen by an enzymatic base would result in the formation of a reactive imidol species

^aChemical Biology Group, Stratingh Institute for Chemistry, University of Groningen, Groningen, The Netherlands. E-mail: m.t.c.walvoort@rug.nl

^bBiological Chemistry Group, Institute of Biology Leiden, Leiden University, Leiden, The Netherlands. E-mail: n.i.martin@biology.leidenuniv.nl

^cDepartment of Chemical Biology & Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

^dDepartment of Biology I, Microbiology, Ludwig-Maximilians-Universität München Planegg/Martinsried, Germany

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0sc05823h

* Authors contributed equally.



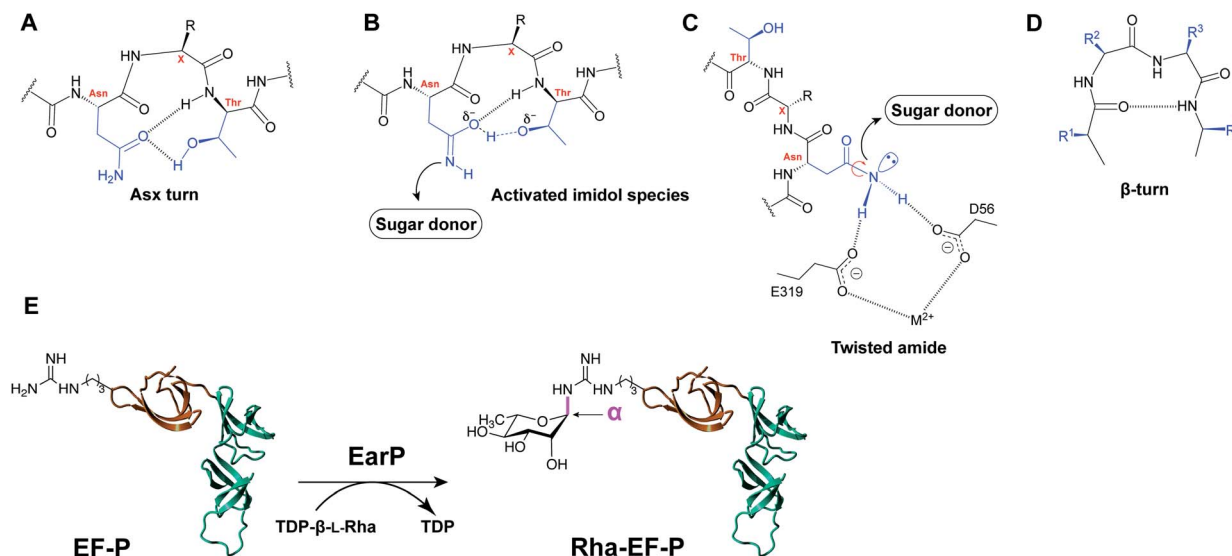


Fig. 1 (A) Asx-turn: proposed secondary structure formed by the sequon amino acids. The amide group of Asn is forming H-bonds with the side chain hydroxyl of the +2 amino acid. (B) Deprotonation by the enzymatic base leads to the formation of the activated imidol species. (C) The twisted amide as alternative hypothesis for Asn activation. (D) β -turn secondary structure, formed by reversing the direction of the chain over four residues, stabilized by interstrand H-bonds. (H-bonds are shown as dashed lines). (E) Rhamnosylation of EF-P by EarP. Domain I of EF-P (amino acids 1–65) is shown in orange.

capable of carrying out the nucleophilic attack (Fig. 1B). An alternative explanation of carboxamide activation in protein glycosylation is the so-called “twisted amide” hypothesis wherein activation of Asn occurs as a result of twisting the nitrogen lone pair out of conjugation with the carbonyl group (Fig. 1C).^{6,9} It is postulated that Asp and Glu residues in the OST active site form H-bonds with amide hydrogens of the acceptor Asn. As a result of this H-bond stabilization, de-conjugation of the nitrogen lone pair electrons occurs, resulting in a more nucleophilic nitrogen primed for attack.

Studies have demonstrated that peptides forced into the Asx turn exhibited increased affinity and faster glycosylation rates in comparison to linear peptides or β -turn peptides (Fig. 1D).¹⁰

Despite the clear three amino acid sequon in the substrate protein, not all predicted N-X-S/T sites are found to be glycosylated which indicates the role of additional recognition elements. A statistical analysis of eukaryotic N-glycosylation sites revealed that there is a preference for β -turns and -bends around the glycosylation site, whereas α -helices are disfavored.¹¹ Additionally, it has been demonstrated that similarly to eukaryotic protein glycosylation, which occurs co-translationally on unfolded polypeptides, bacterial protein glycosylation preferentially takes place on glycosylation sites located in exposed loops and benefits from moderate structure disorder of the acceptor protein.¹² Interestingly, while O-glycosyltransferases do not generally require a specific recognition sequence in their acceptor substrates, several recent examples indicate the preference for properly folded substrate domains, implying a fold recognition mechanism instead.^{13,14}

In addition to the well-established forms of protein glycosylation, novel glycosylation systems have been discovered that are unique to bacteria.¹⁵ Recently, arginine rhamnosylation was identified as a novel type of N-glycosylation.^{16,17} Here the

enzyme EarP transfers a rhamnose moiety from dTDP- β -L-rhamnose (TDP-Rha) to a specific arginine residue in the acceptor translation elongation factor P (EF-P) (Fig. 1E).^{16,18} Arginine glycosylation is a rare modification with only two other examples reported to date, *i.e.* autocatalytic modification of Arg with glucose (Glc) of sweet corn amygdalin,¹⁹ and with N-acetylglucosamine (GlcNAc) of human death receptor domains by the bacterial effector protein NleB.²⁰

Genes associated with the newly discovered EF-P rhamnosylation (*earP*, *efp*, and *rmlABCD* genes for dTDP- β -L-rhamnose (TDP-Rha) donor synthesis) have been identified predominantly in beta- and gamma-proteobacteria, including multiple clinically relevant pathogens, *e.g.* *Pseudomonas aeruginosa*, *Neisseria meningitidis*, and *Bordetella pertussis*.¹⁶ The rhamnose modification has been shown to activate EF-P which alleviates ribosomal stalling during the synthesis of poly-proline stretches in nascent polypeptides.^{21–23} Abolishing rhamnosylation of EF-P in *P. aeruginosa* and *N. meningitidis* led to cellular effects that were detrimental to bacterial fitness and increased susceptibility to antibiotics.^{18,24} It is hypothesized that these severe effects are associated with importance of polyPro-containing proteins and virulence factors of investigated bacterial pathogens for their survival.^{18,24}

Since the discovery of EF-P rhamnosylation in 2015, a number of studies have contributed to an increased understanding of this unique system. The stereochemistry of the α -glycosidic linkage between Arg and Rha was shown by two research groups independently,^{25,26} proving that EarP is an inverting glycosyltransferase. An anti-Arg^{Rha} antibody has also been developed, allowing for facile (*in vitro*) monitoring of arginine rhamnosylation.^{25,27} Several (co)-crystal structures have been reported for EarP (from *P. putida*,²⁷ *N. meningitidis*²⁸ and *P. aeruginosa*²⁹), also in complex with its EF-P substrate, providing insight into the specific amino acid interactions and the catalytic mechanism of rhamnosylation.



As only a single Arg residue in EF-P is modified, an important yet unexplored aspect of this novel glycosylation system is the basis for the observed specificity in recognizing this arginine residue. Previous reports indicate that domain I of EF-P (aa 1–65, Fig. 1E) is sufficient for recognition and rhamnosylation by EarP.²⁷ Domain I, commonly referred to as a “KOW-domain”,²⁷ is a conserved domain in various ribosome-associated proteins involved in transcription and translation,³⁰ and it appears to contain all recognition elements necessary to promote Arg rhamnosylation.^{27–29} A recent study indicates that structural elements are more important than a specific sequon to promote EarP-mediated rhamnosylation.³¹ However, the precise determinants for recognition by EarP are currently not known. Elucidating the necessary substrate recognition elements will allow us to more fully understand this unique bacterial system, an important step towards exploiting bacterial glycosylation systems for the development of novel anti-virulence strategies.³²

Here we report the discovery of a novel β -hairpin recognition element in arginine rhamnosylation of EF-P from *P. aeruginosa*. Using *in vitro* rhamnosylation assays and in-depth NMR studies we demonstrate the importance of this key structural motif in acceptor substrate recognition by EarP. Moreover, we report the shortest peptide fragment known to date to be rhamnosylated by EarP. Next to expanding the current knowledge on structural requirements of protein glycosyltransferases, our results have the potential to inform the development of inhibitors and activity assays to screen for inhibitors for EarP based upon the β -hairpin motif of the EF-P KOW domain.

Results

EarP does not rhamnosylate Arg in linear peptide fragments *in vitro*

A common strategy for studying the activity of *N*-glycosyltransferases *in vitro* is using linear peptide fragments corresponding to their protein substrate. We heterologously expressed the rhamnosyltransferase EarP from *P. aeruginosa* using a previously described procedure,²⁹ and focused our studies on the natural substrate EF-P_{Pa}. As a first step in deciphering the determinants of substrate recognition in arginine rhamnosylation, we investigated a linear peptide fragment comprised of eight amino acids (8mer, Table S1†). As can be seen from the Fig. S1,† this fragment corresponds to the Arg32-containing loop of EF-P. Unexpectedly, this linear peptide did not prove to be a suitable substrate for EarP as no conversion was observed under *in vitro* rhamnosylation conditions (analysis with RP-LCMS). These results suggest that EarP does not rely exclusively on a specific amino acid sequence (primary structure) in the protein substrate for recognition, suggesting that there may be secondary structure requirements involved.

Arginine in an L-Pro–D-Pro-cyclized peptide is rhamnosylated by EarP

As revealed in various structural studies,^{27–29} the acceptor binding site of EarP is unusually large and multiple contacts between active site residues of EarP and amino acids of domain

I of EF-P are necessary for protein substrate recognition. Upon examining the co-crystal structure of EarP and domain I of EF-P from *P. aeruginosa*²⁹ (PDB 6J7M) it is evident that the majority of EF-P residues involved in binding to EarP are located in the β -hairpin with Arg32 at its tip (Fig. 2A). Multiple EarP active site residues are involved in acceptor protein recognition and form both main-chain and side-chain promoted H-bonds, salt bridges, and hydrophobic interactions (Table S2†). As can be seen from Fig. 2A, a selected number of EF-P residues (Arg32, Lys29, Ser30, Asn28, Val36, Phe54, Val53, Lys55) are involved in binding to EarP suggesting that both sequence and shape of the bound motif are recognized, rather than just the target arginine. The β -hairpin secondary structure appears to optimally position Arg32 for binding in the EarP active site. Based on this interaction profile, we decided to explore the importance of secondary structure in Arg-rhamnosylation by preparing peptide mimics of the β -hairpin containing Arg32. Mimics of the β -hairpin secondary structure have been extensively studied over the years, and many structure-inducing templates have been developed.³³ One of the most widely used methods to nucleate a β -hairpin structure is the introduction of an L-proline/D-proline motif.³³ This motif leads to a “kink” in the sequence and brings the strands in close proximity to allow the formation of secondary structure-stabilizing H-bonds between the antiparallel strands.

Based on the sequences of EF-P proteins from *P. aeruginosa*, *Ralstonia solanacearum* (a Gram-negative plant pathogen) and *N. meningitidis*, the corresponding cyclic 11mer peptides depicted in Fig. 2B were prepared using solid-phase peptide synthesis (SPPS) starting from Gly31 loaded onto 2-chlorotrityl resin with the peptides assembled terminating with Arg32 (see ESI† for details). Following mild acid treatment, the side-chain protected peptides were then cyclized by amide bond formation between Gly31 and Arg32 by activation of the C-terminal Gly to avoid any possible racemization. Solution-phase cyclization of these peptides proceeded cleanly after which side chain deprotection and HPLC purification provided the desired cyclic peptides (Table S1†). The peptides were tested in the *in vitro* rhamnosylation reaction with EarP_{Pa} and TDP-Rha, and the rhamnosylated products were identified by an increase in the mass of +146 Da with RP-LCMS. Gratifyingly, the prepared cyclic peptides revealed successful modification by EarP, albeit to varying extents. The best substrate identified was the L-Pro–D-Pro-cyclized 11mer fragment of EF-P from *P. aeruginosa* (11mer_{Pa}, 85% conversion overnight). The extent of rhamnosylation was calculated from ion intensities in the MS spectra and corrected for the relative ionization factor (RIF) values (Table S3†),³⁴ as described in the ESI.† A detailed kinetic analysis of the rhamnosylation of 11mer_{Pa} and native protein substrate EF-P was obtained through a time course study. This analysis revealed that the cyclic 11-mer is indeed rhamnosylated in a time-dependent manner albeit with a lower rate of conversion relative to EF-P (Fig. S2†).

Interestingly, follow-up experiments with both shorter and longer cyclic peptides inspired by the successful 11mer_{Pa} peptide, revealed that the 11mer peptide (nine native amino acids, plus the L/D-Pro motif) was favored as a substrate, as EarP



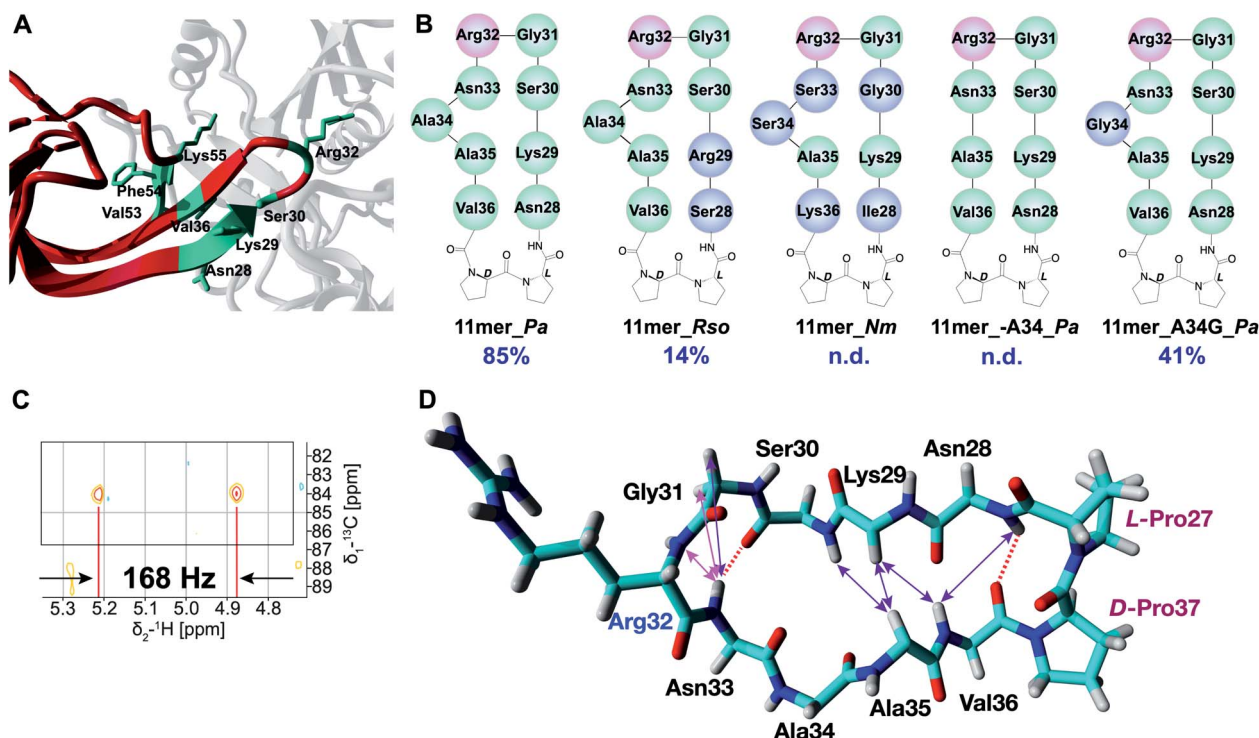


Fig. 2 (A) EarP–EF-P complex from *P. aeruginosa* (generated with YASARA, PDB 6J7M). EarP is depicted in grey, domain I of EF-P is colored red, EF-P residues involved in binding with EarP are in green. (B) Cyclic peptide mimics of the EF-P β -hairpin. Arg32 is shown in pink, altered residues (with respect to 11mer_*Pa*) are shown in blue. (C) Zoom-in of the coupled HSQC spectrum of crude Rha-11mer_*Pa* reaction mixture. JCH = 168 Hz (α -glycosidic bond). (D) Experimentally determined NOE signals that are indicative of a β -hairpin structure are mapped on the 11mer L-Pro–D-Pro_*Pa* fragment (from 3OYY crystal structure). Hydrogen bonds inferred from the NH temperature studies are shown as red dashed lines. NOE signals are shown as double-ended arrows (magenta: medium NOE; purple: weak NOE).

exhibited no activity towards the shorter (7mer_*Pa*) and only low levels of conversion (14%) were achieved with the longer (15mer_*Pa*) fragment. Detailed NMR analysis of the rhamnosylated 11mer_*Pa* peptide (Rha-11mer_*Pa*, prepared enzymatically) revealed that the rhamnose-arginine glycosidic linkage was formed in an α -stereoselective fashion (Fig. 2C and S3[†]), identical to the linkage described for Arg-rhamnosylation of EF-P.^{25,26} Interestingly, our initial attempts at purifying Rha-11mer_*Pa* using anion-exchange under basic conditions, led to full epimerization to the β -linked Arg-Rha species (Fig. S4[†]), in accordance with previous observations by Payne and co-workers.²⁶

EarP allows for sequence promiscuity, but it is sensitive to the cyclization strategy

Encouraged by the successful rhamnosylation of 11mer_*Pa*, we set out to map the promiscuity of EarP for the amino acid sequence surrounding Arg32. To this end, the 11mer fragments of EF-P sequences from *R. solanacearum* and *N. meningitidis* were tested in the rhamnosylation reaction with EarP from *P. aeruginosa* (Fig. 2B). Interestingly, the 11mer_*Rso* peptide with two amino acid mutations compared to 11mer_*Pa* showed low levels of conversion (14% overnight), whereas the 11mer_*Nm* peptide bearing five mutations was not accepted as a substrate. While these cyclic 11mer peptides do not show high conversion

with EarP_*Pa*, this does not exclude the possibility that they may be substrates for their associated native enzymes EarP_*Rso* and EarP_*Nm*, respectively.

The co-crystal structure of EF-P bound to EarP revealed that Ala34 in EF-P undergoes a significant conformational change.²⁹ It appears that upon binding to EarP, the “bulge” formed by Ala34 is significantly reduced, leading to a more narrowly shaped and structured loop than that in free EF-P. To investigate whether Ala34 and the concomitant conformational movement are important for binding, the residue was completely removed in peptide 11mer_A34_*Pa*. Interestingly, this substrate was not rhamnosylated by EarP, suggesting that the Ala34 bulge is important for recognition. Whereas the co-crystal structure suggests that Ala34 is not directly involved in binding to EarP,²⁹ it is reasonable to assume that this residue contributes to shaping the β -hairpin, which in turn positions Arg32 in the active site. This was further corroborated by substituting Ala34 with glycine (11mer_A34G_*Pa*), a smaller and more flexible amino acid. This mutant retained its role as a substrate for EarP, albeit with reduced efficiency (41% conversion).

Next, alternate cyclization strategies were compared to assess their impact on recognition by EarP (Table S1, Fig. S5[†]). Specifically, peptides of varying lengths, based upon the same key EF-P amino acid substrate sequence, were synthesized and cyclized using CLIPS-³⁵ (Chemical Linkage of Peptides onto Scaffolds) and disulfide strategies³⁶ to introduce



conformational rigidity and flexibility, respectively. In addition, we also prepared linear peptides containing the same EF-P sequence with Trp and Phe residues to introduce a so-called “tryptophan zipper” motif known to induce interstrand H-bonding as another means of generating a β -turn mimic.³⁷ Notably, the majority of the trp-zip designs (especially longer sequences of 13mers, 17mers and 18mers) led to almost immediate protein precipitation when incubated with EarP, indicating that the hydrophobicity of these substrates mimics is not compatible with forming and maintaining a soluble complex with the enzyme.

NMR studies reveal β -hairpin formation in the active peptide substrates

Suitable EarP substrates were further investigated with several NMR techniques to gain understanding of the secondary structure, with the most pronounced effects depicted in the Fig. 2D. By performing temperature studies of the chemical shift of amide-hydrogens, two NH groups (Asn28 and Asn33) with low temperature coefficients were identified (Table S4[†]), indicative of the β -hairpin-forming interstrand hydrogen bonds (dashed lines). Additionally, a number of characteristic Nuclear Overhauser Effect (NOE) signals (medium and weak) were observed, indicative of the β -hairpin secondary structure (Table S5[†]). Interestingly, the majority of the observed NOE signals that are characteristic of a β -hairpin structure (Asn28NH-Val36NH, Lys29HA-Val36NH, Lys29HA-Ala35HA, Ser30NH-Ala35HA) are clustered in close proximity to the L-Pro-D-Pro motif, which supports the ability of this template to induce the twist of the peptide structure and consequently bring the strands together for H-bonding. As can be seen from the Fig. 2D this structure-inducing effect of the L-Pro-D-Pro template tapers off towards the Ala35 residue, as the bulge most likely does not allow strands to come close together. Finally, the NOE signals and another H-bond (Asn33) re-appear around Arg32, presumably induced by the Gly31-Arg32 β -turn.

Compared to the original 11mer_{Pa} peptide, the 11mer_{-A34G}_{Pa} variant features only one of the two H-bonds (NH of Asn28, Table S4[†]) and several NOE signals are missing (K29HA-V36NH, S30NH-A35HA, Table S5[†]). It was also shown to exhibit more flexibility, presumably due to the inclusion of the inherently more flexible glycine in the loop. In the case of 11mer_{-A34}_{Pa}, that showed no conversion to product, only very few structural elements were preserved (sequential NOE signals of P27HA-N28NH and R32NH-N33NH, Table S5[†]), and the structure was found to lack one of the two H-bonds (NH of Asn33, Table S4[†]) present in the 11mer_{Pa} that enable the formation of the β -hairpin structure. The 11mer_{Rso} variant structure has both H-bonds found in 11mer_{Pa} (NH of Ser28 and Asn33, Table S4[†]) and more of β -hairpin characteristic NOEs, although S28NH-V36NH was absent (Table S5[†]). On the other hand, it was not possible to determine the structure of the 11mer_{Nm} peptide, as it appears to be a mixture of several structures due to *cis/trans* isomerization of Pro residues. It is therefore difficult to conclude whether absence of conversion for 11mer_{Nm} peptide stems from the sequence variation or lack of structure. NMR

analysis of the structures of the linear peptide, CLIPS-, disulfide- and Trp-rich peptides showed that these peptides tend to form disordered structures and show close to none of the β -sheet structure (see ESI[†] NMR spectra of these peptides).

EarP has a low binding affinity for 11mer_{Pa}

In order to study the binding between EarP and 11mer_{Pa} peptide different experimental methods were employed. Saturation-Transfer Difference (STD) NMR is a widely employed strategy to observe binding between proteins and ligands, even at low binding affinities.³⁸ However, several attempts at measuring STD-NMR for the EarP-11mer_{Pa} peptide complex, in the presence and absence of TDP, provided no definitive proof of binding affinity (data not shown). Similarly, attempts to identify the amino acid residues of 11mer_{Pa} involved in binding to EarP with (TR)-NOE measurements proved unsuccessful (data not shown), presumably due to the low affinity between the protein and the peptide substrate. In a final attempt to quantify the binding affinity of EarP for the best substrate 11mer_{Pa}, Isothermal Titration Calorimetry (ITC) studies were performed. Tight binding of the native protein substrate EF-P was measured with a binding constant K_d of 473 ± 94 nM for EarP (Fig. 3A), in agreement with reported values.²⁹ Conversely, binding of 11mer_{Pa} to EarP could not be detected by ITC using a range of increasing concentrations (Fig. 3B). Even a displacement experiment, in which EF-P was titrated into a solution of pre-formed 11mer_{Pa}-EarP-TDP complex, did not reveal binding of 11mer_{Pa} to EarP (data not shown). The absence of clear results from the STD-NMR and ITC suggest that the binding affinity of EarP for 11mer_{Pa} is too low (mM range or higher) to clearly visualize and quantify using such techniques.

The rhamnosylated β -hairpin peptide is recognized by anti-Arg^{Rha} antibodies

Finally, we investigated the structural similarity between rhamnosylated EF-P (Rha-EF-P), and the *in vitro* rhamnosylated cyclic peptide (Rha-11mer_{Pa}). In previous studies, antibodies against the Rha-Arg modification were raised using synthesized linear Rha-peptide coupled to BSA.²⁷ To determine binding of the anti-Arg^{Rha} antibodies to the rhamnosylated cyclic peptide, we performed a dot blot affinity assay using freshly rhamnosylated 11mer_{Pa} (Rha-11mer_{Pa}) and Rha-EF-P as a control. As can be seen in Fig. 3C, a clear fluorescent signal belonging to the formation of Rha-11mer_{Pa} (C2) was observed well into low micromolar concentration (up to 4 μ M). As can be seen from the Fig. 3C, a similar, albeit more intense signal was obtained for the native substrate (Rha-EFP, C1). At the same time, identical concentrations of non-modified substrates were not detected in this assay (Fig. 3, C3 and C4). Interestingly, a similar dot blot affinity study with the β -linked Rha-11mer_{Pa} product, serendipitously obtained after complete anomerization during anion exchange chromatography under basic conditions (*vide supra*), revealed very little binding to anti-Arg^{Rha} antibodies (Fig. S6[†]). This establishes the selectivity of the antibodies for the α -anomeric linkage in the Arg-Rha glycosidic bond.



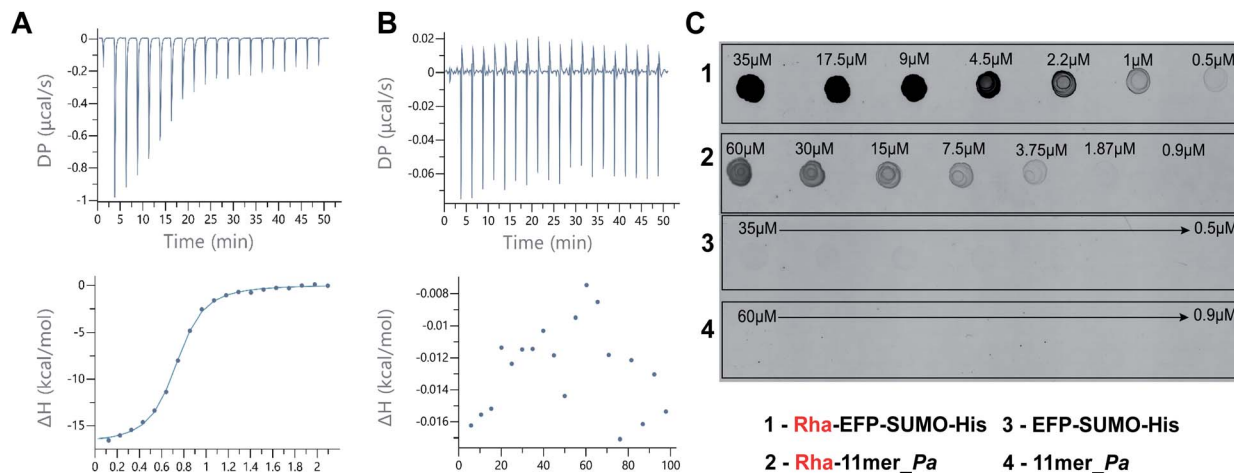


Fig. 3 (A) ITC studies reveal strong binding between EF-P (233 μM) and EarP (20 μM) in the presence of 60 μM TDP. (B) No apparent binding was observed between 11mer_*Pa* (10 mM) and EarP (20 μM) with ITC in the presence of 60 μM TDP. (C) Dot blot affinity assay of a two-fold serial dilution of 35 μM Rha-EFP (C1) and 60 μM Rha-11mer_*Pa* (C2) binding to the anti-Arg^{Rha} antibody, and visualized using anti-rabbit Alexa488 secondary antibody. Non-rhamnosylated substrates are not recognized by the anti-Arg^{Rha} antibody (C3 and C4).

Discussion

As protein glycosylation is a non-templated process, it is generally governed by co-localization of the necessary enzymes and substrates, and specific motifs in the protein substrate that dictate glycosylation. A thorough understanding of the molecular basis underlying site-specific glycosylation is paramount to predicting protein glycosylation, and to combining that knowledge with functional effects to fully understand the impact of protein glycosylation on health and disease. As novel bacterial protein glycosylation systems are identified at a steady pace, knowledge of the chemical and structural requirements at play is also increasing. Bacterial glycoproteins are often involved in cell homeostasis and the initiation of infection, therefore the enzymes involved in production of bacterial glycoproteins are interesting targets for the development of novel antimicrobial strategies.

Here we report the successful rhamnosylation of a cyclic peptide fragment, 11mer_*Pa*, designed to mimic the native EF-P substrate of the EarP rhamnosyltransferase. This 11mer L-Pro-D-Pro-*Pa* peptide is the smallest fragment of EF-P reported to date to be successfully rhamnosylated by EarP. The combination of enzyme activity assays and NMR structural analysis reveals that activity is directly linked to the propensity of the cyclic peptide to form a structured β -hairpin motif. Moreover, the glycosidic linkage is formed in an α -stereoselective fashion, analogous to the native Rha-EF-P and the resulting α -Rha-11mer epitope is successfully recognized by anti-Arg^{Rha} antibody. The results of the various 11mer peptide mutants and different cyclization strategies indicate that substrate recognition by the EarP rhamnosyltransferase is highly dependent on the conformation of the substrate and that some sequence variation is tolerated. Developing a successful structural mimic of EF-P beta-hairpin has proved to be challenging, as the majority of the strategies for secondary structure stabilization investigated

led to inactive substrates. It would appear that many of the peptide mimics generated in this project are not able to recapitulate important enzyme-substrate contacts in the enzyme active site. In this regard, our study reveals that both specific amino acid residues and an optimal secondary structure are important for recognition by EarP.

The success of the substrate mimics bearing the L-Pro-D-Pro motif may be attributed to the right-handed twist that brings the strands together to allow the formation of H-bonds that stabilize the secondary structure.³⁹ In contrast, CLIPS-bearing peptides may be too bulky to fit in the narrow active site of EarP, whereas disulfide cyclization may induce too much rotational freedom and a less defined β -hairpin. Trp zippers, although widely reported in literature to form β -sheet structures, did not induce measurable secondary structures in the peptides in our investigation, and proved to be too hydrophobic to be suitable EarP substrates. Alternative methods of secondary structure stabilization, such as *N*-methylation to reduce flexibility of the structure may be explored in the future.⁴⁰

As apparent from the co-crystal structure, there are many points of contact between EarP and EF-P that are likely to contribute to substrate recognition (Fig. 2A). Interestingly, while the majority of the EarP-EF-P contacts resides around the Arg32-containing β -hairpin that we chose to mimic, several residues of the neighboring β -strand have also been shown to be involved in binding to EarP.²⁹ Consequently, the low affinity of 11mer_*Pa* for EarP may be increased by extending the current scaffold to include more contact points. Notably, residue Lys55 is a promising residue to consider, particularly as replacing it by alanine resulted in a 200-fold decrease in affinity for EarP.²⁹ One strategy for improving the affinity for EarP and the rates of conversion of future peptidomimetic substrates might include incorporation of either a fragment of the third strand, or a single Lys residue into the peptide. While it can be expected that expanding the structure to include the β -sheet increases



the substrate affinity, β -sheets and other secondary structures are difficult to design as isolated motifs. They are stabilized by a larger protein structure they are found in, and without it, they tend to misfold and aggregate.

Conclusion

Whereas the structural determinants for asparagine-linked protein glycosylation are largely based on the primary sequence (consensus sequence), our results suggest that for bacterial arginine-rhamnosylation a specific secondary structural motif is required. The clear importance of secondary structure, and more specifically a β -hairpin motif, as the minimal structural epitope for protein glycosylation may be a unique characteristic of this class of enzymes. Taken together these findings show that the propensity of the (cyclic) peptides to form a β -hairpin structure is an important substrate prerequisite for EarP rhamnosyltransferase and can be directly correlated to activity of the enzyme towards various peptide substrates. This work provides important insights into the recognition motifs for bacterial arginine rhamnosylation which will be useful for the development of future substrate mimics or structure-guided design of peptide inhibitors.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank Dr Ralph Krafczyk for help with optimization of the antibody-based assay. This work was financially supported by the Dutch Organization for Scientific Research (VENI 722.016.006) and the European Union through the Rosalind Franklin Fellowship COFUND project 60021 (both to M. T. C. W.), and the European Research Council (ERC consolidator grant to N. I. M., grant agreement no. 725523) and J. L. is grateful for funding from the Deutsche Forschungsgemeinschaft (LA 3658/1-1).

References

- G. A. Khoury, R. C. Baliban and C. A. Floudas, *Sci. Rep.*, 2011, **1**, 90.
- D. J. Kelleher and R. Gilmore, *Glycobiology*, 2006, **16**, 47–62.
- C. M. Szymanski, R. Yao, C. P. Ewing, T. J. Trust and P. Guerry, *Mol. Microbiol.*, 1999, **32**, 1022–1030.
- A. Dell, A. Galadari, F. Sastre and P. Hitchen, *Int. J. Microbiol.*, 2010, 148178.
- A. Yan and W. J. Lennarz, *J. Biol. Chem.*, 2005, **280**, 3121–3124.
- C. Lizak, S. Gerber, S. Numao, M. Aebi and K. P. Locher, *Nature*, 2011, **474**, 350–355.
- E. Bause, *Biochem. Soc. Trans.*, 1984, **12**, 514–517.
- B. Imperiali, K. L. Shannon, M. Unno and K. W. Rickert, *J. Am. Chem. Soc.*, 1992, **114**, 7944–7945.
- C. Lizak, S. Gerber, G. Michaud, M. Schubert, Y.-Y. Fan, M. Bucher, T. Darbre, M. Aebi, J.-L. Reymond and K. P. Locher, *Nat. Commun.*, 2013, **4**, 2627.
- B. Imperiali, *Acc. Chem. Res.*, 1997, **30**, 452–459.
- A. J. Petrescu, A. L. Milac, S. M. Petrescu, R. A. Dwek and M. R. Wormland, *Glycobiology*, 2004, **14**, 103–114.
- J. M. Silverman and B. Imperiali, *J. Biol. Chem.*, 2016, **291**, 22001–22010.
- Z. Li, M. Fischer, M. Satkunarajah, D. Zhou, S. G. Withers and J. M. Rini, *Nat. Commun.*, 2017, **8**, 185.
- Z. Li, K. Han, J. E. Pak, M. Satkunarajah, D. Zhou and J. M. Rini, *Nat. Chem. Biol.*, 2017, **13**, 757–763.
- H. Nothhaft and C. M. Szymanski, *Curr. Opin. Chem. Biol.*, 2019, **53**, 16–24.
- J. Lassak, E. C. Keilhauer, M. Fürst, K. Wuichet, J. Gödeke, A. L. Starosta, J.-M. Chen, L. Søgaard-Andersen, J. Rohr, D. N. Wilson, S. Häussler, M. Mann and K. Jung, *Nat. Chem. Biol.*, 2015, **11**, 266–270.
- D. Gast, F. Koller, R. Krafczyk, L. Bauer, S. Wunder, J. Lassak and A. Hoffmann-Röder, *Org. Biomol. Chem.*, 2020, **18**, 6823–6828.
- A. Rajkovic, S. Erickson, A. Witzky, O. E. Branson, J. Seo, P. R. Gafken, M. A. Frietas, J. P. Whitelegge, K. F. Faull, W. Navarre, A. J. Darwin and M. Ibba, *mBio*, 2015, **6**, e00823.
- D. G. Singh, J. Lomako, W. M. Lomako, W. J. Whelan, H. E. Meyer, M. Serwe and J. W. Metzger, *FEBS Lett.*, 1995, **376**, 61–64.
- S. Li, L. Zhang, Q. Yao, L. Li, N. Dong, J. Rong, W. Gao, X. Ding, L. Sun, X. Chen, S. Chen and F. Shao, *Nature*, 2013, **501**, 242–246.
- S. Ude, J. Lassak, A. L. Starosta, T. Kraxenberger, D. N. Wilson and K. Jung, *Science*, 2013, **339**, 82–85.
- L. Peil, A. L. Starosta, J. Lassak, G. C. Atkinson, K. Virumäe, M. Spitzer, T. Tenson, K. Jung, J. Remme and D. N. Wilson, *Proc. Natl. Acad. Sci. U.S.A.*, 2013, **110**, 15265–15270.
- J. Lassak, D. N. Wilson and K. Jung, *Mol. Microbiol.*, 2016, **99**, 219–235.
- T. Yanagisawa, H. Takahashi, T. Suzuki, A. Masuda, N. Dohmae and S. Yokoyama, *PLoS One*, 2015, **11**, e0147907.
- X. Li, R. Krafczyk, J. Macošek, Y.-L. Li, Y. Zou, B. Simon, X. Pan, Q.-Y. Wu, F. Yan, S. Li, J. Hennig, K. Jung, J. Lassak and H.-G. Hu, *Chem. Sci.*, 2016, **7**, 6995–7001.
- S. Wang, L. Corcilius, P. P. Sharp, A. Rajkovic, M. Ibba, B. L. Parker and R. J. Payne, *Chem. Sci.*, 2017, **8**, 2296–2302.
- R. Krafczyk, J. Macošek, P. K. A. Jagtap, D. Gast, S. Wunder, P. Mitra, A. K. Jha, J. Rohr, A. Hoffman-Röder, K. Jung, J. Hennig and J. Lassak, *mBio*, 2017, **8**, e014112.
- T. Sengoku, T. Suzuki, N. Dohmae, C. Watanabe, T. Honma, Y. Hikida, Y. Yamaguchi, H. Takahashi, S. Yokoyama and T. Yanagisawa, *Nat. Chem. Biol.*, 2018, **14**, 368–374.
- C. He, N. Liu, F. Li, X. Jia, H. Peng, Y. Liu and Y. Xiao, *J. Bacteriol.*, 2019, **201**, e00128.
- N. C. Kyrpides, C. R. Woese and C. A. Ouzounis, *Trends Biochem. Sci.*, 1996, **21**, 425–426.
- W. Volkwein, R. Krafczyk, P. K. A. Jagtap, M. Parr, E. Mankina, J. Macošek, Z. Guo, M. J. L. J. Fürst, M. Pfab,



- D. Frishman, J. Hennig, K. Jung and J. Lassak, *Front. Microbiol.*, 2019, **10**, 1148.
- 32 Q. Lu, S. Li and F. Shao, *Trends Microbiol.*, 2015, **23**, 630–641.
- 33 J. A. Robinson, *Acc. Chem. Res.*, 2007, **41**, 1278–1288.
- 34 W. Kightlinger, L. Lin, M. Rosztoczy, W. Li, M. P. DeLisa, M. Mrksich and M. C. Jewett, *Nat. Chem. Biol.*, 2018, **14**, 627–635.
- 35 P. Timmerman, J. Beld, W. C. Puijk and R. H. Melen, *ChemBioChem*, 2005, **6**, 821–824.
- 36 C. M. Santiveri, E. León, M. Rico and M. A. Jiménez, *Chem.–Eur. J.*, 2008, **14**, 488–499.
- 37 A. G. Cochran, N. J. Skelton and M. A. Starovasnik, *Proc. Natl. Acad. Sci. U.S.A.*, 2001, **98**, 5578–5583.
- 38 A. Viegas, J. Manso, F. L. Nobrega and E. J. Cabrita, *J. Chem. Educ.*, 2011, **88**, 990–994.
- 39 J. A. Robinson, *Chimia*, 2013, **67**, 885–890.
- 40 M. P. Gimeno, A. Glas, O. Koch and T. N. Grossman, *Angew. Chem., Int. Ed.*, 2015, **54**, 8896–8927.



5 Analysis of substrate specificity of the rhamnosyltransferase EarP reveals an unprecedented mode of target recognition in bacterial N-glycosylation

Manuscript

Ralph Krafczyk*¹, Franziska Koller*¹, Laura Lindenthal², Alina Sieber¹, Bernd Simon³, Jannis Brehm⁴, Daniel Gast⁵, Amit Kumar Jha⁶, Jürgen Rohr⁶, Anja Hoffmann-Röder⁵, Janosch Hennig³, Kirsten Jung¹ and Jürgen Lassak¹.

¹ Department of Biology I, Microbiology, Ludwig-Maximilians-Universität München, Planegg/Martinsried, Germany

² MPI of biochemistry, Planegg/Martinsried, Germany

³ Structural and Computational Biology Unit, EMBL Heidelberg, Heidelberg, Germany

⁴ Mikrobiologie und Weinforschung, Universität Mainz, Mainz, Germany

⁵ Department of Chemistry, Ludwig-Maximilians-Universität München, München, Germany

⁶ University of Kentucky College of Pharmacy, Lexington, Kentucky, USA

Abstract

Protein glycosylation is a ubiquitous post-translational modification that is involved in various important processes within and outside of the cell. N-linked glycosylation is generally carried out in the endoplasmic reticulum of eukaryotes and the periplasm of prokaryotes. Most proteins are modified at a highly conserved amino acid sequence. N-protein glycosyltransferases that target specific enzymes are thought to exhibit a high degree of substrate specificity. Here we characterized substrate specificity of the EF-P arginine rhamnosyltransferase EarP with respect to its acceptor and donor substrates. We show that target recognition by EarP is not sequence specific. Instead, we found that a structural motif – the strand-loop-strand motif of the EF-P acceptor loop – is sufficient for modification by EarP. Synthetic proteins carrying this motif were successfully modified by the rhamnosyltransferase *in vivo*. We also show that – upon overexpression – EarP displays an higher degree of acceptor substrate promiscuity; modifying even C-terminal arginine residues. We also found that EarP is capable of transferring sugars from donor substrates structurally similar to TDP- β -L-rhamnose and should be capable of utilizing UDP-activated sugars for transfer. Our results demonstrate that protein rhamnosylation might be a more widespread intracellular modification than previously thought. They also show that monoglycosyltransferases could also be used for synthetic protein modification within the cell, especially for synthesis of initial amino acid-sugar linkages.

Keywords: TDP-Phenol, glycoengineering, *Pseudomonas aeruginosa*, *Neisseria meningitidis*

Introduction

Glycosylation affects several crucial properties of proteins and is involved in the regulation of multiple cellular processes. Glycosyltransferases (GTs) constitute a group of ubiquitous enzymes that catalyze the synthesis of glycoconjugates by transferring an activated sugar residue to a corresponding acceptor. Glycosylation of proteins is one of the most common post-translational modification and is found to be structurally highly diverse. Usually, glycans are attached to the amid nitrogen of an asparagine side chain (N-glycosylation) or to the oxygen of a hydroxyl group of serine and threonine side chains (O-glycosylation). Eukaryotic and prokaryotic glycosylation mainly differs in structure and composition of glycans: eukaryotic diversity is based on monosaccharide building blocks, linkages, and branching patterns whereas prokaryotic glycans contain more than one hundred sugars (1).

Understanding of the function of GTs is of interest from the perspective of a potential drug target as well as a key enzyme for the development of different therapeutics and vaccines which often depend on regio- and stereospecifically attached sugar moieties (2-5). To achieve desired precise glycosylation, which provide different properties and functions like solubility, folding, and resistance to proteases, prokaryotic glycoengineering has opened a new route (6-8). As prokaryotic glycans differ remarkably from eukaryotic, engineering of prokaryotic glycosylation is necessary to allow the expression of glycosylated therapeutics. Intensive studies on substrate specificity and recognition motive on the acceptor side of different glycosylation systems laid the basis for glycoengineering.

In the last years, great efforts had been made with the N-linked glycosylation system (*pgl*) of *Campylobacter jejuni*, which is similar to the eukaryotic pathway (9). Here, eukaryotic glycans are built by either enzymatic remodeling of the prokaryotic glycan structures on proteins or assembly of synthetic glycosylation pathways. The initial step is to build the key linkage of eukaryotic N-glycosylation; N-acetylglucosamine (GlcNac) attached to asparagine. This is achieved by either *de novo* synthesis or a highly modified *pgl* operon. Further, the required glycan is synthesized on a lipid carrier and then flipped to the periplasm. The final transfer to the target protein is mediated by PglB, the oligosaccharyltransferase (OST) of the Pgl system. These approaches provide defined glycan structures, though it often needs further *in vitro* modification of the glycan structure and is complicated and expensive (10). A more direct way to build eukaryotic-like glycans is to use bottom-up construction. A promising study combined yeast biosynthesis enzymes, which produce the eukaryotic precursor glycan, with PglB, transferring the glycan to the target protein (11). The major disadvantage of this approach is the strikingly low transfer efficiency which is lower than one percent.

Due to intensive research in engineering of N-linked glycosylation, various glycosylations of non-target proteins are now possible. Investigation of alternative glycosylation pathways will further help to expand the substrate promiscuity and target sites. An interesting candidate is the GT EarP. In 2015, it was reported that EarP activates bacterial translation elongation factor P (EF-P) by monorhamnosylation of an arginine residue at the tip of the protein (12). EarP uses the activated sugar donor TDP- β -L-Rhamnose (TDP-Rha) as substrate and transfers the sugar directly to the arginine residue of the acceptor protein (13, 14). The glycosylation reaction takes place in the cytosol which makes EarP an interesting target for engineering. The subcellular location of PglB in the periplasm requests translocation of donor and acceptor substrates to the intermembrane space. Cytosolic glycosylation would facilitate present synthetic glycosylation approaches and additionally increase the range of acceptor targets. Further, monoglycosylation of EarP could be useful in bottom-up constructions thereby avoiding polysaccharide trimming. Thanks to work of various research groups EarP has turned into one of the best characterized GTs from a structural perspective (14-16) providing a unique opportunity for glycoengineering. To assess the potential of EarP as a synthetic GT, the natural substrate specificity with respect to donor and acceptor was investigated in this study.

Results and discussion

EarP recognizes a strand loop strand structural motif with high sequence variability as acceptor substrate

Generally, N-linked glycosylation occurs on highly specific sequence motifs. However, previous studies showed that subtle changes on elongation factor P from *Escherichia coli* are sufficient to make it a target of glycosyltransferase EarP (17). Although that might seem obvious, one has to note that EF-P_{Eco} belongs to a phylogenetic subgroup that is distinct from the one being activated by EarP: *E. coli* EF-P and its ortholog form from *Pseudomonas putida* KT2440 share only 30 % identity at the sequence level. Given their structural similarities on the other hand (18) suggest that sequence recognition is not the recognition mode for EarP. We were therefore interested whether changes in the *P. putida* EF-P acceptor loop composition would affect modification by EarP. Thus, we generated variants of *P. putida* EF-P that harboured the acceptor amino acid arginine in any position of the acceptor loop (29-31, 33-35) while replacing arginine at position 32 by alanine.

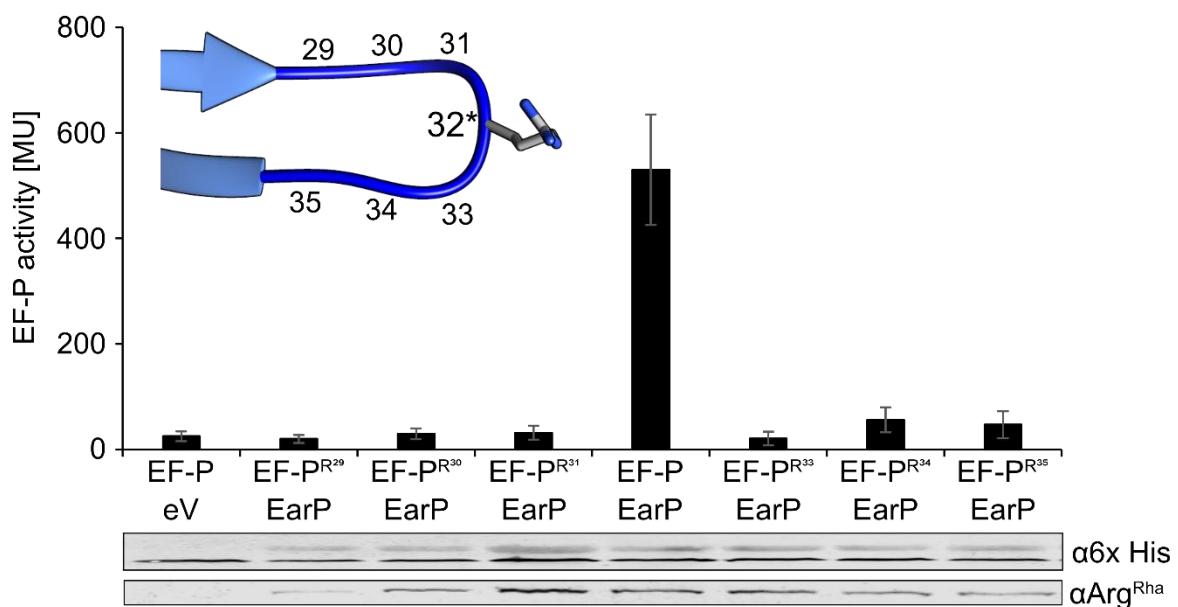


Figure 1: **Arginine walking challenges sequence context model.** Top: Test for functionality of EF-P variants. *In vivo* activities were determined by measuring the β -galactosidase activities of *E. coli* MG1655 $P_{\text{cadBA}}::\text{lacZ} \Delta\text{efp}$ heterologously expressing pBAD24 *P. putida* EF-P variants, harbouring the acceptor amino acid arginine in any position of the acceptor loop (29-31, 33-35) while replacing arginine at position 32 by alanine. Additionally, strains encoded for the *earP* on a pBAD33 vector or sole empty vector. All strains were grown o/n in LB pH 5.8 and activity is given in Miller Units (MU). Means of three independent measurements are shown. Standard deviations from three independent experiments were determined. Bottom: Test for modification of EF-P variants. SDS-PAGE and subsequent Western blot analysis of strains listed in A) was performed. Expression of EF-P variants was verified using 0.1 $\mu\text{g/ml}$ $\alpha 6\text{x His}$, while rhamnosylation was detected using 0.25 $\mu\text{g/ml}$ $\alpha\text{Arg}^{\text{Rha}}$.

Interestingly, while none of these EF-P variants was capable of alleviating ribosome pausing (Figure 1A), all of them showed clear modification after co-expression with the glycosyltransferase EarP (Figure 1B). This result shows that unlike other N-

glycosyltransferases (19), EarP does not target a fixed sequence of amino acids. Previous studies indicated that regions beyond the acceptor loop of EF-P might get into contact with EarP (15, 16, 20). These contacts especially involve amino acids within beta strands β -3 and β -4 that precede and follow the acceptor loop. In addition, studies on the AIDA GT demonstrated recognition of a structural element that – much like the EF-P acceptor loop – consists of strand-loop-strand motif (21). We therefore wondered, whether the larger context of the KOW-like EF-P domain I is involved in directing the modification enzyme towards the target site without specific recognition of a sequence motif. We constructed variants of *E. coli* EF-P in which we replaced the unstructured region of domain III by the *P. putida* EF-P acceptor loop together with the surrounding beta strands (Figure 2A).

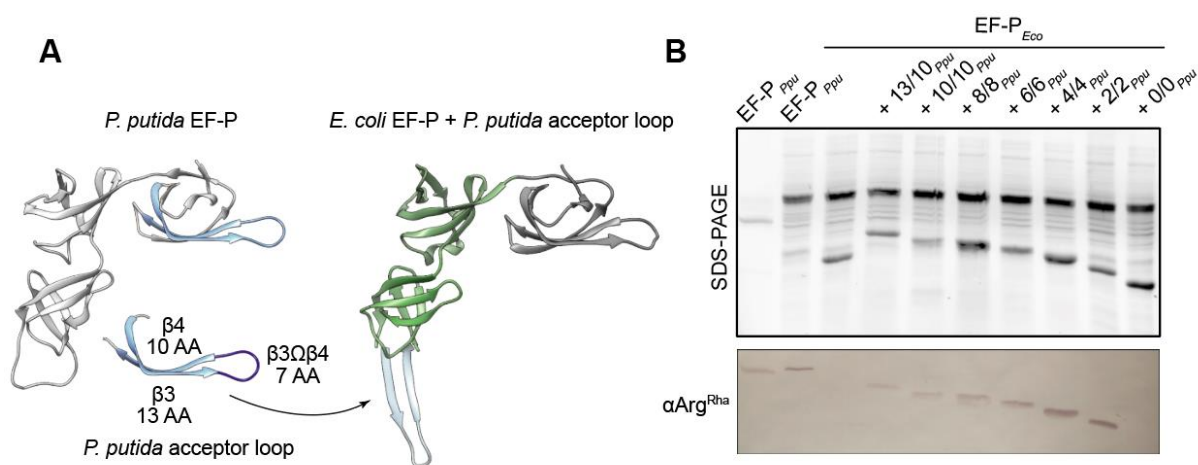


Figure 2: **Sequential truncation of EF-P_{Ppu} peptides reveals minimal acceptor motif.** **A)** Schema illustrating the structure of *P. putida* EF-P (left) and *E. coli* EF-P (right) carrying the EF-P_{Ppu} acceptor loop (blue). The *P. putida* EF-P acceptor loop (Ω) including different portions of the surrounding β -strands 3 (β 3) and 4 (β 4) replaced the unstructured region of domain III of EF-P_{Eco}. **B)** Test for modification of EF-P_{Eco} carrying truncated versions of the EF-P_{Ppu} acceptor loop. O/n cultures of *E. coli* BW25113 cells heterologously expressing EF-P_{Ppu} or EF-P_{Eco}- EF-P_{Ppu} acceptor loop constructs (pBAD24) together with *earP* (pBAD33) were subjected to SDS-PAGE (top) and subsequent Western blot (bottom) analysis using 0.25 μ g/ml α Arg^{Rha}.

We chose this construct to facilitate formation of the strand-loop-strand motif by providing a surrounding protein scaffold. The initial construct contained all 13 amino acids of β -3 and all 10 amino acids of β -4 and showed clear modification by EarP after immunostaining using rhamnosylarginine specific antibodies (Figure 2A, right) (13, 14). To narrow down relevant regions for target recognition within this construct we successively shortened both of the β -strands. Interestingly, all constructs that contained at least two amino acids of beta strands β -3 and β -4 showed a modification signal after immunoblotting. Only the construct that contained the EF-P acceptor loop without any of the surrounding amino acids did not show a signal of modification (Figure 2A, right). This is in line with observations that EarP relies on structured peptides for target recognition (20).

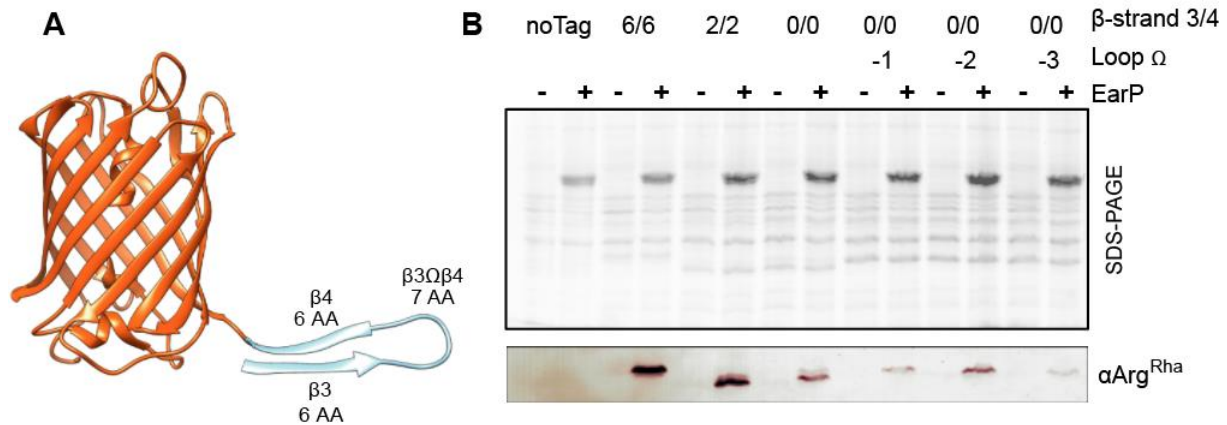


Figure 3: **Truncated EF-P loop fusions lead to rhamnosylation of non-target proteins.** **A)** Schema illustrating the structure of the mCherry (orange) carrying the EF-P acceptor loop (blue). Mcherry carried the EF-P acceptor loop (Ω) of *P. putida* together with different portions of the surrounding β -strands 3 (β 3) and 4 (β 4). **B)** Test for modification of mCherry carrying truncated versions of the EF-P acceptor loop. *O/n* cultures of *E. coli* LMG194 cells heterologously expressing mCherry – EF-P acceptor loop constructs (pBAD-HisA) together with or without *earP* (pBAD33) were subjected to SDS-PAGE (top) and subsequent Western blot (bottom) analysis using 0.25 μ g/ml α Arg^{Rha}.

Having identified a minimal acceptor motif that is sufficient for arginine rhamnosylation by EarP we wanted to know, whether this motif could also be recognized and modified outside of a protein scaffold and terminal of an amino acid chain. We therefore constructed mCherry variants that carried the target motif together with different portions of the surrounding β -strands (Figure 3A). Interestingly, the constructs allowed for even more drastic shortening of the acceptor motif, as even addition of the *P. putida* acceptor loop was sufficient to achieve EarP-dependent rhamnosylation of arginine (Figure 3B). To determine the minimal requirements for target modification at the terminal position, we continued to shorten the acceptor loop by one amino acid from each side. To our surprise, all of these shortened variants – including the one with only a terminal arginine - were still modified by EarP, albeit with an apparently lower efficiency than constructs with longer target motifs (Figure 3B, right). While these results indicate a remarkable and quite unexpected acceptor substrate promiscuity of the rhamnosyltransferase EarP, it has to be noted that both the unnatural target protein and the modification enzyme were overproduced in this context. Despite this, our findings are still a clear hint that – contrary to current belief – EarP might be able to modify more than just a single target protein.

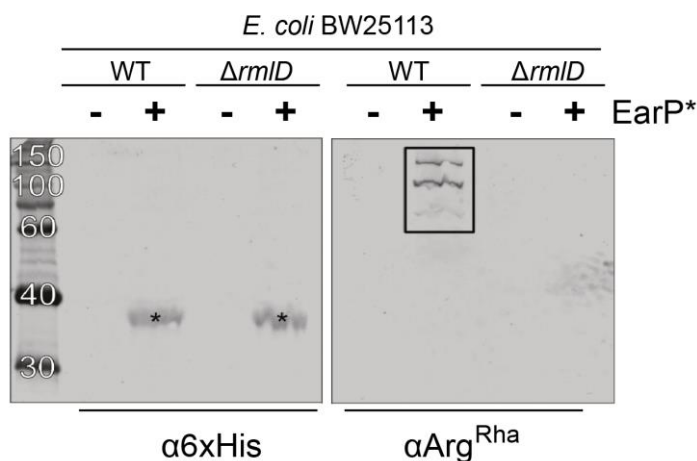


Figure 4: **EarP overexpression in *E. coli* leads to rhamnosylation of several high-molecular proteins.** *In vivo* rhamnosylation in *E. coli* BW25113 WT and $\Delta rmID$ heterologously expressing EarP (*) was detected using anti-Arg^{Rha} (right). Expression of EarP was verified using 0.1 μ g/ml α 6x His (left).

The effector arginine glycosyltransferase NleB from enteropathogenic *E. coli* (EPEC) has targets within the bacterial cell (22). Given the potentially larger than expected acceptor substrate spectrum, we were therefore interested, if EarP also engages in intracellular protein modification. Thus, we overexpressed the rhamnosyltransferase in *E. coli* BW25113 and JW2025 cells to test for both EarP and TDP-Rha dependent arginine modification. Immunoblotting using rhamnosylarginine specific antibodies showed that overexpression of EarP indeed leads to modification of several high-molecular proteins (Figure 4). The absence of a rhamnosylation signal in *E. coli* JW2025 cells overexpressing EarP confirmed that these modifications are dependent on the presence of TDP-Rha. Taken together, these results clearly show that the acceptor substrate spectrum of EarP is not limited to the cognate target protein EF-P. Accordingly, bacteria naturally expressing the rhamnosyltransferase might harbour additional rhamnoproducts that so far evaded detection by any of the common methods. This again highlights the urgent need for novel tools that enable efficient enrichment and detection of glycoproteins such as our rhamnoproduct specific antibodies (23).

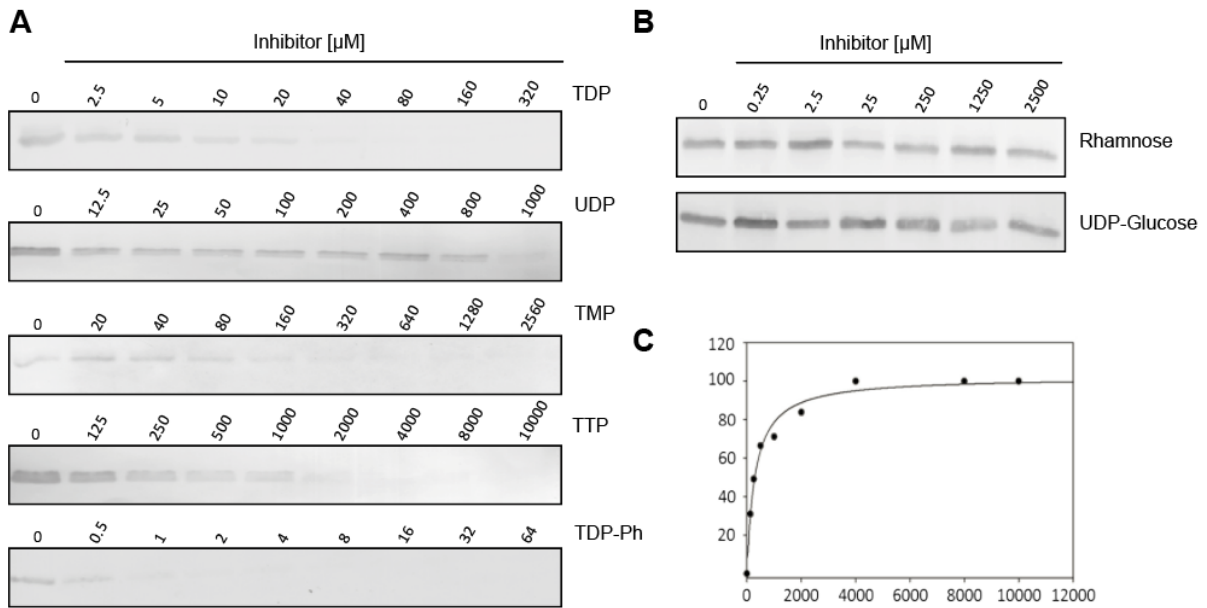
EarP is promiscuous in the nucleotide moiety of its donor substrate

To better understand the substrate specificity of a GT, it is also interesting to investigate substrate analogues. As Leloir-type glycosyltransferase, EarP uses activated nucleotide sugars as donor substrates. The binding of the donor substrates TDP-Rha occurs in the protein C-domain and is mediated by several conserved amino acids (14-16). Both, the sugar and the nucleotide, are potentially involved in binding to EarP. Hence, we first tested related nucleotides and sugars separately in a competitive inhibition assay (Fig. 5). In this *in vitro* assay, unmodified EF-P and EarP were incubated with TDP-Rha and different potential inhibitors of the rhamnosylation reaction (Fig. 5A).

First, nucleotides comprising pyrimidine bases (TDP, UDP, CDP), as well as thymine and thymidine were examined for their inhibitory potential. The rhamnosylation reaction was prevented by TDP and UDP with a calculated K_i of 1.6 μM and 21 μM , respectively (Fig. 5C, table 1). These K_i values are in the same range as the K_M of EarP with the natural substrate TDP-Rha (5 μM). The third pyrimidine derived nucleotide CDP was not able to inhibit the reaction. Thymine and thymidine did not show any inhibitory effect, indicating the importance of the phosphate group. To reveal the influence of the phosphate backbone in particular, nucleotide mono- and triphosphates of promising nucleotides were further examined (TMP, TTP and UTP). The thymidine nucleotides show minor inhibiting potential at high concentrations revealing the K_i of 118 μM for TMP and 160 μM for TTP whereas UTP did not inhibit the reaction. The α -phosphate group seem to a prerequisite for binding, whereas a triphosphate group strongly weakens the binding maybe due to steric hindrance.

Second, we tested the inhibitory potential of the sugar moiety (Fig. 5B). Interestingly, rhamnose was not able to compete with the natural substrate. This indicates that the binding to EarP is rather regulated by the nucleotide- than the sugar-moiety of the substrate confirming structural studies (14). We further investigated the nucleotide sugar UDP-glucose, which did not show any inhibitory effect. Whereas the sole UDP nucleotide exhibited binding potential, the sugar moiety seems to be rather a steric prerequisite than part of binding to EarP.

Lastly, the substrate analogue TDP-Phenol was tested due to its inhibitory potential which was shown for the TDP-Rha biosynthesis protein dTDP-6-deoxy-d-xylo-4-hexulose 3,5 epimerase (RmlC) (24). Here, we determined the lowest K_i value (0,32 μM). This shows not only the promiscuity towards the sugar moiety but also that EarP can be utilized as potential target for antimicrobial strategies. Inhibition of the glycosyltransferase could lead to loss of pathogenicity as this modification was shown to be important for pathogenicity of human pathogens (12, 25, 26).



Conclusion

Our findings challenge the current model for EF-P recognition and binding by the modification enzyme EarP. On the one hand, we were able to show that the natural donor substrate specificity of EarP is promiscuous towards UDP and variation in the phosphate backbone. On the other hand, we demonstrated that the acceptor substrate spectrum of EarP is not as narrow as was believed and goes beyond rhamnosylation of elongation factor P. This is unexpected for a dedicated glycosyltransferase and offers great potential both for gathering insights into cytoplasmic bacterial N-glycosylation and synthetic rhamnosylation of non-natural target proteins. Our finding that EarP modifies several peptides upon overproduction allows us to identify the corresponding proteins by mass spectrometry analysis. Subsequent search for homologous proteins in bacteria naturally encoding the rhamnosyltransferase such as *Neisseria* and *Pseudomonas* species will allow the identification of further intracellularly glycosylated proteins. Furthermore, the analysis of the corresponding target sites of these proteins will provide additional information on the exact mode of EarP target site recognition and modification. This in turn will allow the rational design of more sophisticated unnatural glycoproteins. Apart from this, analysis of naturally occurring motifs that are already similar to the ones targeted by EarP will allow the redesign of those proteins into glycoproteins. This synthetic addition of sugar molecules will provide additional insight into the function of such post-translational modifications and also offer functional opportunities. For example, the acceptor asparagine in the FC-region of IGG antibodies is located at the tip of a strand-loop-strand structure that is highly similar to the EF-P acceptor region. Amino acid substitutions within the FC-region might therefore allow EarP-mediated rhamnosylation and ultimately the synthesis of precursor glycans that could be used for downstream processing into functional humanized sugar modifications. This prospect is especially tempting in combination with our data on the EarP donor substrate promiscuity. As the rhamnosyltransferase is already capable of accepting UDP-activated sugars, synthetic modification into a *N*-acetylglucosamin-transferase through rational design of the active center appears to be feasible. In conclusion, our data on the substrate promiscuity of the rhamnosyltransferase EarP not only revealed hints for a more widespread distribution of this specific post-translational modification but also provides a framework for the rational design of novel glycoproteins.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft research grant LA 3658/1-1 and Research Training Group GRK2062/1 (Molecular Principles of Synthetic Biology).

Contributions

Ralph Krafczyk and Franziska Koller performed protein purification, in vitro rhamnosylation assay, and western blot analysis. Alina Sieber assisted in acceptor substrate specificity studies. Laura Lindenthal, Bernd Simon, Janosch Hennig, Jannis Brehm, Daniel Gast, and Hoffmann-Röder were involved in donor substrate specificity studies. Amit Kumar Jha and Jürgen Rohr synthesized TDP-Phenol. Kirsten Jung contributed to study design. Ralph Krafczyk and Jürgen Lassak designed the study. The manuscript was written by Ralph Krafczyk, Franziska Koller, and Jürgen Lassak.

Supplement

Table 1:

Substrate	calculated K _i (μM)
Thymidine	N./A.
Thymine	N./A.
TMP	118
TDP	1.6
UDP	21
CDP	N./A.
TTP	160
UTP	N./A.
TDP-Phenol	0.32
Rhamnose	N./A.
UDP-Glucose	N./A.
GDP-Fucose	N./A.

Methods

Growth Conditions

All strains and plasmids used in this study are listed and described in table 2 and 3. *P. putida* and *E. coli* were routinely grown in lysogeny broth (LB) (28, 29) at 30°C (for *P. putida*) and 37°C (for *E. coli*) aerobically under agitation, if not indicated otherwise. When required, media were solidified by using 1.5 % (wt/vol) agar. The medium was supplemented with antibiotics at the following concentrations when indicated: 100 mg/ml ampicillin sodium salt, 50 mg/ml kanamycin sulfate or 30 mg/ml chloramphenicol. Plasmids carrying the P_{BAD} promoter (30) were induced with L-arabinose at a final concentration of 0.2% (w/v). Plasmids comprising the lac operator sequences were induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) (Sigma-Aldrich).

Table 2:

Strain	Feature/ Genotype	Reference
<i>E. coli</i> DH5αpir	F- 80lacZΦM15 (<i>lacZYA-argF</i>)U196 <i>recA1 hsdR17 deoR thi-1 supE44 gyrA96 relA1/pir</i>	(31)
<i>E. coli</i> BW25113	F- λ - Δ(<i>araD-araB</i>)567 Δ <i>lacZ4787</i> (::rrnB-3) <i>rph-1</i> Δ(<i>rhaDrhaB</i>) 568 <i>hsdR514</i>	(32)
<i>E. coli</i> LMG194	F- Δ <i>lacX74 galE galK thi rpsL ΔphoA</i> (Pvull) Δ <i>ara714 leu::Tn10</i>	(30)
<i>E. coli</i> MG1655	wild-type; F- lambda- <i>ilvG rfb50 rph-1</i>	(33)
<i>E. coli</i> BW25113 Δ <i>rmID</i>	MG1655 P <i>cadBA::lacZ</i> Δ(<i>cadBA</i>) Δ <i>efp</i> Δ <i>rmID</i>	(34)
<i>E. coli</i> P <i>cadBA::lacZ</i> Δ <i>efp</i>	MG1655 P <i>cadBA::lacZ</i> Δ(<i>cadBA</i>) Δ <i>efp</i>	(34)

Table 3:

Plasmid	Feature/ Genotype	Reference
pBAD24	AmpR-cassette, pBBR322 origin, <i>araC</i> coding sequence, <i>ara</i> operator	(30)
pBAD33	CamR-cassette, p15A origin, <i>araC</i> coding sequence, <i>ara</i> operator	(30)
pBAD-HisA	AmpR-cassette, pBR322-derived expression vector, promoter PBAD of the arabinose operon <i>araBAD</i> from <i>E. coli</i> and its regulatory gene <i>araC</i>	Invitrogen
pBAD33 <i>earP_{Ppu}</i>	C-terminal His6-Tag <i>earP</i> version from <i>P. putida</i> KT2440	(14)
pBAD24 <i>efp_{Eco}</i>	C-terminal His6-tagged <i>E. coli efp</i>	(35)
pBAD24 <i>efp_{Ppu}</i>	C-terminal His6-tagged <i>P. putida efp</i>	(35)
pBAD-HisA <i>mcherry</i>	N-terminal His6-tagged <i>mcherry</i>	This study
Arginine walking <i>efp_{Ppu}</i>		
pBAD24 <i>efp_{Ppu}_29R</i>	pBAD24 <i>efp_{Ppu}</i> , loop substitution variant K29R + R32A	this study
pBAD24 <i>efp_{Ppu}_30R</i>	pBAD24 <i>efp_{Ppu}</i> , loop substitution variant S30R + R32A	this study
pBAD24 <i>efp_{Ppu}_31R</i>	pBAD24 <i>efp_{Ppu}</i> , loop substitution variant G31R + R32A	this study
pBAD24 <i>efp_{Ppu}_33R</i>	pBAD24 <i>efp_{Ppu}</i> , loop substitution variant N33R + R32A	this study
pBAD24 <i>efp_{Ppu}_34R</i>	pBAD24 <i>efp_{Ppu}</i> , loop substitution variant A34R + R32A	this study
pBAD24 <i>efp_{Ppu}_35R</i>	pBAD24 <i>efp_{Ppu}</i> , loop substitution variant A35R + R32A	this study
Sequential truncation <i>efp_{Eco}/efp_{Ppu}loop</i>		
pBAD24 <i>efp_{Eco}_efp_{Ppu}β13/10</i>	pBAD24 <i>efp_{Eco}</i> , substitution of domain III of EF-P _{Eco} by the <i>P. putida</i> EF-P acceptor loop with the surrounding β-strands (<i>efp_{Ppu}</i> AA16-45, β3 13/β4 10)	this study
pBAD24 <i>efp_{Eco}_efp_{Ppu}β10/10</i>	pBAD24 <i>efp_{Eco}</i> , substitution of domain III of EF-P _{Eco} by the <i>P. putida</i> EF-P acceptor loop with the surrounding β-strands (<i>efp_{Ppu}</i> AA19-45, β3 10/β4 10)	this study
pBAD24 <i>efp_{Eco}_efp_{Ppu}β8/8</i>	pBAD24 <i>efp_{Eco}</i> , substitution of domain III of EF-P _{Eco} by the <i>P. putida</i> EF-P acceptor loop with the surrounding β-strands (<i>efp_{Ppu}</i> AA21-43, β3 8/β4 8)	this study
pBAD24 <i>efp_{Eco}_efp_{Ppu}β6/6</i>	pBAD24 <i>efp_{Eco}</i> , substitution of domain III of EF-P _{Eco} by the <i>P. putida</i> EF-P acceptor loop with the surrounding β-strands (<i>efp_{Ppu}</i> AA23-41, β3 6/β4 6)	this study
pBAD24 <i>efp_{Eco}_efp_{Ppu}β4/4</i>	pBAD24 <i>efp_{Eco}</i> , substitution of domain III of EF-P _{Eco} by the <i>P. putida</i> EF-P acceptor loop with the surrounding β-strands (<i>efp_{Ppu}</i> AA25-39, β3 4/β4 4)	this study

pBAD24 <i>efp_{Eco}_efp_{Ppu}</i> β2/2	pBAD24 <i>efp_{Eco}</i> , substitution of domain III of EF-P _{Eco} by the <i>P. putida</i> EF-P acceptor loop with the surrounding β-strands (<i>efp_{Ppu}</i> AA27-37, β3 2/β4 2)	this study
pBAD24 <i>efp_{Eco}_efp_{Ppu}</i> β0/0	pBAD24 <i>efp_{Eco}</i> , substitution of domain III of EF-P _{Eco} by the <i>P. putida</i> EF-P acceptor loop with the surrounding β-strands (<i>efp_{Ppu}</i> AA29-35, β3 0/β4 0)	this study
Mcherry constructs loop <i>efp_{Ppu}</i>		
pBAD-HisA <i>mcherry_efp_{Ppu}</i> β6/6	pBAD-HisA <i>mcherry</i> , C-terminal fusion of <i>P. putida</i> EF-P acceptor loop (<i>efp_{Ppu}</i> AA23-41, β3 6/β4 6)	this study
pBAD-HisA <i>mcherry_efp_{Ppu}</i> β2/2	pBAD-HisA <i>mcherry</i> , C-terminal fusion of <i>P. putida</i> EF-P acceptor loop (<i>efp_{Ppu}</i> AA27-37, β3 2/β4 2)	this study
pBAD-HisA <i>mcherry_efp_{Ppu}</i> β0/0	pBAD-HisA <i>mcherry</i> , C-terminal fusion of <i>P. putida</i> EF-P acceptor loop (<i>efp_{Ppu}</i> AA29-35, β3 0/β4 0)	this study
pBAD-HisA <i>mcherry_efp_{Ppu}</i> β0/0 Ω-1/-1	pBAD-HisA <i>mcherry</i> , C-terminal fusion of <i>P. putida</i> EF-P acceptor loop (<i>efp_{Ppu}</i> AA30-34, β3 0/β4 0 and Ω-1/-1)	this study
pBAD-HisA <i>mcherry_efp_{Ppu}</i> β0/0 Ω-2/-2	pBAD-HisA <i>mcherry</i> , C-terminal fusion of <i>P. putida</i> EF-P acceptor loop (<i>efp_{Ppu}</i> AA31-33, β3 0/β4 0 and Ω-2/-2)	this study
pBAD-HisA <i>mcherry_efp_{Ppu}</i> β0/0 Ω-3/-3	pBAD-HisA <i>mcherry</i> , C-terminal fusion of <i>P. putida</i> EF-P acceptor loop (<i>efp_{Ppu}</i> AA32, β3 0/β4 0 and Ω-3/-3)	this study

Molecular biology methods

Enzymes and kits were used according to the manufacturers' directions. Plasmid DNA was isolated using a Hi Yield plasmid minikit (Süd-Laborbedarf GmbH). DNA fragments were purified from agarose gels by employing a Hi Yield PCR cleanup and gel extraction kit (Süd-Laborbedarf). All restriction enzymes, DNA modifying enzymes and the Q5® high fidelity DNA polymerase for PCR amplification were purchased from New England BioLabs.

All constructs were analyzed by Sanger sequencing (LMU Sequencing Service). Standard methods were performed according to the instructions of Sambrook and Russel (36).

β-Galactosidase activity assay

Cells expressing lacZ under the control of the cadBA promoter were grown in buffered LB (pH 5.8) overnight (o/n) and harvested by centrifugation. β-Galactosidase activities were determined as described in reference (37) in biological triplicates and are given in Miller units (MU) (38). Standard deviations from three independent experiments were determined.

SDS-PAGE and Western Blotting

Electrophoretic separation of proteins was carried out using SDS-PAGE as described by Laemmli (39). Separated proteins were visualized in gel using 0.5% (vol/vol) 2-2-2-trichloroethanol (40) and detected within a Gel Doc™ EZ gel documentation system (Bio-Rad). The proteins were transferred onto a nitrocellulose membrane by vertical Western blotting. Antigens were detected using 0.1 g/ml anti-His₆ tag (Abcam, Inc.) or 0.25 g/ml of anti-Arg^{Rha} (13). Primary antibodies (rabbit) were targeted by 0.2 g/ml alkaline phosphatase-conjugated anti-rabbit IgG (H&L) (goat) antibody (Rockland). Target proteins were visualized by addition of substrate solution (50 mM sodium carbonate buffer, pH 9.5, 0.01 % [wt/vol] nitroblue tetrazolium, 0.045 % [wt/vol] 5-bromo-4-chloro-3-indolylphosphate (BCIP)).

Protein purification

Protein overproduction of His₆-tagged EF-P and EarP was performed in *E. coli* LMG194 cells, grown in LB Miller at 37 °C, harboring the respective pBAD plasmid (table 3). During exponential growth, 0.2 % (w/v) was added to induce protein production. After two hours, cells were harvested by centrifugation and the resulting pellet was resuspended in 100 mM NaPi (pH 7.6). Cell lysates were lysed by sonication and proteins were purified from lysate using Ni-nitrilotriacetic acid (Ni-NTA; Qiagen) according to the manufacturer's instruction. 250 mM imidazole was used for elution of His₆-tagged proteins and subsequently removed by dialysis (o/n + 4 h at 4 °C) in 100 mM NaPi (pH 7.6). The resulting proteins were used for *in vitro* rhamnosylation assays.

***In vitro* rhamnosylation**

The inhibiting effects of different substrates were measured in an *in vitro* rhamnosylation assay. Therefore, unmodified EF-P (0.05 M) and EarP (0.005 M) were incubated at 30 °C for 5 min in 100 mM NaPi (pH 7.6). To start the reaction, TDP-Rha (2.5 μM) and the respective inhibitor substrate at various concentrations were added. The reaction was stopped after 20 s of incubation at 30 °C by the addition of one volume twofold Laemmli buffer (39) and incubation at 95 °C for 5 min. Samples were subjected to SDS-PAGE and rhamnosylated EF-P was detected as described above. Band intensities were quantified using ImageJ (27). *K_i* values were determined by fitting reaction rates (in nanomoles per milligram per second) to the Michaelis-Menten equation.

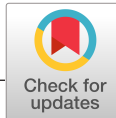
References

1. C. J. Thibodeaux, C. E. Melançon, 3rd, H. W. Liu, Natural-product sugar biosynthesis and enzymatic glycodiversification. *Angew. Chem. Int. Ed. Engl.* **47**, 9814-9859 (2008).
2. J. Härle, A. Bechthold, Chapter 12. The power of glycosyltransferases to generate bioactive natural compounds. *Methods Enzymol.* **458**, 309-333 (2009).
3. M. Dicker, R. Strasser, Using glyco-engineering to produce therapeutic proteins. *Expert Opin. Biol. Ther.* **15**, 1501-1516 (2015).
4. C. Krauth, M. Fedoryshyn, C. Schleberger, A. Luzhetskyy, A. Bechthold, Engineering a function into a glycosyltransferase. *Chem. Biol.* **16**, 28-35 (2009).
5. N. Sethuraman, T. A. Stadheim, Challenges in therapeutic glycoprotein production. *Curr. Opin. Biotechnol.* **17**, 341-346 (2006).
6. A. Naegeli, M. Aebi, Current Approaches to Engineering N-Linked Protein Glycosylation in Bacteria. *Methods Mol. Biol.* **1321**, 3-16 (2015).
7. T. G. Keys, M. Aebi, Engineering protein glycosylation in prokaryotes. *Curr. Opin. Syst. Biol.* **5**, 23-31 (2017).
8. W. Kightlinger, K. F. Warfel, M. P. DeLisa, M. C. Jewett, Synthetic Glycobiology: Parts, Systems, and Applications. *ACS Synth. Biol.* **9**, 1534-1562 (2020).
9. M. Wacker *et al.*, N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **298**, 1790-1793 (2002).
10. F. Schwarz *et al.*, A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. *Nat. Chem. Biol.* **6**, 264-266 (2010).
11. J. D. Valderrama-Rincon *et al.*, An engineered eukaryotic protein glycosylation pathway in *Escherichia coli*. *Nat. Chem. Biol.* **8**, 434-436 (2012).
12. J. Lassak *et al.*, Arginine-rhamnosylation as new strategy to activate translation elongation factor P. *Nat. Chem. Biol.* **11**, 266-270 (2015).
13. X. Li *et al.*, Resolving the alpha-glycosidic linkage of arginine-rhamnosylated translation elongation factor P triggers generation of the first Arg(Rha) specific antibody. *Chem. Sci.* **7**, 6995-7001 (2016).
14. R. Krafczyk *et al.*, Structural Basis for EarP-Mediated Arginine Glycosylation of Translation Elongation Factor EF-P. *mBio* **8**, (2017).
15. T. Sengoku *et al.*, Structural basis of protein arginine rhamnosylation by glycosyltransferase EarP. *Nat. Chem. Biol.* **14**, 368-374 (2018).
16. C. He *et al.*, Complex Structure of *Pseudomonas aeruginosa* Arginine Rhamnosyltransferase EarP with Its Acceptor Elongation Factor P. *J. Bacteriol.* **201**, (2019).
17. W. Volkwein *et al.*, Switching the Post-translational Modification of Translation Elongation Factor EF-P. *Front. Microbiol.* **10**, 1148 (2019).
18. S. Choi, J. Choe, Crystal structure of elongation factor P from *Pseudomonas aeruginosa* at 1.75 Å resolution. *Proteins* **79**, 1688-1693 (2011).
19. F. Schwarz, M. Aebi, Mechanisms and principles of N-linked protein glycosylation. *Curr. Opin. Struct. Biol.* **21**, 576-582 (2011).
20. L. Yakovlieva *et al.*, A β -hairpin epitope as novel structural requirement for protein arginine rhamnosylation. *Chem. Sci.* **12**, 1560-1567 (2021).
21. M. Charbonneau *et al.*, A structural motif is the recognition site for a new family of bacterial protein O-glycosyltransferases. *Mol. Microbiol.* **83**, 894-907 (2012).
22. S. El Qaidi *et al.*, An intra-bacterial activity for a T3SS effector. *Sci. Rep.* **10**, 1073 (2020).
23. D. Gast *et al.*, A set of rhamnosylation-specific antibodies enables detection of novel protein glycosylations in bacteria. *Org. Biomol. Chem.* **18**, 6823-6828 (2020).
24. M. F. Giraud, G. A. Leonard, R. A. Field, C. Berlind, J. H. Naismith, RmlC, the third enzyme of dTDP-L-rhamnose pathway, is a new class of epimerase. *Nat. Struct. Biol.* **7**, 398-402 (2000).
25. A. Rajkovic *et al.*, Cyclic Rhamnosylated Elongation Factor P Establishes Antibiotic Resistance in *Pseudomonas aeruginosa*. *mBio* **6**, e00823 (2015).

26. T. Yanagisawa *et al.*, *Neisseria meningitidis* Translation Elongation Factor P and Its Active-Site Arginine Residue Are Essential for Cell Viability. *PLoS one* **11**, e0147907 (2016).
27. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671-675 (2012).
28. G. Bertani, Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J. Bacteriol.* **186**, 595-600 (2004).
29. G. Bertani, Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J. Bacteriol.* **62**, 293-300 (1951).
30. L. M. Guzman, D. Belin, M. J. Carson, J. Beckwith, Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of bacteriology* **177**, 4121-4130 (1995).
31. D. R. Macinga, M. M. Parojcic, P. N. Rather, Identification and analysis of aarP, a transcriptional activator of the 2'-N-acetyltransferase in *Providencia stuartii*. *Journal of bacteriology* **177**, 3407-3413 (1995).
32. K. A. Datsenko, B. L. Wanner, One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6640-6645 (2000).
33. M. S. Guyer, R. R. Reed, J. A. Steitz, K. B. Low, Identification of a sex-factor-affinity site in *E. coli* as gamma delta. *Cold Spring Harbor symposia on quantitative biology* **45 Pt 1**, 135-140 (1981).
34. S. Ude *et al.*, Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science* **339**, 82-85 (2013).
35. W. Volkwein, C. Maier, R. Krafczyk, K. Jung, J. Lassak, A Versatile Toolbox for the Control of Protein Levels Using N(ε)-Acetyl-L-lysine Dependent Amber Suppression. *ACS synthetic biology* **6**, 1892-1902 (2017).
36. J. Sambrook, D. W. Russell, *Molecular Cloning: A Laboratory Manual, Band 2*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY., 2001), vol. 3rd edition.
37. L. Tetsch, C. Koller, I. Haneburger, K. Jung, The membrane-integrated transcriptional activator CadC of *Escherichia coli* senses lysine indirectly via the interaction with the lysine permease LysP. *Mol. Microbiol.* **67**, 570-583 (2008).
38. J. H. Miller, *A Short Course in Bacterial Genetics – A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria.*, Cold Spring Harbor Laboratory Press. (Cold Spring Harbor, 1992).
39. U. K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685 (1970).
40. C. L. Ladner, J. Yang, R. J. Turner, R. A. Edwards, Visible fluorescent detection of proteins in polyacrylamide gels without staining. *Anal. Biochem.* **326**, 13-20 (2004).

6 Transcriptional regulation of the *N*ε-fructoselysine metabolism in *Escherichia coli* by global and substrate-specific cues

Benedikt Graf von Armansperg*, [Franziska Koller](#)*, Nicola Gericke*, Michael Hellwig, Pravin Kumar Ankush Jagtap, Ralf Heermann, Janosch Hennig, Thomas Henle and Jürgen Lassak (2020). Transcriptional regulation of the *N*ε-fructoselysine metabolism in *Escherichia coli* by global and substrate-specific cues. *Mol Microbiol.* 2020; 00:1–16.



Transcriptional regulation of the N_{ϵ} -fructoselysine metabolism in *Escherichia coli* by global and substrate-specific cues

Benedikt Graf von Armansperg ¹ | Franziska Koller ¹ | Nicola Gericke ¹ | Michael Hellwig ² | Pravin Kumar Ankush Jagtap ³ | Ralf Heermann ⁴ | Janosch Hennig ³ | Thomas Henle ² | Jürgen Lassak ¹

¹Department of Biology I, Microbiology, Ludwig-Maximilians-Universität München, Munich, Germany

²Chair of Food Chemistry, Technische Universität Dresden, Dresden, Germany

³Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany

⁴Institute of Molecular Physiology, Microbiology and Wine Research, Johannes Gutenberg University Mainz, Mainz, Germany

Correspondence

Jürgen Lassak, Department of Biology I, Microbiology, Ludwig-Maximilians-Universität München, Munich, Germany. Email: juergen.lassak@lmu.de

Present address

Benedikt Graf von Armansperg, Max von Pettenkofer Institute of Hygiene and Medical Microbiology, Ludwig-Maximilians-Universität München & German Center for Infection Research (DZIF), Partner Site Munich, Munich, Germany

Nicola Gericke, Interfaculty Institute of Microbiology and Infection Medicine, Eberhard Karls Universität Tübingen, Tübingen, Germany

Funding information

J.L. gratefully acknowledges financial support by the DFG research grant LA 3658/1-1 and GRK2062/1. J.H. acknowledges support from the European Molecular Biology Laboratory (EMBL). PKAJ acknowledges EMBL and the EU Marie Curie Actions Cofund grant for an EIPOD fellowship.

Abstract

Thermally processed food is an important part of the human diet. Heat-treatment, however, promotes the formation of so-called Amadori rearrangement products, such as fructoselysine. The gut microbiota including *Escherichia coli* can utilize these compounds as a nutrient source. While the degradation route for fructoselysine is well described, regulation of the corresponding pathway genes *frlABCD* remained poorly understood. Here, we used bioinformatics combined with molecular and biochemical analyses and show that fructoselysine metabolism in *E. coli* is tightly controlled at the transcriptional level. The global regulator CRP (CAP) as well as the alternative sigma factor σ_{32} (RpoH) contribute to promoter activation at high cAMP-levels and inside warm-blooded hosts, respectively. In addition, we identified and characterized a transcriptional regulator FrIR, encoded adjacent to *frlABCD*, as fructoselysine-6-phosphate specific repressor. Our study provides profound evidence that the interplay of global and substrate-specific regulation is a perfect adaptation strategy to efficiently utilize unusual substrates within the human gut environment.

KEYWORDS

1-(ϵ -N-lysyl)-1-deoxy-D-fructose, Amadori rearrangement product, fructosyllysine, glycation, GntR transcriptional regulator

Benedikt Graf von Armansperg, Franziska Koller, and Nicola Gericke contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors. Molecular Microbiology published by John Wiley & Sons Ltd

1 | INTRODUCTION

Glycation is a non-enzymatic form of glycosylation and means a spontaneous reaction of amino compounds with reducing sugars such as glucose (Lassak et al., 2019; Ulrich and Cerami, 2001). The phenomenon was first described by Louis-Camille Maillard in 1912 being predominantly responsible for the taste, aroma, and appearance of thermally processed food (Maillard, 1912a; 1912b). Simple condensation products of primary amino groups at the N-termini of polypeptides or the ϵ -amino group of lysine and reducing sugars such as glucose are the most prevalent Maillard reaction products in food (Henle, 2003). These “sugar-amino acids” are also called “Amadori rearrangement products” (ARPs) (Amadori, 1925; Hodge, 1955). Protein-bound Maillard reaction products can be the result of a condensation reaction, taking place between an aldose or ketose and a primary amine either in form of an α -amino group at the N-terminus or an ϵ -amino group of lysine residues within the polypeptide chain. Bacteria, including *Escherichia coli*, *Bacillus subtilis*, and *Salmonella enterica* have evolved efficient strategies to use these ARPs as sole carbon source (Ali et al., 2014; Miller et al., 2015; Wiame et al., 2002; 2004; 2005). Notably, the wide distribution of the ARP catabolism among the gut microbiota (Sabag-Daigle et al., 2018) further suggests that this carbon source plays an important role in colonizing the intestinal environment (Barroso-Batista et al., 2020).

While there are numerous uptake mechanisms and diverse specificities toward glycation products in distinct microorganisms (Miller et al., 2015; Sabag-Daigle et al., 2018; Wiame et al., 2004), degradation follows a conserved route and can be illustrated by the *E. coli* N_{ϵ} -fructoselysine (ϵ -FrK) metabolism. Upon uptake—presumably by the putative permease FrIA—a kinase FrID phosphorylates the sugar moiety at the C6-position (Figure 1) (Wiame et al., 2002). In a second step, the deglycase FrIB hydrolyses fructoselysine-6-phosphate (FrK-6P) into glucose-6-phosphate and lysine to be further processed via glycolysis and amino acid metabolism, respectively. Together with an additional N_{ϵ} -psicoselysine/ ϵ -FrK epimerase FrIC (Wiame and Van Schaftingen, 2004), the pathway is encoded in one single operon *frIABCD* of thus far unknown regulation. In the present study we show that the *E. coli* ϵ -FrK catabolism is tightly controlled by positive and negative regulation. On the one hand, the global transcription factor CRP (CAP) as well as the sigma factor σ^{32} (RpoH) contribute to promoter activation. On the other hand, we identified the previously elusive regulator FrIR_{Eco}, encoded adjacent to *frIABCD*, as an ϵ -FrK specific roadblock repressor. However, ϵ -FrK is not recognized directly but only upon phosphorylation. FrIR_{Eco} itself is a member of the GntR/HutC family of transcriptional regulators recognizing the consensus sequence 5'-(N)_yGT(N)_xAC(N)_y-3'. Binding of ϵ -FrK-phosphate presumably leads to structural rearrangements in the FrIR_{Eco} transcriptionally active dimer, which in turn weakens DNA binding, and subsequently permits transcription of the *frIABCD* operon. Thus, we conclude that the interplay of global and substrate-specific regulation combined with a σ^{32} mediated transcription activation as response of colonization of a warm-blooded host is a perfect adaptation to utilize thermally processed food within the human gut environment.

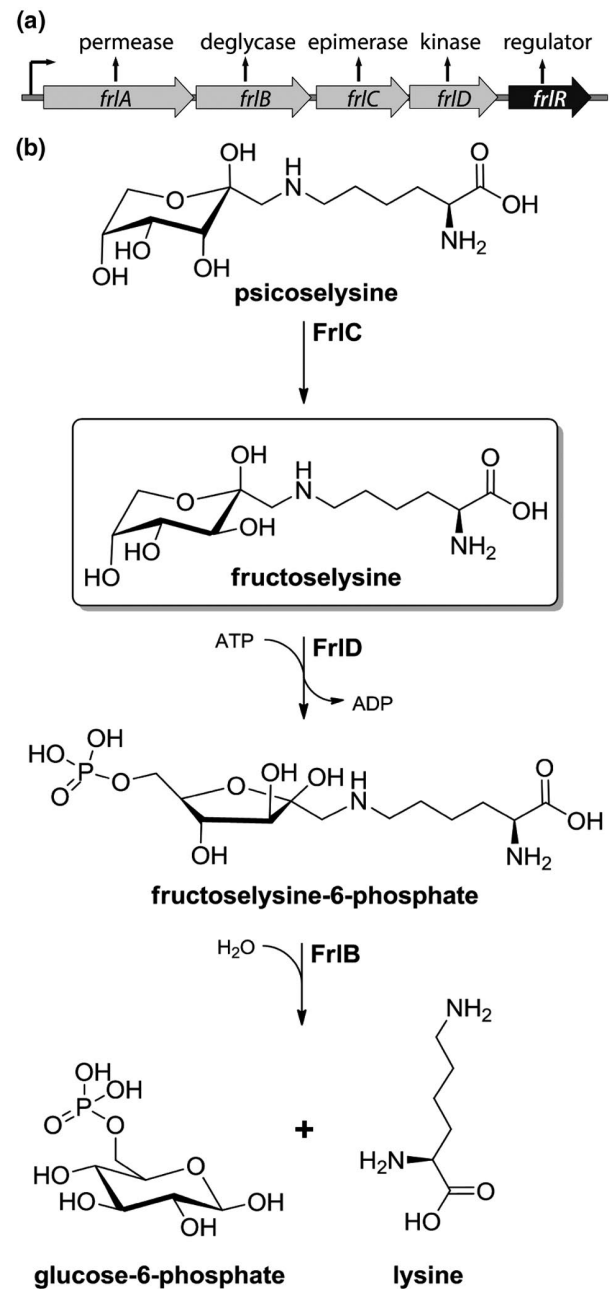


FIGURE 1 N_{ϵ} -fructoselysine metabolism in *Escherichia coli*. (a) Organization of the *frIABCDR* genome region. (b) Degradation of the ARP N_{ϵ} -fructoselysine is a two-step reaction: First, a kinase FrID phosphorylates the amino sugar, thereby forming fructoselysine-6-phosphate. In the second step, the deglycase FrIB hydrolytically cleaves sugar phosphate and amino acids into its two building blocks lysine and glucose-6-phosphate, the latter of which is directly shuffled into glycolysis. A further enzyme in the pathway, termed FrIC, is capable in catalyzing the epimerization of psicoselysine into fructoselysine

2 | RESULTS

2.1 | FrIR_{Eco} is a putative GntR like transcriptional regulator

Wiame and coworkers noticed the presence of a putative regulator, termed FrIR_{Eco}, in the genomic vicinity of the ϵ -FrK degradation

pathway (Wiame et al., 2002). However, its role in controlling the *frlABCD* operon remained enigmatic. To elucidate the function of FrIR_{Eco}, we started with a bioinformatic comparison and performed a multiple sequence alignment (Figure 2). This revealed sequence similarities to the *B. subtilis* orthologous regulator FrIR of the *frl-BONMD* operon (Deppe et al., 2011), encoding the genes to metabolize various α -glycated amino acids (Wiame et al., 2004). We also identified a distinct ortholog in *S. enterica*, termed FraR, being encoded immediately upstream of the *fraBDAE* operon, a gene cluster that is needed for degradation of fructoseasparagine (Ali et al., 2014). Taken together, *E. coli* FrIR is likely to be involved in substrate-specific regulation of ARP metabolism. The outcome of our blast search also suggests a common regulatory theme that applies to all ARP metabolizing organisms despite their distinct substrate spectra.

We next used Phyre2 (Kelley et al., 2015) as well as the iTASSER suite (Yang et al., 2015) and performed a homology modelling to gain first molecular insights into a putative mode of action (Figure 3).

This approach revealed the cytoplasmic regulator NagR (YvoA) of *B. subtilis* as structural homolog of FrIR_{Eco} (identity: 29%/ similarity: 51%) (Fillenberg et al., 2015; 2016; Resch et al., 2010). NagR_{Bsu} is a GntR type transcription factor, negatively regulating the genes from the *N*-acetylglucosamine-degrading pathway (Resch et al., 2010). Its C-terminal domain binds the effectors *N*-acetylglucosamine-6-phosphate (GlcNAcP) or glucosamine-6-phosphate (GlcNP) (Fillenberg et al., 2015) and adopts a chorismate lyase fold, thus belonging to the UbiC transcription regulator-associated (UTRA) protein family (PF07702). The NagR_{Bsu} N-terminal part, moreover, is a winged helix-turn-helix (wHTH) DNA-binding domain. It is proposed, that in an allosteric coupling mechanism GlcNAcP binding to the UTRA domain promotes a loop-to-helix transition. This ultimately leads to a 122° rotation of the wHTH-domains in a so-called jumping-jack-like motion (Resch et al., 2010). As a result, DNA-binding is weakened and repression is abolished. Based on their structural similarities we thus speculate that NagR_{Bsu} and FrIR_{Eco} might share a common molecular mechanism.

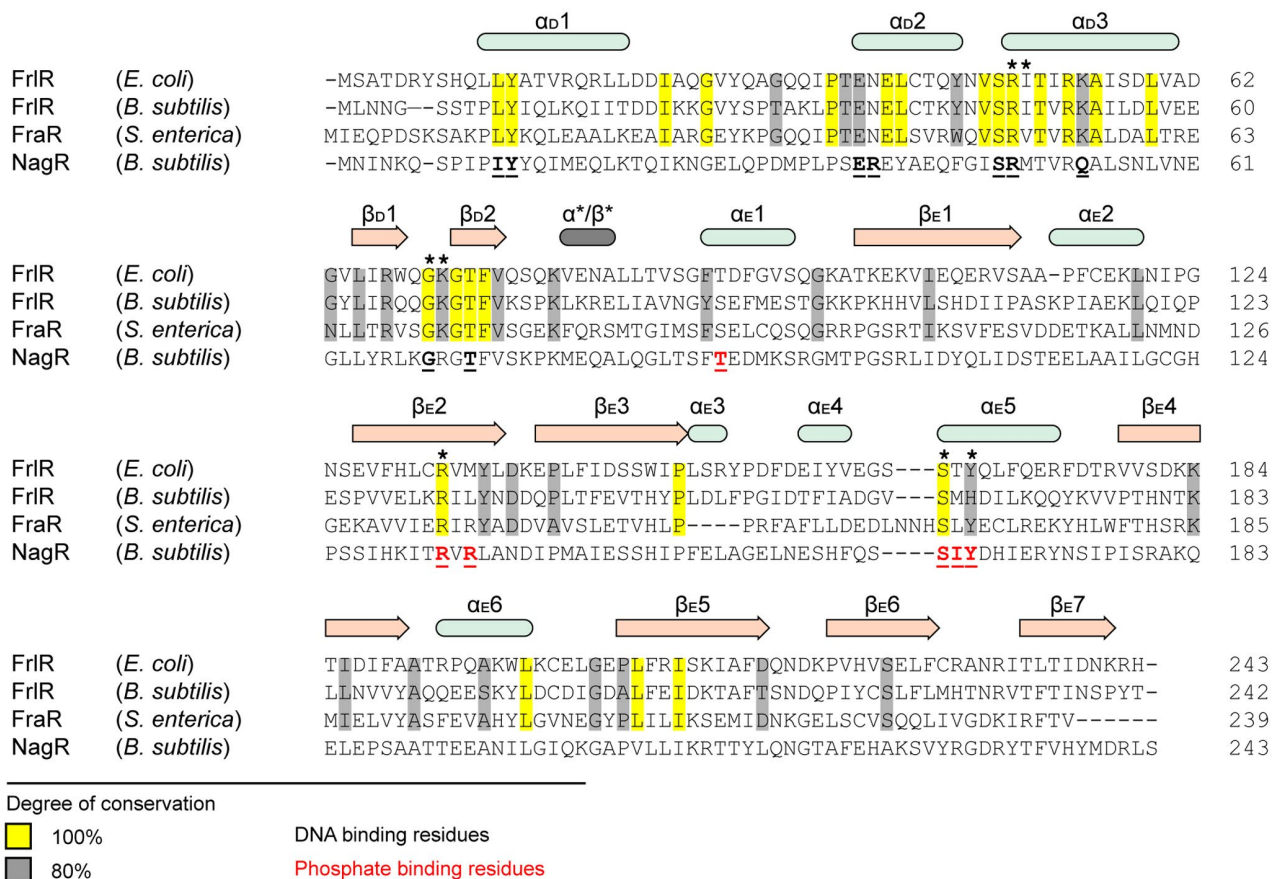


FIGURE 2 Evolutionary conservation of amino acids in FrIR homologs. Multiple sequence alignment of FrIR/FraR/NagR proteins from *Escherichia coli*, *Bacillus subtilis* and *Salmonella enterica*. The multiple sequence alignment was generated using Clustal Omega (Sievers et al., 2011). Secondary-structure elements are shown and based on the NagR_{Bsu} crystal structure (4UOW): α -Helices and β -strands are colored in mint, salmon, and grey, respectively. N- and C-domain elements are numbered and marked with a subscript “D” (for DNA binding domain) and “E” (for effector-binding and oligomerization domain), respectively. Conserved residues are colored according to their degree of conservation with yellow (100%) and grey ($\geq 80\%$). Degree of conservation is based on an alignment of 20 protein sequences that were collected from the NCBI database (Table S4). Amino acids important for function of NagR_{Bsu} are underlined either in black—being relevant for DNA-binding—or red—being relevant for coordination of the phosphate moiety of *N*-acetylglucosamine-6-phosphate. “**” depict amino acids whose substitution leads to an impaired FrIR_{Eco} functionality

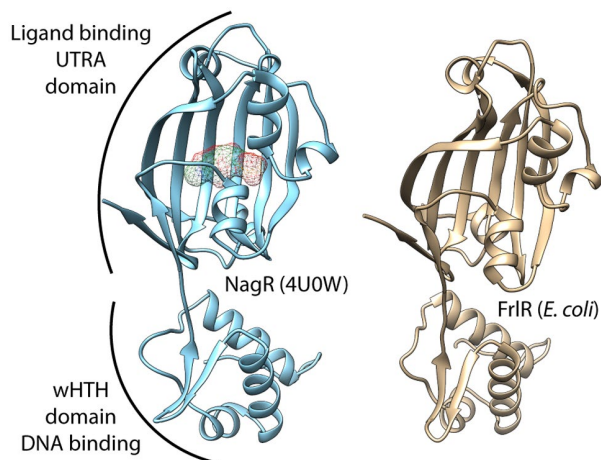


FIGURE 3 Structural model of FrIR_{Eco} and comparison with NagR of *Bacillus subtilis*. Left: Ribbon representation of the crystal structure of NagR_{Bsu} (aa9–aa243) of *B. subtilis* in complex with *N*-acetylglucosamine-6-phosphate (mesh) (4U0W). Right: Ribbon representation of the i-TASSER homology model of FrIR of *Escherichia coli* (aa9–aa243). Illustrations were generated with UCSF Chimera

2.2 | FrIR_{Eco} delays growth on ϵ -fructoselysine

To see whether FrIR_{Eco} functions as a repressor similarly to NagR_{Bsu}, we first performed a series of growth experiments with FrK. Initially, we reinvestigated the capability of *E. coli* K12 to utilize this ARP as sole carbon source. To this end, we monitored bacterial growth in M9 minimal medium (Miller, 1972) supplemented with either 1 mM ϵ -FrK or 1 mM glucose, the latter serving as positive control (Figure 4a left). As demonstrated earlier (Griffiths and Pridham,

1980; Wiame et al., 2002), *E. coli* can grow on both carbon sources yielding similar biomass as can be concluded from their same maximal optical density at 600 nm. The major difference between the two curves is a prolonged lag phase and an increased doubling time of 210 min for ϵ -FrK supplemented cells, being around 20% longer compared to glucose (180 min). As further control, we tested a strain Δ frlD lacking the FrK kinase, which catalyzes the first step in the degradation pathway. Expectedly, this mutant was no longer able to grow on ϵ -FrK (Figure 4a right). *Escherichia coli* FrK kinase FrID shows only little enzymatic activity toward α -glycated amino acids in vitro (Wiame et al., 2004) and hence we were curious whether α -FrK can substitute for the ϵ -glycated lysine in vivo. In line with the previously determined substrate specificity of FrID, we observed neither significant growth with the α -glycated ARP nor any degradation (Figure 4a left, 4b). Having confirmed these previous findings, we went on to investigate whether FrIR_{Eco} is involved in the regulation of ϵ -FrK utilization (Figure 4b). We compared growth of an *E. coli* wild type with a Δ frlR strain and found the phenotype with respect to total biomass yield indistinguishable from each other. At the same time, we noticed a shortened lag-phase, giving the first hint that FrIR_{Eco} acts as a repressor. In line with this assumption, the overproduction of FrIR_{Eco} (*frlR*⁺⁺) prevents efficient ϵ -FrK utilization (Figure 4a right, 4b, Figure S1). Presumably, the increased protein copy number decouples DNA binding from substrate recognition, and thus reveals FrIR_{Eco} to be a transcriptional repressor. In parallel, we quantified metabolization of both derivatives of FrK by the wild type during 24 hr of incubation and using direct amino acid analysis. Whereas the concentration of α -FrK remained unchanged during the experiment, ϵ -FrK was almost completely degraded while at the same time lysine was formed in an equimolar amount. The results

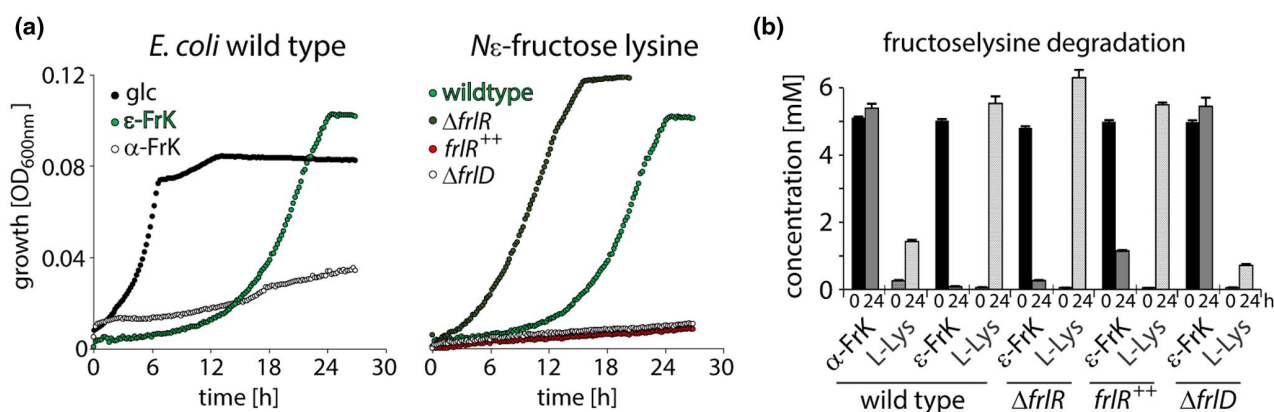


FIGURE 4 Growth analysis and fructoselysine degradation in *Escherichia coli* wild-type and *frl* mutant strains. (a) Left: Growth of *E. coli* BW25113 wild type in M9 minimal medium supplemented with either 1 mM glucose (glc) (black circles), 1 mM N_{ϵ} -fructoselysine (ϵ -FrK) (green circles), or 1 mM N_{α} -fructoselysine (α -FrK) (white circles). Right: Growth of *E. coli* BW25113 on 1 mM ϵ -FrK in comparison to isogenic mutant strains JW3337 (Δ frlD) (white circles) and JW5698 (Δ frlR) (black circles). BW25113 cells ectopically overexpressing FrIR_{Eco} (*frlR*⁺⁺) (red circles). Shown is a representative curve from three independent biological replicates. (b) FrK degradation in *E. coli* BW25113 wild-type cells in comparison to isogenic mutant strains JW3337 (Δ frlD) and JW5698 (Δ frlR) as well as BW25113 cells ectopically overexpressing FrIR_{Eco} (*frlR*⁺⁺). The educts α / ϵ -FrK as well as the product L-lysine (L-Lys) were quantified after 0 hr and 24 hr incubation time in M9 Minimal medium supplemented with 5 mM of the respective carbon source using cation-exchange chromatography with post-column ninhydrin derivatization and UV-detection of the reaction products (“amino acid analysis”). The mean values and standard deviations of two technical replicates are shown

for ϵ -FrK were similar in the $\Delta frlR$ mutant. Less ϵ -FrK was degraded by $frlR^{++}$ cells, while no degradation and only little lysine formation was observed in the $\Delta frlD$ strain lacking the kinase. We conclude that growth of *E. coli* is retarded when metabolization of ϵ -FrK is inhibited.

2.3 | *frlABCD* transcription is positively controlled by σ_{32} and CRP as well as negatively regulated by $FrIR_{Eco}$

Having shown that ϵ -FrK utilization is subject to $FrIR_{Eco}$ regulation, we rigorously analyzed the 5'-UTR (untranslated region) of *frlABCD* employing $P_{frlABCD}$ promoter fusions to uncover further control elements. It is predicted that *frlABCD* transcription starts 75 nucleotides 5' of the *frlA* open reading frame (+1) and depends on the housekeeping sigma factor σ^{70} (RpoD) (Figure 5a) (Huerta and Collado-Vides, 2003). We further noticed a sequence matching the CRP/CAP (cAMP Response Protein/Catabolite Activator Protein) binding site centered around position -41.5 and thus being in perfect distance to constitute a class II promoter (Lawson et al., 2004). To test our hypotheses on transcriptional regulation of the $P_{frlABCD}$

promoter we initially fused 294 bp (-219/+75) 5' of *frlABCD* with the lux-operon *luxCDABE* of *Photobacterium luminescens* (Volkwein et al., 2017) and measured the light output over a time course of 24 hr in *E. coli* BW25113 wild-type cells grown in Lysogeny broth (LB/Miller) (Bertani, 2004). We reached a maximal luminescence per OD₆₀₀ of about 1×10^6 RLU, showing that this region comprises a fully functional promoter (Figure 5b-d). We note that a further sequence extension to 397 bp (-322/+75) did not change this emission significantly (Figure 5b), demonstrating that the -219/+75 lux fusion encompasses all required elements for transcriptional control. A truncation to 140 bp (-65/+75) reduced the light output by a factor of five to 2×10^5 RLU showing that this construct lacks an important part of the promoter. Using RegulonDB (Santos-Zavaleta et al., 2019) we became aware of putative -10 and -35 regions of the heat shock sigma factor σ^{32} (RpoH). To investigate their involvement in $P_{frlABCD}$ activation, we ectopically expressed *rpoH* and measured the effect on the promoter lux fusions. A similar strategy was described earlier and is favored as the sigma factor is essential at growth temperatures higher than 20°C (Zhao et al., 2005). With the -219/+75 construct σ^{32} overproduction led to a significant increase in luminescence compared to the wild-type situation (Figure 5c). Interestingly, -65/+75 also responded positively to σ^{32} expression

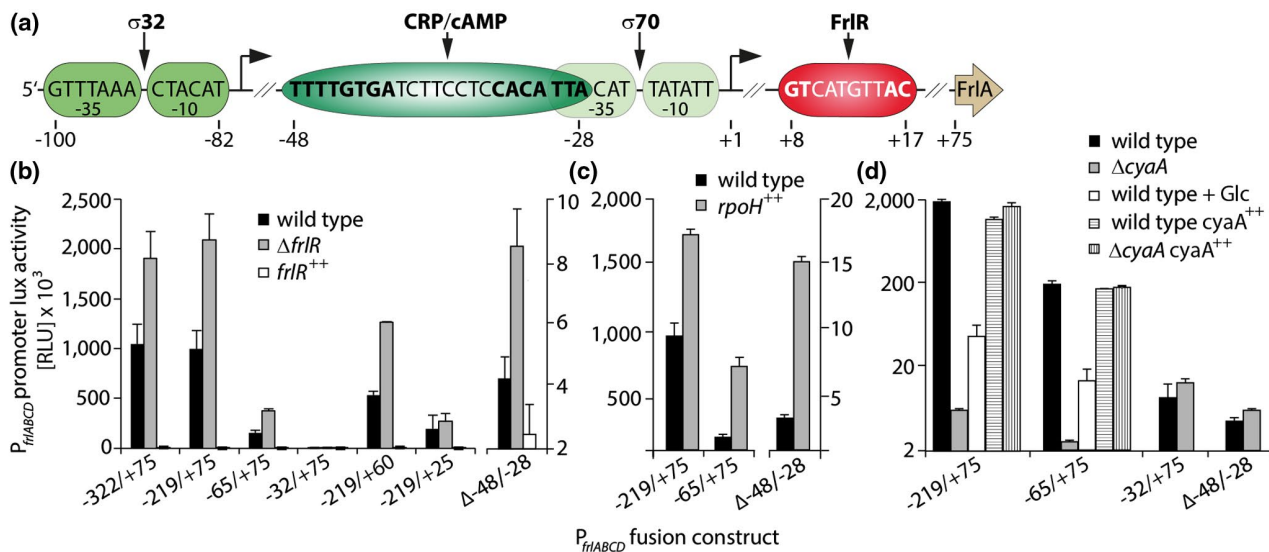


FIGURE 5 In vivo analysis of the *frlABCD* promoter region. (a) Illustration of the *frlABCD* promoter region with +1 as the putative transcriptional start site. DNA-recognition sites for RNA-polymerase are boxed in green (dependent on RpoH/ σ^{32}) and light green (dependent on RpoD/ σ^{70}). The binding motif of the cAMP-activated global transcriptional regulator CRP is boxed in dark green. The DNA-binding site of $FrIR_{Eco}$ is boxed in red. Consensus sequences for the transcription factors are highlighted with bold letters. (b-d) In vivo analyses of $P_{frlABCD}$ promoter lux fusions in *E. coli* cells containing the respective reporter plasmid. The maximal light emission from a 24 hr time course experiment of cells grown in LB (Miller) is given in RLU. The mean values and standard deviations of at least three biological replicates are shown. Naming of the $P_{frlABCD}$ -lux promoter truncations depicted on the abscissas gives information about the sequence length and numbering and relates to the illustration in (a). *E. coli* BW25113 wild-type cells (black bars) in comparison to the isogenic mutant strain JW5698 ($\Delta frlR$; grey bars) (Baba et al., 2006) and cells where $FrIR_{Eco}$ overexpression was induced by the addition of 0.2% (w/v) arabinose ($frlR^{++}$; white bars) from a pBAD33 backbone were analyzed in (b). To highlight the differences with Δ -48/-28 a second ordinate was introduced. (c) *E. coli* BW25113 wild-type cells (black bars) in comparison to cells where *rpoH* overexpression was induced by the addition of 0.2% (w/v) arabinose (*rpoH*⁺⁺; grey bars) from a pBAD33 backbone. To highlight the differences with Δ -48/-28 a second ordinate was introduced. (d) *E. coli* BW25113 wild-type cells (black bars) in comparison to an isogenic mutant strain, which lacks the adenylate cyclase CyaA, JW3778 ($\Delta cyaA$) and cells ectopically overexpressing *cyaA* (*cyaA*⁺⁺; dashed lined bars). A wild-type cell culture supplemented with 20 mM glucose (glc; white bars) was also analyzed

(two-fold increase in light output), implying that the sigma factor also binds within the presumed σ^{70} promoter region. Therefore, we included another reporter construct which lacks base pairs -28 to -48 ($\Delta-48/-28$). Cells harboring $\Delta-48/-28$ turned almost dark unless σ^{32} was overproduced which increased light production five-fold. Thus, we can conclude on the presence of two promoters both of which are needed to fully induce $P_{frIABCD}$: One in close proximity to the transcription start site and presumably the major driver of $frIABCD$ expression, the other further upstream and important to reach high level activation. While both promoter regions respond to increased levels of σ^{32} , only the downstream one encompasses the DNA-binding elements for the house keeping sigma factor σ^{70} as well as CRP.

Accordingly, we next assessed the regulation by catabolite repression. In an inverse correlation to the extracellular glucose concentration, CRP activates metabolic processes that facilitate the conversion of alternative carbon sources. Notably, the glucose content is not measured directly but is derived from the intracellular cAMP concentration which in turn relies on the activity of a sole cAMP-synthetizing adenylate cyclase CyaA (Lawson et al., 2004). Accordingly, luminescence was analyzed in the $-65/+75$ promoter fusion which should comprise the full CRP DNA-binding motif. Here, LB/Miller growth conditions, with amino acids as sole carbon source, were compared on the one hand to a culture supplemented with 10 mM glucose and on the other hand to a $\Delta cyaA$ strain (Baba et al., 2006). For the latter two conditions, the light output was strongly diminished (Figure 5d) corroborating our assumption about carbon catabolite repression. This is in line with an earlier global study that suggested CRP-dependent regulation of the $frIABCD$ operon (Shimada et al., 2011). We also note that glucose and cAMP dependency is lost with the truncated promoter fusion $-32/+75$ (Figure 5d). However, with this construct the overall luminescence is strongly decreased. As $-32/+75$ no longer encompasses the full CRP-binding site, RNA-polymerase recruitment is impaired (Figure 5a). Similarly, $\Delta-48/-28$ does no longer respond to cAMP ($\Delta cyaA$), demonstrating that there is no second CRP-binding site. Next, we investigated whether inducer exclusion (Deutscher et al., 2014; Osumi and Saier, 1982; Sondej et al., 2002) might also play a role in $P_{frIABCD}$ regulation. Hence, we analyzed luminescence development with $-219/+75$ and in a mutant lacking component EIIA/Crr (Δcrr) of the phosphotransferase system (Figure S2) (Deutscher et al., 2014). Compared to the wild type, the light output in the mutant was diminished (Figure S2a). This was expected as in Δcrr cells CyaA adenylate cyclase activity can no longer be stimulated by phosphorylated EIIA. At the same time, an external supplement with 1 mM ϵ -FrK led to elevated light production showing that regulation by $FrIR_{Eco}$ was retained. The negative regulation by glucose (Figure 5d), however, was mitigated in the mutant (Figure S2b). This in turn gives a first hint that EIIA might prevent ϵ -FrK internalization by inhibiting the assumed transport activity of FrIA, similar to those what was reported for other sugar transporters (Deutscher et al., 2014). Additional experiments are needed, to further corroborate this hypothesis.

Lastly, we employed the lux reporter to investigate transcriptional regulation by $FrIR_{Eco}$. Our initial growth experiments (Figure 4a) indicated the role of the protein as negative regulator repressing $P_{frIABCD}$ under non-inducing conditions. However, the strong light emission for wild-type cells harboring $P_{frIABCD}$ -lux fusions in LB/Miller seems to contradict this assumption. We hypothesized that the plasmid-borne nature of the reporter might imbalance the ratio of $FrIR_{Eco}$ and promoter abundance in favor of the latter. Accordingly, increasing the $FrIR_{Eco}$ copy number—by introducing an arabinose inducible ectopic copy of $frIR_{Eco}$ (pBAD33- $FrIR$ (Eco))—should be enough to silence the promoter and this was in fact what we observed (Figure 5b). We also note that cells lacking $frIR_{Eco}$ produced 1.5-fold more the light of wild-type cells, further supporting that $FrIR_{Eco}$ is a transcriptional repressor. We ultimately analyzed whether this repression is ϵ - α -FrK dependent. Consequently, bioluminescence development was now measured in $\Delta frIR$ cells in the concomitant presence of a P_{BAD} controlled, plasmid-encoded copy of $frIR_{Eco}$ (pBAD33 $FrIR$ (Eco)) and the $-219/+75$ $P_{frIABCD}$ -lux reporter construct. In this scenario, light emission was strongly reduced (even without arabinose induction) but becomes elevated again when adding 1 mM ϵ -FrK (Figure 6). In contrast, promoter repression was maintained with 1 mM α -FrK. These data confirm that specificity for the ϵ -glycated form is not limited to the metabolism, but also extends, directly or indirectly, to regulation. To define the sensitivity of the system we performed a titration series with ϵ -FrK (10 μ M-10 mM), and thus determined that the system responds to concentrations as low as 10 μ M (Figure 6, Figure S5). The strength of depression is gradual with a maximum in the low mM range. Considering that the ϵ -FrK uptake with our diet can reach such concentrations (Henle, 2003; 2005), *E. coli* is perfectly adapted to its natural human gut habitat.

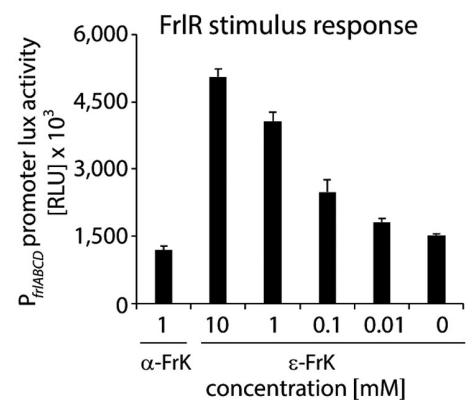


FIGURE 6 $FrIR_{Eco}$ sensitivity analysis toward ϵ -fructoselysine. In vivo analyses of the $P_{frIABCD}$ promoter lux fusion $-219/+75$ in *Escherichia coli* JW5698 cells ($\Delta frIR$) that concomitantly express $frIR_{Eco}$. The maximal light emission from a 24 hr time course experiment of cells grown in LB (Miller) (without the addition of the P_{BAD} inducing agent L-arabinose) is given in RLU. The mean values and standard deviations of at least three biological replicates are shown

2.4 | FrIR_{Eco} is a repressor that binds to TTGT CATGTT AC CTT via its winged helix-turn-helix domain

Knowing that FrIR_{Eco} is a transcriptional repressor, we went on to identify its cognate DNA-binding motif. Therefore, we analyzed various P_{frIABCD}-lux fusions on FrIR_{Eco} dependent repression (Figure 5B). These were truncated 5' (-322/+75, -219/+75, -65/+75, -33/+75 and Δ-48/-28) or 3' (-219/+60 and -219/+25) relative to the transcriptional start site (+1). To our surprise we saw a strong drop in light output with -219/+60 and -219/+25. As the ribosome binding site driving LuxC production was kept the same in all fusion constructs, the reduction might be attributed to structural alterations of the mRNA that potentially interfere with proper binding to the ribosome.

Upon ectopic frIR_{Eco} expression luminescence decreased dramatically irrespective of the promoter truncation tested. This in turn narrows down the putative binding site to a region of about 50 nucleotides. The proposed structural similarities of FrIR_{Eco} to NagR_{Bsu} (Figure 3) implied that both transcription factors belong to the same family of regulators, namely the GntR/HutC related ones (Hoskisson and Rigali, 2009; Rigali et al., 2002). Members of this group bind to a motif 5'-(N)_yGT(N)_xAC(N)_y-3' and so do NagR_{Bsu} and FrIR_{Bsu} (Deppe et al., 2011; Fillenberg et al., 2015). We recognized a putative operator 5'-(N)_yGT CATGTT AC (N)_y-3' (frIO_{Eco}) immediately downstream of the transcription start site indicating that RNA polymerase is occluded from binding to the σ⁷⁰ dependent promoter when frIO_{Eco} is occupied by FrIR_{Eco}. To prove this hypothesis, the putative FrIR_{Eco} binding motif was placed between the T7-polymerase promoter (P_{T7}) and a synthetic ribosome-binding site, all of which precede the ORF of the super folder green fluorescent protein (sfGFP) (Figure 7a). Fluorescence intensity was measured as means of the transcriptional activity in the absence and presence of a plasmid-borne copy of frIR_{Eco}. While sfGFP expression was high without FrIR_{Eco}, cells turned dark with the repressor. Thus, the artificially introduced motif comprises frIO_{Eco}.

When substituting the conserved GT/AC by CC/GG, sfGFP expression was no longer controlled by FrIR_{Eco}. Similarly, the extension of the 6 bp long spacer by one additional "A" heavily interfered with DNA-binding as can be concluded from the heterogeneous fluorescence signal. As FrIR_{Eco} is expected to work as a dimer (Fillenberg et al., 2015), a spacer length of six is indicative for a binding of each monomer at opposite sides of the operator region.

We were also curious what happens, when we mirror the core motif. Here, FrIR_{Eco} was still able to diminish sfGFP expression, however not as efficient as with the native binding site. This led us to conclude that, beside the bracketing GT/AC and spacer length, their upstream and downstream residues—which were kept in the original order—play also a crucial role in frIO_{Eco} recognition.

Our in vivo analysis was complemented by testing the operator sequence on FrIR_{Eco} binding in vitro. We employed thermal shift assays (Huynh and Partch, 2015) to assess heat stability of FrIR_{Eco} in

dependence of frIO_{Eco} (frIO⁺) (Figure S3). As controls we measured melting temperatures in the presence of a random DNA-fragment of similar size (frIO⁻) as well as in the absence of any DNA. In theory, DNA-binding to FrIR_{Eco} should stabilize the protein and in turn, the melting temperature increases. In fact, we observed a FrIR_{Eco} melting temperature of 62°C in combination with frIO_{Eco}, whereas it was 3°C lower in the mixture with the random DNA-fragment.

As these data are only qualitative, we next utilized Surface Plasmon Resonance spectroscopy (SPR) in order to determine the binding kinetics (association rate k_a and dissociation rate k_d) as well as the affinity (K_D) of FrIR_{Eco} and its cognate recognition motif (Figure 7b, Figure S3). To this end, different concentrations of FrIR_{Eco} were combined with two immobilized DNA fragments (see the experimental procedures for details). One fragment includes the FrIR_{Eco} binding site, the other was of random sequence composition and served as negative control. Before that, the absolute and "active" fraction of FrIR_{Eco} was determined using calibration free concentration analysis (CFCA) to approximately 50% of the total protein concentration. A clear and stable binding could be observed for FrIR_{Eco} to frIO_{Eco} with a high association ($k_a = 7.0 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$) and low dissociation ($k_d = 3.4 \times 10^{-2} \text{ s}^{-1}$) rate, whereas only a weak binding of FrIR was seen with the control DNA. Calculations were based on the association and dissociation rates and with this we derived an affinity of FrIR_{Eco} for frIO_{Eco} of 4.9 nM. In comparison, the GntR/HutC transcription regulator NagR_{Bsu} has a 250 times lower dissociation constant of around 20 pM for the native nagAB operator (Fillenberg et al., 2015) which might reflect an adaptation strategy to respond appropriately to specific stimuli.

We were also curious whether the *E. coli* FrIR binding motif is recognized by the orthologous regulators FrIR of *B. subtilis* and FraR of *S. enterica*. In this regard, the corresponding genes were placed under control of the arabinose inducible promoter P_{BAD} analogous to frIR of *E. coli*. NagR of *B. subtilis* (recognition motif: 5'-GTGGTCTAGACCAC-3') was also included in the study and served as negative control. Binding to P_{frIABCD} was tested once more employing the -219/+75 lux reporter. Despite significant differences in sequence composition of their cognate DNA-binding motifs (Figure 7c), the two FrIR_{Eco} homologs FrIR_{Bsu} and FraR but not NagR_{Bsu} negatively regulated P_{frIABCD} dependent luxCDABE expression under inducing conditions (Figure 7d). By contrast, leaky expression (no arabinose for promoter activation added) was only sufficient for efficient silencing by FrIR_{Eco}. While FrIR_{Bsu} retained a mild repressing phenotype, FraR completely lost its regulatory capability. Reportedly, such copy number dependent differences hint to alterations in DNA-binding affinity (Schlundt et al., 2017). The strength of repression matches the phylogenetic distance to FrIR_{Eco} and decreases the more distantly related the transcription factor is (Figure 2).

It is notable, that the sequence length between the bracketing GT/AC differs among the three FrIR homologs. Whereas, in frIO_{Eco} and frIO_{Bsu} the spacer encompasses six bp, in the putative fraO it is seven bp long. This distinguishing feature might provide one rationale explaining the differences in repression efficiency.

To assess DNA binding by FrIR_{Eco}, we first verified DNA binding by the putative wHTH-DNA binding domain (Figure 3). DNA-binding to the recognition sequence can be enforced by overproduction of the wHTH-domain solely (Schlundt et al., 2017). Accordingly, we truncated FrIR_{Eco} and compared the fragment encompassing amino acids 1-77 with the full-length protein (244 aa). In *E. coli* cells, that

simultaneously harbor the P_{frIABCD}-lux⁻219/+75 reporter plasmid, luminescence can be suppressed only when transcription of the FrIR_{Eco} 1-77 aa fragment was arabinose induced while with the full-length FrIR_{Eco} even low protein levels diminished the light output (Figure 7d). These data strongly imply that indeed the first 77 residues fold into a wHTH-domain that is sufficient to recognize the

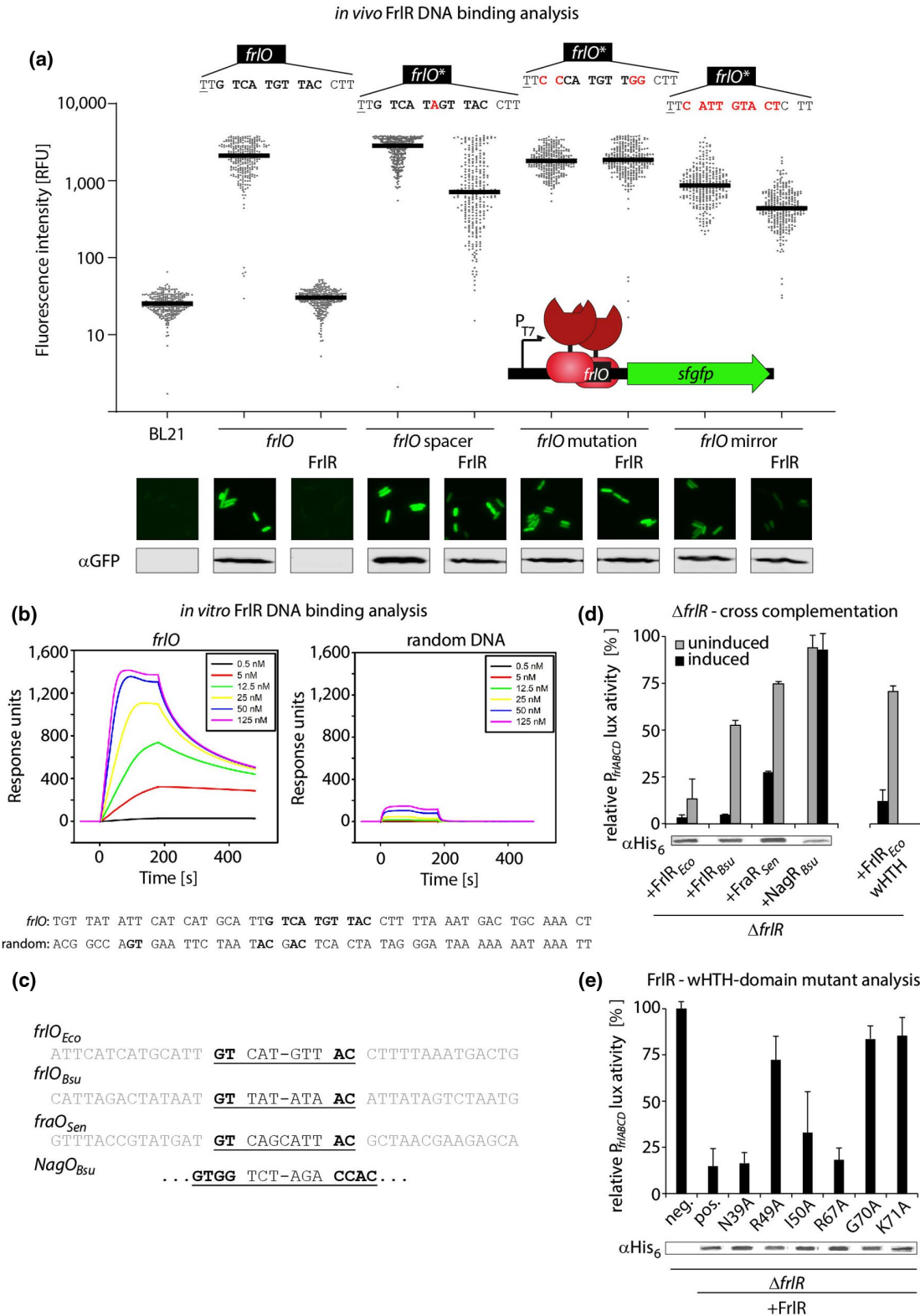


FIGURE 7 In vivo and in vitro analysis of protein/DNA interactions. (a) $frlO_{Eco}$ -dependent sfGFP production: *sfgfp* expression is driven by the T_7 promoter: A sequence was inserted either comprising the putative FrIR_{Eco} recognition motif (*frlO*) or a mutated version (*frlO**). The sequence of the motif inserted downstream to the T_7 promoter is shown with bold letters indicating the *Escherichia coli* *frlO* core motif. Mutations are colored in red and the transcription start site is underlined. The box plot graph shows the fluorescence intensity (given in relative fluorescence units RFU) of 300 individual cells in the presence or absence of FrIR_{Eco}. Pictures were analyzed using ImageJ (Schneider et al., 2012). sfGFP production was visualized by Western blot analysis using anti-GFP specific antibodies (α -GFP). (b) In vitro DNA binding of FrIR_{Eco} to *frlO*_{Eco} analyzed by surface plasmon resonance spectroscopy (SPR) with the biotin-labeled DNA fragment *frlO*_{Eco} (left), and the control fragment without the *frlO*_{Eco} (right) using different concentrations of purified FrIR_{Eco}. (c) Sequence comparison of the operator of *frlO*_{Eco}, *frlO*_{BsuI}, and *nagO*_{BsuI} as well as the putative DNA-binding motif of FraR of *Salmonella enterica*. (d,e) Depicted is the relative $P_{frlABCD}$ -lux activity in % which was compared to the light output of JW5698 cells ($\Delta frlR$) harboring the lux fusion -219/+75. (e) Analyzed are the binding capabilities to *frlO*_{Eco} by the FrIR orthologs depicted in (d) (left) and the *E. coli* FrIR winged helix-turn-helix DNA-binding domain (wHTH) (right). The corresponding genes were ectopically expressed utilizing P_{BAD} with (black bars) and without (grey bars) the addition of 0.2% L-arabinose. The mean values and standard deviations of at least three biological replicates are shown. Protein production was confirmed by Western blot analysis using anti-His₆ specific antibodies (α -His₆). (e) wHTH mutant analysis. Assessed is the capability of FrIR_{Eco} mutant variants to repress the *frlABCD* promoter. Neg.: $\Delta frlR$ Pos.: $\Delta frlR$ +FrIR^{wt}. Protein production was confirmed by Western blot analysis using anti-His₆ specific antibodies (α -His₆)

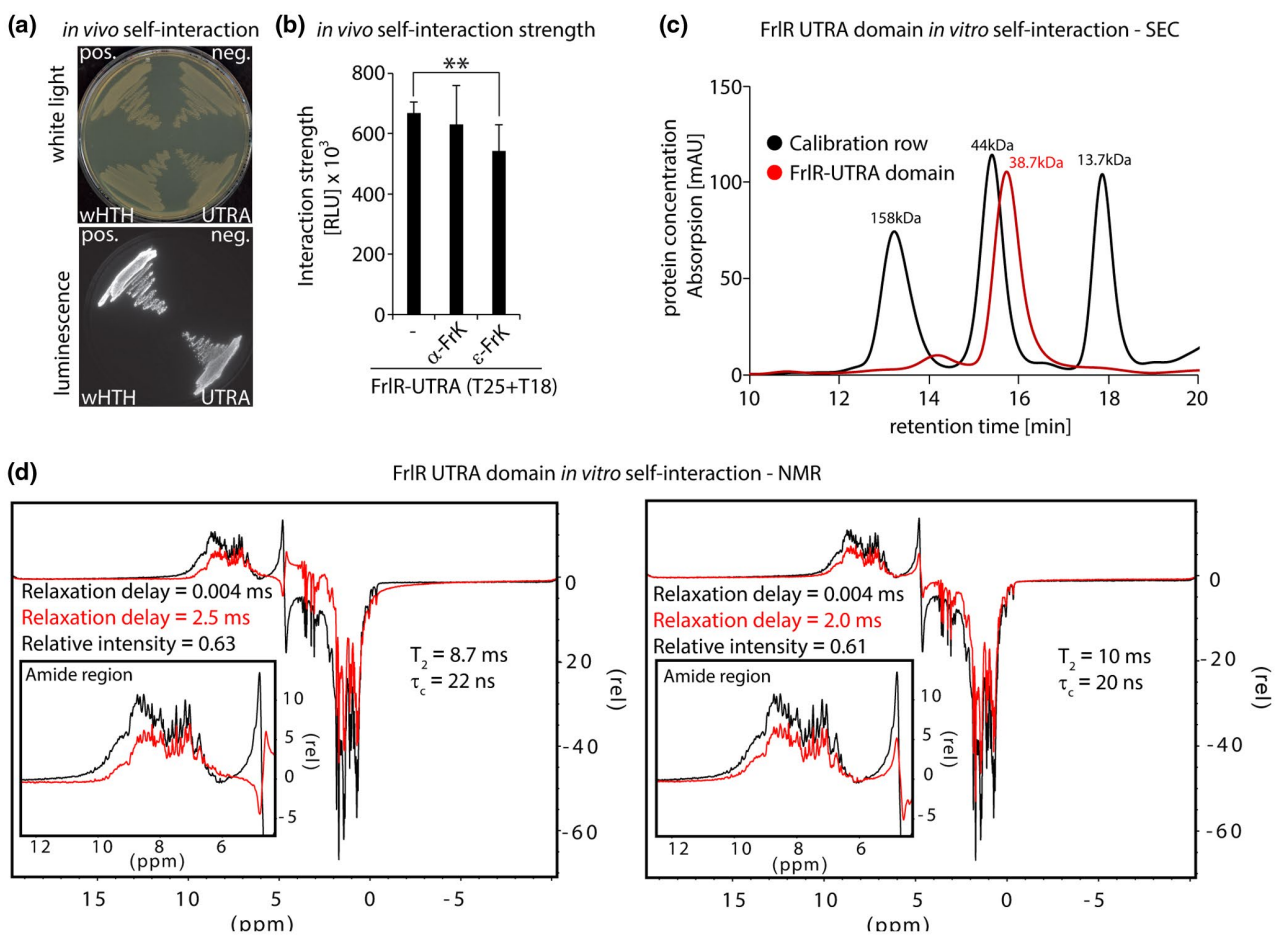


FIGURE 8 FrIR_{Eco} domains in vivo/ in vitro interaction analyses. (a) Qualitative in vivo self-interaction analysis of a T18/T25-FrIR UTRA domain fusion as well as a T18/T25-wHTH-domain fusion. (b) Quantitative in vivo self-interaction analysis of a T18/T25-FrIR UTRA domain fusion with and without (-) supplement of α -FrK or ϵ -FrK. The maximal light emission from a 40 hr time course experiment is given in RLU. Ninety-five percent confidence intervals of at least six replicates are shown. Asterisks indicate significant ($p < .05$) differences in the maximal light emission determined by a T-test for paired samples between cells without supplement and those being exposed to ϵ -FrK. (c) Size exclusion chromatography of purified FrIR_{Eco} UTRA domain. The elution profile is depicted as solid red line. The black line represents a calibration curve of three proteins with varying size (158 kDa, 44 kDa, and 13.7 kDa). Based on this the FrIR_{Eco} UTRA domain migration behavior corresponds to a size of 38.7 kDa. (d) 1D-jump-and-return NMR experiments including 0.004, 2.0, and 2.5 ms relaxation delays from which the transverse relaxation time T_2 is estimated to be 9 ms, allowing to infer a rotational correlation time τ_c of 21 ns assuming isotropic tumbling.

promoter. However, full-length FrIR_{Eco} is needed for high affinity binding.

Next, we generated FrIR_{Eco} mutants based on the sequence alignment with NagR of *B. subtilis* (Figure 2). In NagR_{Bsu}, the wing motif reaches into the minor groove and contacts the flanking guanosine via the carbonyl oxygen of Gly69 (Fillenberg et al., 2015). In addition, the two conserved arginines Arg38 and Arg48 specifically recognize further guanosines of the operator by forming bidental contacts with their corresponding guanine base. Taken together, these three amino acids are assumed to be the hallmark feature of the specific interaction of NagR_{Bsu} with DNA (Fillenberg et al., 2015). Two of the contacts—Gly69 and Arg48—are also conserved throughout FrIR/FraR homologs (Figure 2). Accordingly, we mutated the corresponding residues—Arg49 and Gly70—into alanine resulting in the FrIR_{Eco} variants R49A and G70A. We additionally mutated the putative nucleic acid interacting residues N39, I50, R67, and K71 into alanine, being in the equivalent positions to NagR's Arg38, Met49, and Arg70, respectively. The functionality of the FrIR_{Eco} variants was tested in vivo on their capability to abolish luminescence in $\Delta frlR$ cells encoding the $-219/+75 P_{frIABCD}$ -lux reporter (Figure 7e). Three out of the six mutants—R49A, G70A, and K71A—clearly lost their repressing capability, indicating a role in DNA binding. This led us to conclude that FrIR_{Eco} binds to its recognition motif in a way similar as described for NagR_{Bsu}.

2.5 | FrIR_{Eco} dimerizes via its C-terminal UTRA domain

The predicted structural similarities to NagR_{Bsu} combined with our data on FrIR_{Eco} DNA-binding implies that substrate binding might also occur analogously. Canonically, members of the GntR/HutC family of transcriptional regulators form antiparallel dimers via their UTRA domain to accommodate the two half sites of their palindromic GT/AC flanked DNA-binding motif (Fillenberg et al., 2015; Rigali et al., 2002; Suvorova et al., 2015). Whether this mode of action also applies for FrIR_{Eco}, we examined the dimerization tendency in vitro and in vivo. For the in vivo analysis, the bacterial two-hybrid system (BTH) described by Karimova *et al.* was employed (Karimova et al., 1998) with the exception that all experiments were carried out with our recently published reporter strain, that has a luminescence reporter readout in addition to the original LacZ based colorimetric one (Volkwein et al., 2019). Based on the FrIR_{Eco} homology modeling (Figure 3), we split the protein into two parts, one comprising the N-terminal DNA-binding domain (aa 1-77), the other one encompassing the C-terminal UTRA domain (aa 78-244) and fused them to the T25 and T18 fragments of *Bordetella pertussis* adenylate cyclase. Bioluminescence was recorded qualitatively on agar plates together with the GCN4 leucine zipper as positive (bright phenotype) and T25/T18 solely as negative control (dark phenotype) (Figure 8a). When assessing the self-interaction of the UTRA domain we saw the bright phenotype, whereas clones with the wHTH-domain remained dark, confirming our initial assumption. We also tested whether FrK influences interaction strength or might

even interfere with dimer formation. Consequently, we measured bioluminescence development quantitatively by recording the light output in a 40 hr time course experiment. The maximal RLU are plotted as a bar diagram (Figure 8b). Neither the addition of the α - nor the ϵ -glycated FrK abolished light emission. However, we observed a slight but significant change in interaction strength with the latter. Reportedly, NagR_{Bsu} derepression is not achieved by monomerization upon signal perception (Resch et al., 2010) as, for example, in the case of other one-component systems such as CadC (Buchner et al., 2015; Lindner and White, 2014). Instead, substrate binding to the UTRA domain is transduced into a structural rearrangement of the two wHTH-domains, which hinders proper DNA binding. Similarly, such movement might also reposition the two fragments T18 and T25 and could explain the reduction in bioluminescence with ϵ -FrK. This is plausible as the loop, which is transitioned into a helix upon GlcNAcP/GlcNP binding to NagR_{Bsu} (Fillenberg et al., 2015; Resch et al., 2010), was kept in our fusion construct.

To recapitulate our in vivo observation in vitro, we investigated the oligomerization behavior of the UTRA domain (FrIR-UTRA) by determining its molecular weight by size exclusion chromatography (Figure 8c). In its monomeric form, this would correspond to 19.3 kDa. As expected, and in line with the BTH data, FrIR-UTRA runs at the size of dimers. The addition of ϵ -FrK did not change the migration behavior of FrIR-UTRA (data not shown) further supporting that the regulation does not alter the oligomeric state of the protein. In addition to this, nuclear magnetic resonance (NMR) spectroscopy also confirmed the dimeric state of FrIR-UTRA. The line width of 1H - ^{15}N resonances in 1H - ^{15}N -HSQC spectra is larger than what would be expected for a monomeric FrIR-UTRA (Figure S4a). Receptor-based titration with ϵ -FrK did not change the resonance line width in 1H - ^{15}N -HSQC spectra and we could only observe small chemical shift perturbations (Figure S4a) even with six-fold excess of ϵ -FrK, indicating that interaction with ϵ -FrK is weak and FrIR-UTRA does not become monomeric upon titration with ϵ -FrK. This qualitative assessment of FrIR-UTRA dimerization was followed up with 1D T_2 experiments, in which a relaxation delay has been included in a 1D- 1H experiment to estimate the transverse relaxation time T_2 . From three different relaxation delays (0.004, 2.0 and 2.5 ms) we could estimate T_2 to be 9 ms (Figure 8d). Assuming isotropic tumbling, we estimate a rotational correlation time of 21 ns which would correspond to a molecular weight of 42 kDa. This indicates that FrIR_{Eco} tumbles as a dimer in solution. The weak interaction with ϵ -FrK is confirmed by ligand-based titration using saturation transfer difference (STD)-NMR, where no magnetization transfer could be observed from protein to ligand suggesting an affinity weaker than 10^{-3} M for the FrIR- ϵ -FrK interaction (Figure S4b,c).

2.6 | Fructoselysine-6-phosphate is the cognate effector substrate of FrIR

We have conclusively shown, that ϵ -FrK efficiently induces *E. coli* FrIR dependent derepression in vivo. However, our in vitro FrIR_{Eco}

interaction analysis with the ligand showed only a weak response (Figure S3a) or even failed (Figure S3b,c). Thus, we hypothesized that not ϵ -FrK directly but instead one of its metabolic derivatives is the cognate substrate. This is a plausible explanation as the structural homolog of FrIR_{Eco}—NagR_{Bsu}—recognizes the phosphorylated form of GlcNAc—GlcNAcP (Resch et al., 2010; Rigali et al., 2002). Analogous to this, the only rational candidate is FrK-6P, which is generated by the kinase FrID as first step of the degradation pathway (Figure 1). If the assumption is true, loss of FrID will prevent from ϵ -FrK mediated derepression of P_{frIABCD}. We, therefore, utilized an *E. coli* Δ frID strain containing the $-219/+75$ P_{frIABCD}-lux fusion and measured the light output in the presence and the absence of the ARP. Indeed, the frID⁻ strain lost its ability to respond to ϵ -FrK, as can be inferred from the maximum light emission, which—unlike frID⁺ cells—did not increase with external supplementation of the ARP (Figure 9).

The sequence comparison of FrIR_{Eco} and NagR_{Bsu} revealed high conservation of the phosphate-binding pocket (Figure 2). Therefore, we decided to investigate whether FrK-6P might be accommodated analogously as GlcNAcP in NagR_{Bsu} (Figure 3). Accordingly, we generated the substitution variants located in the UTRA domain, namely T91A, R133A, S166A, and Y168A, and investigated them on their capability to regulate promoter activity in an ϵ -FrK dependent manner once more utilizing the $-219/+75$ P_{frIABCD}-lux fusion. With the

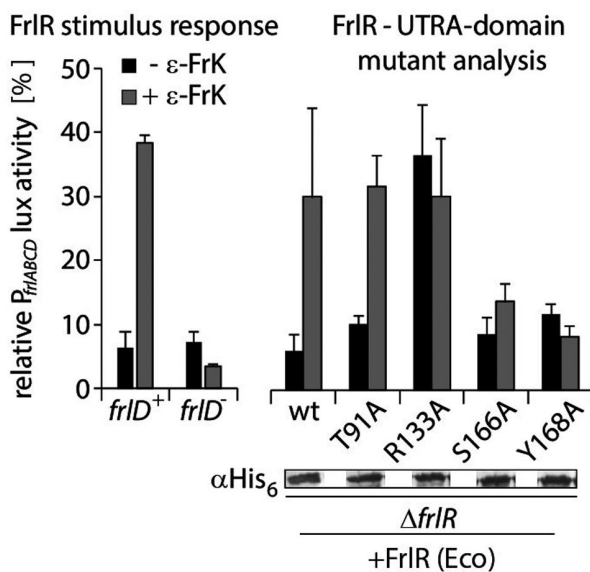


FIGURE 9 FrIR_{Eco} mutant analysis of the putative fructoselysine-6-phosphate binding pocket. Depicted is the relative P_{frIABCD}-lux activity in %, which was compared to the light output of JW5698 cells (Δ frIR) harboring the lux fusion $-219/+75$. *Escherichia coli* JW5698 cells (frID⁺) and JW3337 (frID⁻) that concomitantly express frIR_{Eco} and mutant derivatives from a pBAD33 backbone were analyzed (Guzman et al., 1995). The maximal light emission from a 24 hr time course experiment of cells grown in LB (Miller) (without the addition of the P_{BAD} inducing agent L-arabinose) is given in RLU. The mean values and standard deviations of at least three biological replicates are shown. Protein production of the FrIR_{Eco} variants was confirmed by Western blot analysis using anti-His₆ specific antibodies (α -His₆)

wild-type protein and in the absence of ϵ -FrK we observed a strong repression, exhibited by an about 95% diminished light output compared to cells lacking frIR_{Eco} (Figure 9). The luminescence increased by more than five-fold when the culture was supplemented with the ARP. Deviating from this behavior S166A and Y168A did not respond to the presence of ϵ -FrK with P_{frIABCD} remaining repressed. Similarly, R133A became blind to the inductor but at the same time also lost its DNA-binding capability. We hypothesize that this variant is locked in its inactive state that precludes efficient operator recognition. In summary, these data strongly indicate that FrK-6P is the cognate substrate of FrIR_{Eco} and is recognized similar to GlcNAcP by NagR_{Bsu}.

3 | DISCUSSION

In the present study, the transcriptional regulation of FrK metabolism in *E. coli* was investigated. Our data show that the expression of the frIABCD operon is controlled by global and specific stimuli. We identified two promoter regions, both stimulated by the alternative sigma factor RpoH (σ 32). The induction of the other σ 70-dependent promoter is subject to cAMP/CRP triggered catabolite repression. A repressor FrIR_{Eco} further prevents transcription in the absence of FrK. However, this substrate has to be processed to FrK-6P by the kinase FrID to be recognized as the cognate stimulus. Accordingly, a basal frIABCD gene expression is necessary even under repressive conditions. Several scenarios are conceivable to achieve such a goal. For instance, the copy number of the FrIR_{Eco} protein could be limited to such an extent that not every cell has a sufficient amount for effective repression. In this case, an unequal distribution of the permease FrIA and kinase FrID occurs within the population, which enables an individual response to the carbon source. The resulting heterogeneity might be particularly pronounced because of the complex regulation and the necessity to transport and modify ϵ -FrK to induce the cell response. The so-called "bet hedging" strategy (van Vliet and Ackermann, 2015) would be particularly advantageous here: Limiting the FrK metabolism to a subset of cells allows the simultaneous utilization of different carbon sources by the entire population. In the intestine, *E. coli* is confronted with exactly such a situation and is, thus, able to compete with other bacteria of the gut microbiota. In the same context, the specialization of the metabolism toward certain ARPs is also beneficial. While *E. coli* exclusively utilizes ϵ -fructose/psicoselysine, *B. subtilis* favors various α -glycated amino acids (Wiame et al., 2004, Wiame and Van Schaftingen, 2004), thus creating distinct niches and minimizing competition. However, such selectivity also demands for species specific regulation patterns. Global transcriptional control seems to be adjusted to the respective organism with *B. subtilis* being dependent on CodY, a transcriptional regulator that helps to adapt to changes in nutrient availability (Deppe et al., 2011) and *E. coli* where ARP metabolism is activated by CRP/cAMP and σ ³² instead. By contrast, the substrate-specific FrIR proteins from both organisms are 63% similar and even recognize the same operator sequence. It is therefore possible, that both FrIRs also respond to the same stimulus. However, in our assays

we saw a highly specific response of *E. coli* FrIR toward ϵ -FrK. As only its phosphorylated form is the cognate signal for FrIR_{Ec}, an elegant way to achieve regulatory specificity could come through distinct substrate specificities of certain enzymes in the degradation pathways. Notably, the FrID kinases from *E. coli* and *B. subtilis* differ in their enzymatic properties (Wiame et al., 2004). The K_M of ϵ -FrK for instance is 20 μ M with FrID_{Ec}, the corresponding one for FrID_{Bsu} (YurL) is three orders of magnitude higher (14 mM). Conversely, α -fructosevaline is processed efficiently by the *B. subtilis* enzyme at a concentration of about 100 μ M, whereas the *E. coli* counterpart has a K_M above 20 mM. The selectivity of the metabolism can also explain how FrIR_{Bsu} is able to respond to chemically diverse α -glycated substrates. Specifically, substrate recognition could simply be achieved by the combination of sugar 6-phosphate and C₁-N linkage, two structure elements shared by all phosphorylated ARPs derived from glucose.

4 | EXPERIMENTAL PROCEDURES

4.1 | Bacterial strains, plasmids, and growth conditions

Bacterial strains, plasmids, and primers are listed in Tables S1–S3. *E. coli* was routinely cultivated in LB according to the Miller modification (Bertani, 1951; Miller, 1992) or M9 minimal medium (Miller, 1972) unless indicated otherwise. For solidification 1.5% (w/v) agar was added to the medium. If needed, carbon sources and other media supplements were added as indicated. Antibiotics were used at the following final concentrations: 100 μ g/ml ampicillin sodium salt, 50 μ g/ml kanamycin sulfate, 30 μ g/ml chloramphenicol, or 20 μ g/ml gentamycin sulfate. Growth was recorded by measuring the optical density at a wavelength of 600 nm. Plasmids carrying the pBAD (Guzman et al., 1995) or T₇ promoter were induced with L-arabinose at a final concentration of 0.2% (w/v) or Isopropyl- β -D-thiogalactopyranosid (IPTG) at a final concentration of 1 mM.

All kits and enzymes employed for plasmid construction were used according to manufacturer's instructions: Plasmid DNA was isolated using the Hi Yield® Plasmid Mini Kit from Süd-Laborbedarf GmbH. DNA fragments were purified from agarose gels using the Hi Yield® Gel/PCR DNA fragment extraction kit from Süd-Laborbedarf GmbH. All restriction enzymes, DNA modifying enzymes, and the Q5® high fidelity DNA polymerase for PCR amplification were purchased from New England BioLabs GmbH. A detailed description for plasmid construction is given in Table S2.

4.2 | Synthesis and analysis of ARPs

N- ϵ - and *N*- α -FrK were synthesized and isolated according to previous publications (Hellwig et al., 2011; Krause et al., 2003) and met the spectroscopic properties given in those works. Analysis of both ARPs as well as lysine was performed by amino acid analysis with the

analyzer S 433 (Sykam) on a cation-exchange column (LCA K07/Li; 150 mm \times 4.6 mm, 7 μ m). Before analysis, 10 μ l of solutions from microorganism culture was diluted with 190 μ l of loading buffer (0.12 M lithium citrate, pH 2.12), centrifuged (10.000 \times g, 10 min), and 40 μ l of the diluted sample was injected. Separation was accomplished with custom lithium citrate buffers of increasing pH and ionic strength. Amino acids were detected by online post-column derivatization with ninhydrin and UV-detection (570 nm). Calibration was performed by an external standard of proteinogenic amino acids, and both FrK derivatives were quantified as lysine (Hellwig et al., 2011).

4.3 | SDS-PAGE and western blotting

For protein analyses cells were subjected to 12.5% (w/v) sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) as described by Laemmli (Laemmli, 1970). To visualize proteins by UV light 2,2,2-trichloroethanol was added to the polyacrylamide gels (Ladner et al., 2004). Subsequently, the proteins were transferred onto nitrocellulose membranes, which were then subjected to immunoblotting. In a first step the membranes were incubated either with 0.1 μ g/ml anti-6xHis® antibody (Abcam) or with 0.1 μ g/ml anti-GFP (Sigma-Aldrich) antibody. These primary antibodies (rabbit) were targeted with 0.2 μ g/ml anti-rabbit alkaline phosphatase-conjugated secondary antibody (Rockland) or 0.1 μ g/ml anti-rabbit IgG (IRDye® 680RD) (donkey) antibodies (Abcam). Anti-rabbit IgG was detected by adding development solution [50 mM sodium carbonate buffer, pH 9.5, 0.01% (w/v) p-nitro blue tetrazolium chloride (NBT), and 0.045% (w/v) 5-bromo-4-chloro-3-indolyl-phosphate (BCIP)]. Anti-rabbit IgG were visualized via Odyssey® CLx Imaging System (LI-COR, Inc).

4.4 | Bacterial two-hybrid assay

Protein-protein interactions were detected using the bacterial adenylate cyclase two-hybrid system kit (Euromedex) according to product manuals (Karimova et al., 2000). This system is based on functional reconstitution of split *Bordetella pertussis* adenylate cyclase CyaA, which catalyzes the formation of cyclic AMP from ATP. In *E. coli* KV1, the cAMP dependent *lac* promoter P_{lac} precedes a translational fusion of the *lux*-operon and *lacZ*, allowing the indirect measurement of protein-protein-interactions by light emission and colorimetric detection (Volkwein et al., 2019).

For measuring interaction strength, chemically competent (Inoue et al., 1990) *E. coli* KV1 cells were transformed with pKT25-*frlR* FL, pKT25-*frlR* UTRA domain, or pKT25-*frlR* HTH-domain and pUT18C-*frlR* FL, pUT18C-*frlR* UTRA domain, or pUT18C-*frlR* HTH-domain. Transformants containing pUT18-*zip*/pKT25-*zip* and pUT18C/pKT25 vector backbones were used as positive and negative controls, respectively. Single colonies were inoculated in LB (with 50 μ g/ml kanamycin sulfate and 100 μ g/ml ampicillin sodium salt, 5 mM α - ϵ -FrK and 0.5 mM IPTG (w/v)) and grown aerobically at 37°C o/n. The next day, a microtiter plate with fresh LB (with

the appropriate antibiotics, 10 mM substrate and 0.5 mM IPTG (w/v) was inoculated with the cells at an OD_{600} of 0.01. The cells were grown aerobically in the Tecan Infinite F500 system (TECAN) at 30°C. OD_{600} and luminescence were recorded in 10 min intervals over the course of 16 hr. Each measurement was performed at least in triplicate.

For qualitative analysis, the bacterial two-hybrid KV1 strains were plated on LB agar (containing 50 µg/ml kanamycin sulfate, 100 µg/ml ampicillin sodium salt, and 0.5 mM IPTG) and grown overnight at 37°C. Pictures of the plates were taken in a Fusion-SL 3500 WL (PEQLAB) using 10 s of exposure time.

4.5 | Luminescence activity assay

E. coli cells harboring a lux fusion plasmid (Table S2—reporter assays) were inoculated in LB (with appropriate antibiotics and 0.2% arabinose (w/v)) and grown aerobically at 37°C. The next day, a microtiter plate with fresh LB (with the appropriate antibiotics and 0.2% arabinose (w/v)) was inoculated with the cells at an OD_{600} of 0.01. The cells were grown aerobically in the Tecan Infinite F500 system (TECAN) at 37°C. OD_{600} and luminescence were recorded in 10 min intervals over the course of 16 hr. Light units were normalized to OD_{600} and are thus expressed in relative light units (RLU). Each measurement was performed in triplicate.

To examine the regulatory role of rpoH, the growth temperature was adjusted to 20°C and 25°C during the measurement. The binding site for CAP was examined in LB with or without 20 mM glucose, thereby repressing or activating CyaA, respectively.

4.6 | Surface plasmon resonance spectroscopy

Surface Plasmon Resonance Spectroscopy (SPR) spectroscopy and calibration free concentration (CFCA) assays were performed using a Biacore T200 device (GE Healthcare) and streptavidin-precoated Xantec SAD500-L carboxymethyl dextran sensor chips (XanTec Bioanalytics GmbH, Düsseldorf, Germany). All experiments were conducted at 25°C with HBS-EP+ buffer [10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, and 0.05% (v/v) detergent P20].

Before immobilizing the DNA fragments, the chips were equilibrated by three injections using 1 M NaCl/50 mM NaOH at a flow rate of 10 µl min⁻¹. Then, 10 nM of the respective double-stranded biotinylated DNA fragment was injected using a contact time of 420 s and a flow rate of 10 µl min⁻¹. As a final wash step, 1 M NaCl/50 mM NaOH/50% (v/v) isopropanol was injected. Approximately 250–350 RU of each respective DNA fragment were captured onto the respective flow cell. All interaction kinetics of FlrR with the respective DNA fragment were performed in HBS-EP buffer at 25°C at a flow rate of 30 µl min⁻¹ in the presence and absence of 1 mM ε-FrK. The proteins were diluted in HBS-EP buffer and passed over all flow cells in

different concentrations (0.5 nM–125 nM) using a contact time of 180 s followed by a 300 s dissociation time before the next cycle started. After each cycle the surface was regenerated by injection of 2.5 M NaCl for 30 s at 60 µl min⁻¹ flow rate followed by a second regeneration step by injection of 0.5% (w/v) SDS for 30 s at 60 µl min⁻¹. All experiments were performed at 25°C. Sensorgrams were recorded using the Biacore T200 Control software 2.0 and analyzed with the Biacore T200 Evaluation software 2.0. The surface of flow cell 1 was not immobilized with DNA and used to obtain blank sensorgrams for subtraction of bulk refractive index background. The referenced sensorgrams were normalized to a baseline of 0. Peaks in the sensorgrams at the beginning and the end of the injection emerged from the run-time difference between the flow cells of each chip.

Calibration-free concentration analysis (CFCA) was performed using a 1 µM solution of purified FrIR_{Eco}, which was stepwise diluted 1:2, 1:5, 1:10, and 1:20. Each protein dilution was injected two-times, one at 5 µl min⁻¹ as well as 100 µl min⁻¹ flow rate. On the active flow cell DNA fragment including the P_{FlrO}-DNA was used for FlrR binding. CFCA basically relies on mass transport, which is a diffusion phenomenon that describes the movement of molecules between the solution and the surface. The CFCA therefore relies on the measurement of the observed binding rate during sample injection under partially or complete mass transport limited conditions. Overall, the initial binding rate (dR/dt) is measured at two different flow rates dependent on the diffusion constant of the protein. The diffusion coefficient of BceR-P was calculated using the Biacore diffusion constant calculator and converter webtool (https://www.biacore.com/lifesciences/Application_Support/online_support/Diffusion_Coefficient_Calculator/index.html), whereby a globular shape of the protein was assumed. The diffusion coefficient of FlrR was determined as $D = 1.02 \times 10^{-10} \text{ m}^2/\text{s}$. The initial rates of those dilutions that differed in a factor of at least 1.5 were considered for the calculation of the “active” concentration, which was determined as $5 \times 10^{-7} \text{ M}$ (approximately 50% of the total protein concentration) for FlrR. The “active” protein concentration was then used for calculation of the binding kinetic constants and steady-state affinity.

4.7 | Protein purification

His₆-SUMO-tagged FraR from *S. enterica* (pET-SUMO-*fraR* (Sen)), FrIR from *E. coli* (pET-SUMO-*frlR* (Eco)) and *B. subtilis* (pET-SUMO-*frlR* (Bsu)) (Table S2—Protein overexpression) were overproduced in *E. coli* BL21 (DE3) by addition of 1 mM IPTG to exponentially growing cells and subsequent cultivation at 18°C o/n. Cells were lysed by sonication in the respective buffer (Table S4). The proteins were purified using Ni-nitrilotriacetic acid (Ni-NTA; Qiagen) according to the manufacturer's instructions, using 20 mM imidazole for washing and 250 mM imidazole for elution. Subsequently, imidazole was removed by dialysis o/n at 4°C in buffer 1. The His₆-SUMO tag was cleaved by incubation with His₆-Ulp1 (Starosta et al., 2014) overnight.

Subsequently, tag-free FrIR_{Eco} was collected from the flow through after metal chelate affinity chromatography.

Size exclusion chromatography was performed in the respective buffer (Table S4) using a Superdex 200 Increase 10/300-GI column with a flow rate of 0.3 ml/min on an Äkta purifier (GE Healthcare). Four milligrams of protein were loaded in a volume of 0.4 ml (8,7 mg/ml). Eluting protein was detected at 280 nm. Fractions of 0.5 ml were collected.

4.8 | NMR experiments

All NMR spectra were acquired using a Bruker Avance III NMR spectrometer with a magnetic field strength corresponding to a proton Larmor frequency of 700 MHz equipped with a room temperature triple resonance gradient probe head. NMR experiments were performed in a buffer containing 100 mM potassium phosphate, 300 mM NaCl, pH 6.5 at 298 K. Two dimensional ¹H-¹⁵N-HSQC titrations were performed with 200 μM FrIR-UTRA domain in the absence and presence of 1.2 mM excess of ε-FrK. For measuring T₂ relaxation, one dimensional experiments were performed with 200 μM FrIR-UTRA in the presence of 1.2 mM ε-FrK. 1D T₂ experiments were performed using a 1-1 echo pulse sequence with a relaxation delay varying between 0.004, 2.0 and 2.5 ms (Sklenář and Bax, 1987). τ_c was estimated using the equation 1/(5T₂) where T₂ is expressed in seconds (Barbato et al., 1992; Kay et al., 1989). Saturation transfer difference NMR experiments (Mayer and Meyer, 1999) were performed on 10 μM of FrIR-UTRA + 1mM ε-FrK with irradiation at either 0.65 ppm (protein methyl region) or 8.5 ppm (protein amide region), far from ε-FrK signals to only saturate protein, with a relaxation delay of 5 s, which includes an effective saturation time of 4 s and an interscan delay of 1 s. For control experiments only 1 mM ε-FrK was used.

4.9 | Fluorescence microscopy

The DNA-binding site of FrIR_{Eco} was examined by fluorescence microscopy of the *E. coli* strain BL21 (DE3) transformed with pUC19 based fluorescent reporter plasmids (Table S2—Reporter assays). The cells were cultivated overnight at 30°C in LB medium supplemented with ampicillin. Expression of FrIR_{Eco} was induced or repressed by addition of 20 mM arabinose or 20 mM glucose, respectively. The overnight cultures were used to inoculate (OD₆₀₀ of 0.1) fresh LB medium supplemented with ampicillin and arabinose or glucose. Cells were aerobically cultivated at 37°C and harvested by gentle centrifugation after 2 hr. The pellet was washed twice and resuspended in PBS (OD₆₀₀ 0.5). 2 μl of the culture was spotted on 2% (w/v) agarose pads, placed onto microscopic slides and covered with a coverslip. Subsequently, images were taken on a Leica DMi8 inverted microscope equipped with a Leica DFC365 FX camera (Wetzlar). An excitation wavelength of 484 nm and a 535 nm emission filter with a 75-nm bandwidth was used for sfGFP fluorescence

for 1 s, gain 5, and 75% intensity. Fluorescence intensity of at least 300 cells was measured and plotted using ImageJ.

4.10 | Thermal shift assay

About 5 μM of the Protein (FrIR_{Eco}) was pipetted into 30 μl (2.5 μM) of DNA (cleaned up PCR product) on ice. About 0.3 μl of a 1:10 dilution of SYPRO™ orange protein gel stain (Thermo Fisher Scientific) was added to the mix. The mixture was transferred into a 96-well semi-skirted PCR plate. To ensure that the liquid was at the bottom the plate was centrifuged for 30 s at 3,000 rpm. Subsequently, the plate was inserted into iQ™5 Real-Time PCR Detection Systems (Bio-Rad). The 96-well plate was incubated at 8°C for 10 min. Every 30 s the temperature was increased by 0.5°C until 90°C were reached. Fluorescence was recorded at a wavelength of 520 nm.

ACKNOWLEDGMENTS

We thank Ralph Krafczyk and Kirsten Jung for fruitful discussions. Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTION

Biochemical and genetic analyses on FrIR/FraR/NagR were conducted by N.G., F.K., and B.G.v.A. NMR studies were performed by P.K.A.J. and J.H. The corresponding proteins were produced and purified by B.G.v.A., N.G. and F.K. M.H. and T.H. synthesized ARPs and analyzed ε/α-FrK turnover. J.L. designed the study. The manuscript was written by B.G.v.A., N.G. and F.K., P.K.A.J., J.H., M.H. and J.L.

DATA AVAILABILITY STATEMENT

All data generated or analysed during this study are included in this published article (and its supplementary information files).

ORCID

Benedikt Graf von Armansperg  <https://orcid.org/0000-0002-4717-0176>

[org/0000-0002-4717-0176](https://orcid.org/0000-0002-4717-0176)

Franziska Koller  <https://orcid.org/0000-0003-2003-7129>

Nicola Gericke  <https://orcid.org/0000-0002-1500-5053>

Michael Hellwig  <https://orcid.org/0000-0001-8528-6893>

Pravin Kumar Ankush Jagtap  <https://orcid.org/0000-0002-9457-4130>

[org/0000-0002-9457-4130](https://orcid.org/0000-0002-9457-4130)

Ralf Heermann  <https://orcid.org/0000-0003-0631-6156>

Janosch Hennig  <https://orcid.org/0000-0001-5214-7002>

Thomas Henle  <https://orcid.org/0000-0001-6220-7267>

Jürgen Lassak  <https://orcid.org/0000-0003-3936-3389>

REFERENCES

Ali, M.M., Newsom, D.L., Gonzalez, J.F., Sabag-Daigle, A., Stahl, C., Steidley, B., et al. (2014) Fructose-asparagine is a primary nutrient

- during growth of *Salmonella* in the inflamed intestine. *PLoS Pathogens*, 10, e1004209.
- Amadori, M. (1925) Products of condensation between glucose and p-phenetidine I. *Atti dell'Accademia Pontificia de'Nuovi Lincei*, 2, 337–342.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, 2, 2006.0008.
- Barbato, G., Ikura, M., Kay, L.E., Pastor, R.W. and Bax, A. (1992) Backbone dynamics of calmodulin studied by nitrogen-15 relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry*, 31, 5269–5278.
- Barroso-Batista, J., Pedro, M.F., Sales-Dias, J., Pinto, C.J.G., Thompson, J.A., Pereira, H., et al. (2020) Specific eco-evolutionary contexts in the mouse gut reveal *Escherichia coli* metabolic versatility. *Current Biology*, 30, 1049–1062.e7.
- Bertani, G. (1951) Studies on lysogeny. I. The mode of phage liberation by lysogenic *Escherichia coli*. *Journal of Bacteriology*, 62, 293–300.
- Bertani, G. (2004) Lysogeny at mid-twentieth century: P1, P2, and other experimental, systems. *Journal of Bacteriology*, 186, 595–600.
- Buchner, S., Schlundt, A., Lassak, J., Sattler, M. and Jung, K. (2015) Structural and functional analysis of the signal-transducing linker in the pH-responsive one-component system CadC of *Escherichia coli*. *Journal of Molecular Biology*, 427, 2548–2561.
- Deppe, V.M., Klatte, S., Bongaerts, J., Maurer, K.H., O'Connell, T. and Meinhardt, F. (2011) Genetic control of amadori product degradation in *Bacillus subtilis* via regulation of *frlBONMD* expression by *FrIR*. *Applied and Environment Microbiology*, 77, 2839–2846.
- Deutscher, J., Ake, F.M., Derkaoui, M., Zebre, A.C., Cao, T.N., Bouraoui, H., et al. (2014) The bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system: regulation by protein phosphorylation and phosphorylation-dependent protein-protein interactions. *Microbiology and Molecular Biology Reviews*, 78, 231–256.
- Fillenberg, S.B., Friess, M.D., Korner, S., Bockmann, R.A. and Muller, Y.A. (2016) Crystal structures of the global regulator DasR from *Streptomyces coelicolor*: implications for the allosteric regulation of GntR/HutC repressors. *PLoS One*, 11, e0157691.
- Fillenberg, S.B., Grau, F.C., Seidel, G. and Muller, Y.A. (2015) Structural insight into operator dre-sites recognition and effector binding in the GntR/HutC transcription regulator NagR. *Nucleic Acids Research*, 43, 1283–1296.
- Griffiths, D.R. and Pridham, J.B. (1980) The metabolism of 1-(ϵ -N-lysyl)-1-deoxy-D-fructose by *Escherichia coli*. *Journal of the Science of Food and Agriculture*, 31, 1214–1220.
- Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose pBAD promoter. *Journal of Bacteriology*, 177, 4121–4130.
- Hellwig, M., Geissler, S., Matthes, R., Peto, A., Silow, C., Brandsch, M., et al. (2011) Transport of free and peptide-bound glycosylated amino acids: synthesis, transepithelial flux at Caco-2 cell monolayers, and interaction with apical membrane transport proteins. *ChemBioChem*, 12, 1270–1279.
- Henle, T. (2003) AGEs in foods: do they play a role in uremia? *Kidney International*, 63(Suppl. 84), S145–S147.
- Henle, T. (2005) Protein-bound advanced glycation endproducts (AGEs) as bioactive amino acid derivatives in foods. *Amino Acids*, 29, 313–322.
- Hodge, J.E. (1955) The Amadori rearrangement. *Advances in Carbohydrate Chemistry*, 10, 169–205.
- Hoskisson, P.A. and Rigali, S. (2009) Chapter 1: variation in form and function the helix-turn-helix regulators of the GntR superfamily. *Advances in Applied Microbiology*, 69, 1–22.
- Huerta, A.M. and Collado-Vides, J. (2003) Sigma70 promoters in *Escherichia coli*: specific transcription in dense regions of overlapping promoter-like signals. *Journal of Molecular Biology*, 333, 261–278.
- Huynh, K. and Partch, C.L. (2015) Analysis of protein stability and ligand interactions by thermal shift assay. *Current Protocols in Protein Science*, 79, 28.9.1–28.9.14.
- Inoue, H., Nojima, H. and Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, 96, 23–28.
- Karimova, G., Pidoux, J., Ullmann, A. and Ladant, D. (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 5752–5756.
- Karimova, G., Ullmann, A. and Ladant, D. (2000) *Bordetella pertussis* adenylate cyclase toxin as a tool to analyze molecular interactions in a bacterial two-hybrid system. *International Journal of Medical Microbiology*, 290, 441–445.
- Kay, L.E., Torchia, D.A. and Bax, A. (1989) Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, 28, 8972–8979.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10, 845–858.
- Krause, R., Knoll, K. and Henle, T. (2003) Studies on the formation of furosine and pyridosine during acid hydrolysis of different Amadori products of lysine. *European Food Research and Technology*, 216, 277–283.
- Ladner, C.L., Yang, J., Turner, R.J. and Edwards, R.A. (2004) Visible fluorescent detection of proteins in polyacrylamide gels without staining. *Analytical Biochemistry*, 326, 13–20.
- Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227, 680–685.
- Lassak, J., Koller, F., Krafczyk, R. and Volkwein, W. (2019) Exceptionally versatile—arginine in bacterial post-translational protein modifications. *Biological Chemistry*, 400(11), 1397–1427.
- Lawson, C.L., Swigon, D., Murakami, K.S., Darst, S.A., Berman, H.M. and Ebright, R.H. (2004) Catabolite activator protein: DNA binding and transcription activation. *Current Opinion in Structural Biology*, 14, 10–20.
- Lindner, E. and White, S.H. (2014) Topology, dimerization, and stability of the single-span membrane protein CadC. *Journal of Molecular Biology*, 426, 2942–2957.
- Maillard, L.C. (1912a) The action of amino acids on sugar; The formation of melanoidin by a methodic route. *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences*, 154, 66–68.
- Maillard, L.C. (1912b) General reaction of amino acids on sugars: its biological consequences. *Comptes Rendus des Seances de la Societe de Biologie et de Ses Filiales*, 72, 599–601.
- Mayer, M. and Meyer, B. (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angewandte Chemie International Edition in English*, 38, 1784–1788.
- Miller, J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, pp. 221–222 and 263–274.
- Miller, J.H. (1992) *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press N.Y.
- Miller, K.A., Phillips, R.S., Kilgore, P.B., Smith, G.L. and Hoover, T.R. (2015) A mannose family phosphotransferase system permease and associated enzymes are required for utilization of fructoselysine and glucoselysine in *Salmonella enterica* Serovar Typhimurium. *Journal of Bacteriology*, 197, 2831–2839.
- Osumi, T. and Saier, M.H. (1982) Regulation of lactose permease activity by the phosphoenolpyruvate-sugar phosphotransferase system—evidence for direct binding of the glucose-specific enzyme III to the

- lactose permease. *Proceedings of the National Academy of Sciences of the United States of America*, 79, 1457–1461.
- Resch, M., Schiltz, E., Titgemeyer, F. and Muller, Y.A. (2010) Insight into the induction mechanism of the GntR/HutC bacterial transcription regulator YvoA. *Nucleic Acids Research*, 38, 2485–2497.
- Rigali, S., Derouaux, A., Giannotta, F. and Dusart, J. (2002) Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. *Journal of Biological Chemistry*, 277, 12507–12515.
- Sabag-Daigle, A., Wu, J.K., Borton, M.A., Sengupta, A., Gopalan, V., Wrighton, K.C., et al. (2018) Identification of bacterial species that can utilize fructose-asparagine. *Applied and Environmental Microbiology*, 84, e01957–17.
- Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sanchez-Perez, M., Gomez-Romero, L., Ledezma-Tejeda, D., et al. (2019) RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Research*, 47, D212–D220.
- Schlundt, A., Buchner, S., Janowski, R., Heydenreich, T., Heermann, R., Lassak, J., et al. (2017) Structure-function analysis of the DNA-binding domain of a transmembrane transcriptional activator. *Scientific Reports*, 7, 1051.
- Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9, 671–675.
- Shimada, T., Fujita, N., Yamamoto, K. and Ishihama, A. (2011) Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources. *PLoS One*, 6, e20081.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7, 539.
- Sklenář, V. and Bax, A. (1987) Spin-echo water suppression for the generation of pure-phase two-dimensional NMR spectra. *Journal of Magnetic Resonance*, 74, 469–479.
- Sondej, M., Weinglass, A.B., Peterkofsky, A. and Kaback, H.R. (2002) Binding of enzyme IIA^{Glc}, a component of the phosphoenolpyruvate:sugar phosphotransferase system, to the *Escherichia coli* lactose permease. *Biochemistry*, 41, 5556–5565.
- Starosta, A.L., Lassak, J., Peil, L., Atkinson, G.C., Woolstenhulme, C.J., Virumae, K., et al. (2014) A conserved proline triplet in Val-tRNA synthetase and the origin of elongation factor P. *Cell Reports*, 9, 476–483.
- Suvorova, I.A., Korostelev, Y.D. and Gelfand, M.S. (2015) GntR family of bacterial transcription factors and their DNA binding motifs: Structure, positioning and co-evolution. *PLoS One*, 10, e0132618.
- Ulrich, P. and Cerami, A. (2001) Protein glycation, diabetes, and aging. *Recent Progress in Hormone Research*, 56, 1–21.
- van Vliet, S. and Ackermann, M. (2015) Bacterial ventures into multicellularity: Collectivism through individuality. *PLOS Biology*, 13, e1002162.
- Volkwein, W., Krafczyk, R., Jagtap, P.K.A., Parr, M., Mankina, E., Macosek, J., et al. (2019) Switching the post-translational modification of translation elongation factor EF-P. *Frontiers in Microbiology*, 10, 1148.
- Volkwein, W., Maier, C., Krafczyk, R., Jung, K. and Lassak, J. (2017) A versatile toolbox for the control of protein levels using N_ε-acetyl-L-lysine dependent amber suppression. *ACS Synthetic Biology*, 6, 1892–1902.
- Wiame, E., Delpierre, G., Collard, F. and Van Schaftingen, E. (2002) Identification of a pathway for the utilization of the Amadori product fructoselysine in *Escherichia coli*. *Journal of Biological Chemistry*, 277, 42523–42529.
- Wiame, E., Duquenne, A., Delpierre, G. and Van Schaftingen, E. (2004) Identification of enzymes acting on alpha-glycated amino acids in *Bacillus subtilis*. *FEBS Letters*, 577, 469–472.
- Wiame, E., Lamosa, P., Santos, H. and Van Schaftingen, E. (2005) Identification of glucoselysine-6-phosphate deglycase, an enzyme involved in the metabolism of the fructation product glucoselysine. *The Biochemical Journal*, 392, 263–269.
- Wiame, E. and Van Schaftingen, E. (2004) Fructoselysine 3-epimerase, an enzyme involved in the metabolism of the unusual Amadori compound psicoselysine in *Escherichia coli*. *The Biochemical Journal*, 378, 1047–1052.
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y. (2015) The I-TASSER Suite: protein structure and function prediction. *Nature Methods*, 12, 7–8.
- Zhao, K., Liu, M. and Burgess, R.R. (2005) The global transcriptional response of *Escherichia coli* to induced sigma 32 protein involves sigma 32 regulon activation followed by inactivation and degradation of sigma 32 in vivo. *Journal of Biological Chemistry*, 280, 17758–17768.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

How to cite this article: Graf von Armanberg B, Koller F, Gericke N, et al. Transcriptional regulation of the N_ε-fructoselysine metabolism in *Escherichia coli* by global and substrate-specific cues. *Mol Microbiol.* 2020;00:1–16.
<https://doi.org/10.1111/mmi.14608>

7 Concluding discussion and outlook

7.1 Alternative TDP-Rha biosynthesis gene

Monosaccharides have to be activated to act as glycosyl donors in glycosylation reactions. The activated form of the sugar L-rhamnose in bacteria is TDP-Rha, which is biosynthesized originating from glucose-1-phosphate *via* the RmlBDAC pathway. Disruption of genes in the Rml pathway leads to severe growth defects and partially to loss of pathogenicity in human pathogen bacteria (68-72). These findings emphasise the importance of the glycosylation with L-rhamnose. Interestingly, the mutant phenotype is nearly absent in *S. oneidensis* and *P. aeruginosa* when deleting the gene coding for the epimerase RmlC of the pathway. In chapter 2 we therefore investigated whether alternative biosynthesis enzymes exist in these bacteria using two approaches: First, generating a genomic library of the related model organism *P. putida* KT2440 and second, a bioinformatical analyses of potential analogues.

We could identify a *rmlC* homologue named *PP_0265*. They share high amounts of identities in protein sequences (64 %) and consequently fold similarity in predicted structures using Phyre2 (138). Taken together with the fact that they carry out identical function, it can be assumed that they share a common ancestry (139). Gene duplication can immediately result in an increase in gene dosage (140) which might be beneficial when it leads to an increased fitness. Interestingly, the equilibrium for the epimerisation by RmlC lies heavily on the side of the substrate dTDP-4-keto-6-deoxy-D-glucose with only 3 % of conversion to the product dTDP-4-keto-6-deoxy-L-mannose (141, 142). The complex double epimerase reaction consists of four distinct chemical steps which require careful steric control (143). The resulting keto-sugar is known to be unstable and is only available in small amounts. Efficient biosynthesis of TDP-Rha is enabled due to the fact that the formation of dTDP-D-glucose by RmlA is nearly irreversible and that the equilibrium of the reduction catalysed by RmlD lies strongly on the side of TDP-Rha (142). The gene duplication of *rmlC* could lead to increased availability of the epimerase and subsequently higher amount of the substrate dTDP-4-keto-6-deoxy-L-mannose and further TDP-Rha. Interestingly, *P. putida* cells hold high concentrations of intracellular TDP-Rha which are even higher than *P. aeruginosa* (3.5 mM and 2.0 mM, respectively) (91). This is surprising as *P. putida* does not produce rhamnolipids or rhamnose-containing LPS like *P. aeruginosa*. These findings provide evidence that TDP-Rha plays an important role in so far undetected cellular processes. Actually, the rates of biosynthesis of TDP-Rha and other nucleotide sugars were shown to be limiting for glycosylation rates (144, 145). Hence, increased TDP-Rha amounts may be beneficial for *Pseudomonas* species due

to high usage of TDP-Rha in synthesis of cell wall components or secondary metabolites, but also for protein glycosylation.

7.2 Distribution of rhamnosylation

Conversely, in protein glycosylation, rhamnose is considered as rare sugar (59). The widespread distribution of the TDP-Rha biosynthesis genes *rmIBDAC* among bacteria, but also other activated rhamnose nucleotide precursors like UDP- β -L-rhamnose (plants, fungi and algae) or GDP- α -D-rhamnose (*Pseudomonads* and chlorovirus) (49, 146), indicate that glycosylation with rhamnose moieties are not just a curiosity.

In the past, some rhamnose containing glycans were identified in bacteria. Among them, unusual rhamnose-amino acid linkage was discovered. It was first described for an Asn residue in the S-layer Glycoprotein *Geobacillus stearothermophilus* (147). Here, the glycan consists of three rhamnose residues in a row linked to Asn. O-linked rhamnosylation was detected for Ser and Thr residues in flagellar proteins in *P. aeruginosa* and *P. syringae* (63, 64). In *P. aeruginosa*, depending on the strain (JJ692 or PAK) considered, either a single mono-rhamnose or a rhamnose moiety as a link between an oligosaccharide and the protein was shown. In *P. syringae*, the glycan is linked *via* two rhamnose residues to a Ser residue. Recent findings then showed the unexpected monorhamnosylation on an Arg residue of the elongation factor EF-P in around 10 % of all bacteria including *P. aeruginosa* and *N. meningitidis* (76, 77, 148). These few examples of proteins containing the rhamnose-amino acid linkage together with the highly distributed TDP-Rha biosynthesis raised the question whether rhamnosylation represents a general mechanism for protein modification in prokaryotes. We investigated the distribution of modification by immunodetection using a set of antibodies (chapter 3).

So far, anti-Arg^{Rha} antibodies have been developed for the detection of protein monorhamnosylation and proved to be a useful tool to discriminate unmodified and rhamnosylated (activated) EF-P (91, 92). The newly generated antibodies in chapter 3 were designed to cover other, most probable rhamnosylation sites: anti-Asn^{Rha}, anti-Ser^{Rha}, and anti-Thr^{Rha}. Monorhamnosyl amino acid containing peptides fused to bovine serum albumin were used as immunogens. The resulting purified polyclonal antibodies sensitively and specifically detect the rhamnosylated amino acids without any cross-reactivity with the unmodified amino acids. Interestingly, anti-Ser^{Rha} exclusively recognize the β -configuration of the glycopeptide indicating that Ser ^{α -Rha} is poorly immunogenic.

Having these four different antibodies at hand, we were able to detect N- and O-rhamnosylated sites. We found them in various bacterial species in cytosolic as well as in membrane proteins. The number of rhamnosylated proteins in *Corynebacterium glutamicum* and *Mycobacterium*

phlei was striking. This is especially interesting considering that the biosynthesis genes of TDP-Rha (*rmIBDAC*) were shown to be essential in *Mycobacterium* (74, 149). So far, this is explained by the linking rhamnosyl residues in the cell wall of the bacteria (between arabinogalactan and peptidoglycan). Our findings now allow for an alternative or additional explanation: There might be essential proteins whose function is directly linked to rhamnosylation as it is shown for EF-P (76). In this case, identifying these rhamnosylated proteins could contribute to the discovery of new target structures to combat tuberculosis or diphtheria.

The antibody toolbox generated here sheds light on the distribution of the so far rare and unusual protein monorhamnosylation (chapter 3). We were able to show that this post-translational modification appears to be more distributed in bacteria than previously assumed. In general, protein glycosylation is expected to be very widespread. It was even postulated that more than every second protein is a glycoprotein (23). However, a study showed that there are far more putative glycoproteins than experimentally verified (13). Today, there are around 6,800 documented glycoproteins but almost 100.000 putative targets for protein glycosylation. This indicates that there are a lot of glycoproteins which remain to be uncovered. In bacteria, glycosylation was unnoticed for a long time, furthermore, the modification seems to be much more diverse here (11). Additionally, the study of glycans in general is lagging far behind other fundamental macromolecular components like DNA. This is due to the complexity of structures, difficulties in prediction of glycosylation which are not encoded directly by the genome as it is the case for proteins sequences and simply the challenges in identification, isolation, and characterization of glycans (1). The toolbox of antibodies against mono-rhamnosylated proteins contributes to challenges in identification of novel glycosylation sites as they are a prerequisite for enrichment of modified proteins prior to mass spectrometry-based identification (150, 151).

7.3 Rhamnose containing glycans as target for therapeutics

According to the WHO, antibiotic resistance is becoming one of the biggest challenges for humanity looking at global health, food security, and development. Resistance mechanisms are promoted by misuse and overuse of antibiotics but also by broad-spectrum antibiotics as they can increase the spread and uptake of bacterial genetic elements (for example plasmids encoding antibiotic resistance genes) (152, 153). The development of microorganism-specific antibiotics is therefore important as their use can slow down the evolution and spread of antibiotic resistance (154, 155).

Bacterial glycan structures and their corresponding enzymes are an interesting target for therapeutics and antibiotics as they have distinct structures and are often linked to pathogenesis. Interestingly, rhamnose containing glycans and corresponding biosynthesis enzymes are absent in humans making them ideal targets for therapeutics and antibiotics (49, 67). As shown in chapter 3, many monorhamnosylated glycoproteins are unknown today. This discovery lays the basis for further examination of these glycoproteins and their corresponding enzymes. Once these have been identified, they can be used as targets for therapeutics and antibiotics. During the last years, promising studies showed different ways to target rhamnose containing macromolecules or the biosynthesis of TDP-Rha. An interesting study on sugar binding lectins that recognize specific rhamnose structures (rhamnose binding proteins (RBPs)), showed antimicrobial activity of these substrates (156, 157). Naturally, these RBPs are found in various fish eggs (158). A recombinant version of a RBP from horseshoe crab is able to bind to rhamnose-containing components in biofilm and inhibits the biofilm formation and growth of *P. aeruginosa*. They postulate that the recombinant RBP is a prospective anti-biofilm agent which links glycan-recognition to novel anti-biofilm strategies (156).

Not only rhamnosylated proteins can be target of therapeutics, but also the TDP-Rha biosynthesis proteins. A screening for TDP-Rha biosynthesis genes identified 5-(4-chlorophenyl)-2-furoic acid (Ri03) as a potential agent which predominantly inhibits the biosynthesis through interference with RmlB function (159). Ri03 inhibited growth in Gram-positive bacteria containing rhamnose cell wall polysaccharides such as *Streptococcus* and *Mycobacteria*. Further development of Ri03 could lead to a new class of antibiotics that targets dTDP-rhamnose biosynthesis in pathogenic bacteria.

7.4 Donor substrate promiscuity of EarP

The EarP rhamnosylation reaction was intensively studied during the last years. EarP was identified as GT-B fold GT which catalyses an inverting reaction resulting in α -rhamnosyl-arginine (76, 77, 91-93, 148, 160). Typically for GT-B fold Leloir-type GTs, the binding site for the donor substrate TDP-Rha was localized in a highly conserved pocket in the protein C-domain (91, 93, 94, 161). Crystal structure analyses of *P. putida* EarP indicate that the TDP moiety rather than the rhamnose moiety contacts the protein (all amino acid in the following referred to EarP_{Ppu}) (91). Three Phenylalanine (Phe) residues (Phe191, Phe252, and Phe258) form an aromatic pocket surrounding the thymine ring (Fig. 3B). Arg271, tyrosin (Tyr)193, glutamic acid (Glu)273, aspartic acid (Asp)274, and glutamine (Gln)255 recognize the diphosphate and the sugar ring of the nucleotide by forming hydrogen bonds. The rhamnose residue is surface exposed obviously without any contact to EarP (91). This fact indicates that there is flexibility to be expected in the binding of substrates, at least on the sugar moiety. Nevertheless, studies in *N. meningitidis* (94) and *P. aeruginosa* (93) postulate that the sugar moiety does form bonds with EarP. These crystal structure analyses indicate that a certain promiscuity of the glycosyltransferase is conceivable, however, this was not investigated so far.

Our inhibitor assay could show that single TDP but also UDP are able to bind EarP (chapter 5 Fig. 5). Interestingly, the third pyrimidine derivate CDP is unable to bind EarP. In TDP and apparently also UDP, the O4 is involved in hydrogen-bonding (93, 94). CDP harbours an amino group instead of O4, indicating that the amino group is unfavourable for the binding. Further binding to the purine bases adenine or guanine is rather unlikely as the size of the aromatic pocket surrounding the thymidine ring is too small (93, 94).

The phosphate group seems to play a crucial role, as the precursors thymine and thymidine were not able to inhibit binding of TDP-Rha to EarP. Both phosphate groups of TDP hydrogen-bond with the main chain nitrogen atoms of a number of amino acids (91). Interestingly, both TMP and TTP were able to bind to EarP and compete with TDP-Rha albeit exhibiting high K_i values. The α -phosphate group seems to be sufficient whereby the second phosphate strengthens the binding. A triphosphate group strongly weakened the binding compared to TDP, in case of UTP no binding occurred even at high concentrations. The binding pocket is closed by the bulky side chain of Tyr193 (91), this may lead to steric clashes in case of a triphosphate.

The inhibitor assay testing the inhibitory potential of the sugar moiety could confirm what was expected from the studies of *P. putida* EarP (91). A single rhamnose was not able to compete

with TDP-Rha for EarP binding indicating that the binding to EarP is regulated by the nucleotide- and phosphate- rather than the sugar-moiety of the donor substrate. However, as UDP-glucose did not inhibit the rhamnosylation reaction with TDP-Rha, we conclude that the sugar moiety is a steric prerequisite for recognition and binding to EarP.

Taken together, our experiments could confirm relevant amino acid residues proposed by the crystal structure in the binding pocket of EarP. The determined substrate promiscuity is of high interest from the view of a potential antimicrobial strategy as binding without transfer of substrates can inhibit the glycosylation reaction by EarP (162). Suppressed rhamnosylation can lead to loss of pathogenicity as this modification was shown to be important for pathogenicity of human pathogens (76, 77, 148).

7.5 Acceptor substrate promiscuity of EarP

On the other hand, the acceptor substrate specificity of EarP is poorly characterized. The only known substrate of the rhamnosyltransferase is a single Arg (Arg32) residue of EF-P, hence, the glycosylation site possesses unique recognition features. EarP does not show high sequence homologies to neither the Arg glycosylating GT NleB nor to any other structurally characterized GT. Therefore, it is difficult to predict the prerequisites for acceptor substrate binding.

So far, it is known that EarP recognizes and binds the KOW-like EF-P N-domain I (91, 94). The GT interacts with the acceptor mainly through a highly conserved region of the N-terminal domain (94). Crystal studies indicate that the molecular recognition is both structure and sequence specific as the binding is mediated by several specific side chains (94). Interestingly, these interacting side chains are not conserved (93). Further, it was shown that EarP rhamnosylates the non-cognate EF-P orthologue from *E. coli* when Lys was substituted by Arg (163). These facts indicate that the acceptor recognition is rather not sequence specific. This assumption was further supported by an *in vitro* rhamnosylation experiment (chapter 4, Fig. S1). Here, we could show that EarP does not rhamnosylate a linear peptide fragment comprising the amino acids of the Arg32-containing loop of EF-P (acceptor loop). In addition, mutational analyses revealed that Arg is rhamnosylated at any position of the acceptor loop *in vivo* (chapter 5, Fig. 1). Both experiments underline the importance of the loop structure which seems to optimally position the Arg residue for rhamnosylation in the active site of EarP.

However, data from crystal structures indicate that EarP also interacts with other residues besides those next to Arg32 at the loop (94). These residues mainly belong to two neighbouring

β -strands leading to an experiment where the acceptor loop plus the surrounding β -strands of EF-P were fused to *E. coli* EF-P lacking domain III (chapter 5, Fig. 2). Stepwise truncation of both β -strands could narrow down the important region for acceptor recognition to the acceptor loop plus two amino acids of both β -strands. The two residues of both β -strands may be sufficient to build the β -sheet hairpin structure including the loop. The development of a cyclic peptide, which mimics the β -hairpin secondary structure *via* introduction of a L-proline/D-proline motif, provided further insights into the importance of the secondary structure (chapter 4, Fig. 2). The cyclic peptide, which contained the acceptor loop together with one residue of each β -strand, was successfully rhamnosylated *in vitro* with a conversion rate of 85 %. The here shortened β -strand indicates that valine (Val)³⁷ and Phe²⁷ play a minor role in recognition but may be crucial for forming the β -sheet hairpin structure. In the cyclic peptide, the loop structure is alternatively built by the cyclisation strategy. Both experiments, (chapter 4 Fig. 2, chapter 5 Fig. 2) clearly showed that the acceptor loop alone without neighbouring amino acid residues from the β -strand is not an acceptor substrate for EarP. On the basis of this data, we identified the minimal recognition motif (Fig. 6) that is sufficient for arginine rhamnosylation by EarP. The surrounding EF-P context seems to play a minor role in substrate recognition by EarP. This assumption is further supported by the successful rhamnosylation of the EF-P acceptor loop fused to the terminus of mCherry protein (chapter 5, Fig. 3).

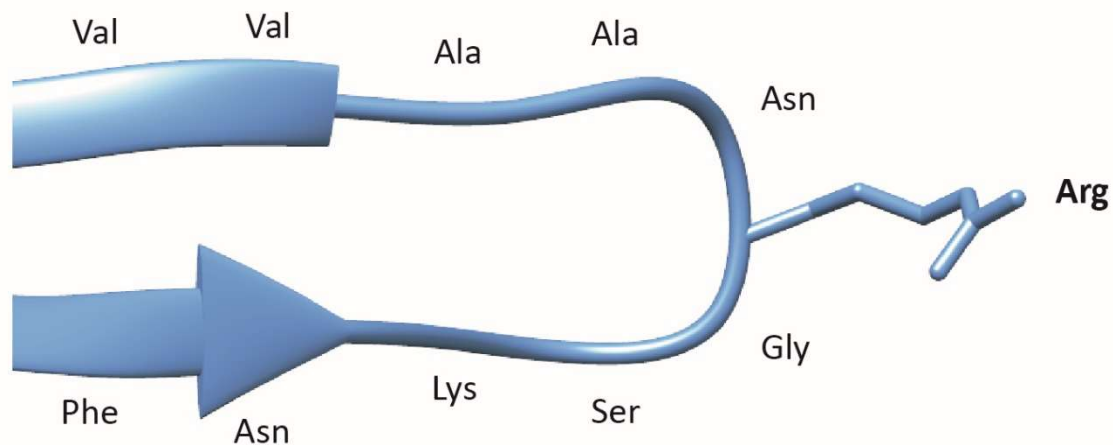


Figure 1: **Minimal recognition motif of EarP.** The glycosyltransferase EarP recognizes a structural motif allowing for sequence variation. The motif consists of two β -strands (two amino acids on each strand: Val/Val and Asn/Phe) forming a loop structure with Arg on its tip. Illustrations were generated with UCSF Chimera (19).

Recognition of structural elements for glycosylation was reported a few years ago for GT AIDA-associated heptosyltransferase (AAH) (164). The substrate of AAH, adhesin involved in diffuse adherence (AIDA-I), is glycosylated in the extracellular domain which comprises imperfect repeats of a consensus sequence forming a “short β -strand–short acceptor loop–short β -strand”. Similar to EarP, AAH recognizes a structural motif, not a specific sequence. As the

glycosylation of AIDA is O-linked (165), the recognition of a structural element does not seem to be restricted to a specific type of glycosylation.

Mutational analysis of this cyclic peptide further revealed that the alanine (Ala)³⁴ plays an important role in shaping the β -hairpin (chapter 4, Fig. 2). Single deletion of this residue prevents rhamnosylation of the cyclic peptide whereas substitution with glycine (Gly) lead to 41 % conversion. Together with further mutational analysis (chapter 4) and the arginine-walking (chapter 5, Fig. 1) we can assume that EarP partially allows for acceptor sequence promiscuity as long as the loop structure is maintained. Interestingly, upon EarP overexpression in *E. coli* wildtype, several rhamnosylated high-molecular proteins are detectable which are dependent on the presence of TDP-Rha (chapter 5, Fig. 4). These off-targets could be rhamnosylated either at a loop region similar to the minimal recognition motif or contain some variations. Identifying these off-targets and their target sites could further contribute to our knowledge of acceptor promiscuity of EarP.

7.6 EarP as synthetic glycosyltransferase

Taken together, our results (chapter 4 and 5) provide further insights into donor and acceptor substrate binding and promiscuity of the GT EarP. On the one hand, we were able to show for the first time that EarP allows for donor substrate promiscuity. On the other hand, we determined the minimal recognition motif which is sufficient for acceptor recognition. Again, EarP showed promiscuity as the GT tolerates amino acid sequence variation in the acceptor.

These findings lay the basis for application of the GT EarP in synthetic biology and medicine. This is an important field as the properties and functionality of many therapeutics (70 % of therapeutic proteins (166)) is dependent on regio and stereo specific glycosylation (167-170). In addition, recombinant therapeutic glycoproteins must resemble human glycosylation to avoid immune response (171). The application of the bacterial glycosylation machinery has opened a new route to achieve these desired glycosylations (10, 172, 173). The major advantages are the more simple and cheap culturing conditions compared to eukaryotic cell cultures (174). However, as the bacterial glycosylation systems differ from eukaryotic, engineering of prokaryotic glycosylation is necessary. EarP in particular seems to be an interesting target for engineering as it is a cytoplasmatic GT facilitating the availability of donor and acceptor substrates compared to recent approaches with the periplasmatic PglB (173). Additionally, the fact that EarP is a mono-GT could be useful in bottom-up constructions. There are different approaches ranging from rational design to domain swapping and combined GTs approaches.

Based on crystal structure analysis, rational design aims at the engineering of a natural GT by different mutations (175). The basis for EarP rational mutagenesis was laid by different crystal structures including donor and acceptor binding states (91, 93, 94). We demonstrated that EarP possesses a natural donor substrate promiscuity (for example UDP) (chapter 5) making the GT an ideal starting point for further engineering (Fig. 7A top). Expanding the promiscuity of EarP by rational mutagenesis of residues in the TDP-Rha binding pocket could allow for altering or switching the donor substrate specificity. The most interesting substrate for a synthetic GT so far would be UDP-GlcNac as GlcNac is the ubiquitous initial linkage sugar in eukaryotic N-linked glycosylation. Originating from the GlcNac linkage, different glycans can be built in bottom-up approaches. EarP seems to be a promising candidate to transfer UDP-GlcNac as it naturally binds UDP (chapter 5, Fig. 5). This would be a more direct way to achieve the precursor GlcNac linkage instead of glycan remodeling by trimming and transglycosylation (176). A further step including the switch from Arg- to Asn-linked glycosylation by EarP would lead to a synthetic GT exactly establishing the eukaryotic N-linked glycosylation precursor. It is also thinkable to reach this switch in the acceptor amino acid or the donor substrate

specificity *via* domain swapping. Here, chimeric glycosyltransferases are generated by combination of the N- and C-terminal domains of characterized GTs (177). In the past, a switch of substrate specificity was successfully developed using domain swapping (177, 178).

The possibility of integrating a minimal recognition motif into other proteins in order to allow rhamnosylation by EarP is another interesting prospect arising from this study (Fig. 7A bottom). Also, already existing loop structures could be adapted to be recognized by EarP. The arginine walking experiment (chapter 5, Fig. 1) showed that there might be a huge range of loop structures that are recognized by EarP. Implementing the minimal recognition motif could be interesting for different therapeutics and vaccines whose functionality is influenced or dependent on glycosylation.

One of the most prominent glycosylated therapeutics are recombinant monoclonal antibodies. Naturally, antibodies are N-glycosylated in the constant domain (Fc) which is postulated to be strongly involved in antibody stability, half-life, secretion, immunogenicity, and function (179). Variation of the attached glycans results in different glycoforms which can differ in biological efficiency (180). Hence, engineering of the Fc glycosylation is nowadays an important part of generating safe and efficient recombinant monoclonal antibodies (181).

Glycosylation also plays an important role in the development of vaccines as it can influence the immunogenicity of an antigen. α -Gal (Gal-R(1,3)-Gal- β (1,4)-GlcNAc/Glc) is one of the most known carbohydrate antigens (182). The epitope is absent in humans whereas they naturally produce a huge amount of the anti-Gal antibody which specifically interacts with α -Gal epitopes. This mechanism is used to highly increase immunogenicity of tumour vaccines for example as the anti-Gal/ α -Gal complex is efficiently presented to antigen presenting cells (183). Interestingly, even higher levels of antibodies against mono-L-rhamnose were also found in humans (184-186). This finding suggests that mono-rhamnosylated vaccines could serve as an alternative to the α -Gal (186). By implementing the minimal recognition motif of the GT EarP to vaccines, we could provide a simple tool to efficiently enhance the immunogenicity of various vaccines.

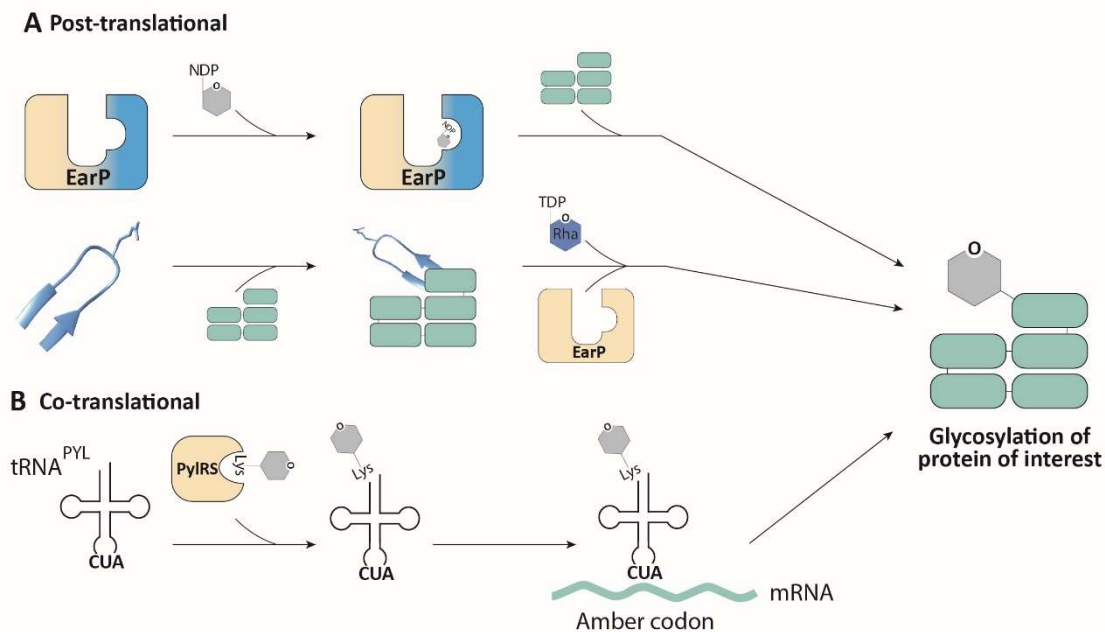


Figure 2: **Post- and co-translational approaches to introduce protein glycosylation to non-target proteins.** **A)** Post-translational glycosylation. Top: Rational design or domain swapping based on EarP (yellow/blue) lead to modification of non-target proteins (green) with desired activated nucleotide sugars (NDP-sugar (grey)). Bottom: Introducing of the structural recognition motif (blue loop structure) of EarP to proteins of interest leads to non-target rhamnosylation by EarP (yellow). **B)** Co-translational glycosylation. The aminoacyl-tRNA synthetase PylRS charges its cognate tRNA_{CUA} with fructoselysine (Lys-sugar (grey)) and integrates the sugar-amino acid at the desired position (amber codon) during translation of the mRNA (green line).

7.7 Co-translational glycosylation using ϵ -Frk as non-canonical amino acid

Engineering natural GTs like EarP is one possibility to achieve desired glycosylation in proteins of interest. Besides post-translational modification, co-translational integration using systems like amber suppression are useful to achieve modification of distinct amino acids (187). Using amber suppression, it is possible to incorporate non-canonical amino acids (ncAAs) site specific into proteins thereby expanding the genetic code (188). Therefore, a stop codon (amber codon UAG) is reassigned: the amber codon is converted into an elongation codon for a ncAAs through the usage and engineering of orthogonal aminoacyl-tRNA synthetase/tRNA_{CUA} pairs. As a result, the ncAA is incorporated in the growing polypeptide after mutating the desired site to an amber stop codon. The pyrrolysine tRNA synthetase-tRNA pair (PylRS)/tRNA_{CUA} from the archaeum *Methanosarcina mazei* is often used as it is one of the most promiscuous pairs found in nature (189, 190).

With this method, a large range of ncAAs including several PTMs was successfully integrated in proteins, whereby most of the ncAAs transferred by PylRS are lysine derivatives (191, 192). Different approaches integrating glycosylation co-translationally were established in cell-free

systems (193, 194). However, a major issue of this technique *in vivo* is the uptake and stability of the desired sugar-amino acid. Therefore, we searched for a natural occurring sugar-amino acid which can be transferred inside the cell without concomitantly processing (for example phosphorylation (195)). A promising candidate is ϵ -Frk, a sugar-amino acid formed *via* the Maillard reaction of glucose and lysine. *E. coli* was shown to be able to grow on ϵ -Frk (196) indicating that the bacterium encodes for an uptake and degradation mechanism. Here, the uptake, presumably by the permease FrlA, is separated from the degradation catalyzed by the kinase FrlD and the deglycase FrlB (131) making ϵ -Frk an interesting target for genetic code expansion. However, nothing is known about the regulation of the operon coding for the uptake and degradation of the sugar-amino acid. To be able to use ϵ -Frk for genetic code expansion, we investigated the ϵ -Frk catabolism regulation as a first step (chapter 6).

In *E. coli*, the operon *frlABCD*, responsible for ϵ -Frk uptake and catabolism, is tightly controlled by positive and negative regulation. Promotor activation is achieved by the global transcription factor catabolite activator protein (CAP) and the sigma factor σ_{32} (RpoH) (chapter 6, Fig. 5). In the absence of ϵ -Frk, the transcriptional regulator FrlR, a member of the GntR/HutC family, recognized a consensus sequence upstream of the start codon (chapter 6, Fig. 7). Upon uptake of ϵ -Frk, FrlR binds the phosphorylated FrlD kinase product fructoselysine-6-phosphate leading to a structural rearrangement of the FrlR dimer (chapter 6, Fig. 8/9). Thereby, the DNA binding is weakened and transcription of the *frlABCD* operon is possible. This fine-tuned regulation allows for an efficient use of unusual substrates within the gut environment.

Knowing the regulation of the ϵ -Frk catabolism, it is left over to investigate the uptake of the sugar-amino acid. The putative permease *frlA* might be crucial for increasing the amount of ϵ -Frk within the cell. High concentrations of ϵ -Frk in the cell would improve the efficiency of genetic expansion approaches and lower the costs by efficient substrate uptake. Once the whole ϵ -Frk catabolism is characterized, it is worth investigating whether ϵ -Frk can be co-translationally incorporated. This would finally enable the co-translational integration of glycosylation *in vivo* (Fib. 7B).

7.8 Outlook

In this thesis, we showed that monorhamnosylation is a widely distributed protein modification in bacteria. Considering the importance of this PTM in regard of pathogenicity, rhamnosylated proteins, their corresponding GT, and nucleotide sugar biosynthesis proteins could be target for novel anti-microbial strategies. Moreover, the substrate promiscuity of the mono-GT EarP was determined, giving the prerequisite of synthetic protein glycosylation with EarP. Engineering of this GT by rational design or domain swapping could lead to a synthetic GT able to achieve site-specific mono-glycosylation matching the needs of glycosylated therapeutics. An alternative approach to build synthetic glycosylation is the co-translational modification using amber suppression. Investigation of ϵ -Frk as non-canonical amino acid could lead to successful *in vivo* glycosylation. These ideas of synthetic glycosylation contribute to the emerging field of synthetic biology and could have major impact on future medical applications.

References for chapters 1 and 7

1. A. Varki, S. Kornfeld, in *Essentials of Glycobiology*, A. Varki et al., Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2015), pp. 1-18.
2. G. W. Hart, R. L. Schnaar, R. S. Haltiwanger, A Quarter Century of Glycobiology. *Glycobiology* **25**, 1321-1322 (2015).
3. A. Varki, Biological roles of glycans. *Glycobiology* **27**, 3-49 (2017).
4. T. J. Tolbert, C.-H. Wong, in *Encyclopedia of Biological Chemistry*. (2013), pp. 362-367.
5. P. H. Seeberger, in *Essentials of Glycobiology*, A. Varki et al., Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2015), pp. 19-30.
6. H. Schachter, H. H. Freeze, Glycosylation diseases: quo vadis? *Biochim Biophys Acta* **1792**, 925-930 (2009).
7. J. D. Marth, A unified vision of the building blocks of life. *Nat Cell Biol* **10**, 1015-1016 (2008).
8. A. Varki, Evolutionary forces shaping the Golgi glycosylation machinery: why cell surface glycans are universal to living cells. *Cold Spring Harb Perspect Biol* **3**, (2011).
9. C. M. Harding, M. F. Feldman, Glycoengineering bioconjugate vaccines, therapeutics, and diagnostics in *E. coli*. *Glycobiology* **29**, 519-529 (2019).
10. W. Kightlinger, K. F. Warfel, M. P. DeLisa, M. C. Jewett, Synthetic Glycobiology: Parts, Systems, and Applications. *ACS Synth Biol* **9**, 1534-1562 (2020).
11. M. Abu-Qarn, J. Eichler, N. Sharon, Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Curr Opin Struct Biol* **18**, 544-550 (2008).
12. J. L. Baker, E. Çelik, M. P. DeLisa, Expanding the glycoengineering toolbox: the rise of bacterial N-linked protein glycosylation. *Trends Biotechnol* **31**, 313-323 (2013).
13. G. A. Houry, R. C. Baliban, C. A. Floudas, Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci Rep* **1**, (2011).
14. C. T. Walsh, S. Garneau-Tsodikova, G. J. Gatto, Jr., Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **44**, 7342-7372 (2005).
15. M. M. Müller, Post-Translational Modifications of Protein Backbones: Unique Functions, Mechanisms, and Challenges. *Biochemistry* **57**, 177-185 (2018).
16. H. Nothaft, C. M. Szymanski, Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* **8**, 765-778 (2010).
17. L. L. Lairson, B. Henrissat, G. J. Davies, S. G. Withers, Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **77**, 521-555 (2008).
18. P. M. Coutinho, E. Deleury, G. J. Davies, B. Henrissat, An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol* **328**, 307-317 (2003).
19. E. F. Pettersen et al., UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
20. V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**, D490-495 (2014).
21. A. Neuberger, Carbohydrates in protein: The carbohydrate component of crystalline egg albumin. *Biochem J* **32**, 1435-1451 (1938).
22. R. H. Nuenke, L. W. Cunningham, Preparation and structural studies of ovalbumin glycopeptides. *J Biol Chem* **236**, 2452-2460 (1961).
23. R. Apweiler, H. Hermjakob, N. Sharon, On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1473**, 4-8 (1999).
24. J. Lombard, The multiple evolutionary origins of the eukaryotic N-glycosylation pathway. *Biol Direct* **11**, 36 (2016).
25. S. P., T. N., A. M., in *Essentials of Glycobiology*. 3rd edition. (Cold Spring Harbor Laboratory Press, 2017).

26. G. Niu *et al.*, Comparative and evolutionary analyses of the divergence of plant oligosaccharyltransferase STT3 isoforms. *FEBS Open Bio* **10**, 468-483 (2020).
27. D. N. Hebert, M. Molinari, Flagging and docking: dual roles for N-glycans in protein quality control and cellular proteostasis. *Trends Biochem Sci* **37**, 404-410 (2012).
28. J. J. Caramelo, A. J. Parodi, A sweet code for glycoprotein folding. *FEBS Lett* **589**, 3379-3387 (2015).
29. A. Corfield, Eukaryotic protein glycosylation: a primer for histochemists and cell biologists. *Histochem Cell Biol* **147**, 119-147 (2017).
30. A. Dell, A. Galadari, F. Sastre, P. Hitchen, Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *Int J Microbiol* **2010**, 148178 (2010).
31. J. H. Prestegard, J. Liu, G. Widmalm, in *Essentials of Glycobiology. 3rd edition.* (Cold Spring Harbor Laboratory Press, 2017).
32. I. Brockhausen, P. Stanley, in *Essentials of Glycobiology*, A. Varki *et al.*, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2015), pp. 113-123.
33. C. M. Szymanski, B. W. Wren, Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Microbiol* **3**, 225-237 (2005).
34. U. B. Sleytr, K. J. Thorne, Chemical characterization of the regularly arranged surface layers of *Clostridium thermosaccharolyticum* and *Clostridium thermohydrosulfuricum*. *J Bacteriol* **126**, 377-383 (1976).
35. U. B. Sleytr, Heterologous reattachment of regular arrays of glycoproteins on bacterial surfaces. *Nature* **257**, 400-402 (1975).
36. F. E. Aas, A. Vik, J. Vedde, M. Koomey, W. Egge-Jacobsen, *Neisseria gonorrhoeae* O-linked pilin glycosylation: functional analyses define both the biosynthetic pathway and glycan structure. *Mol Microbiol* **65**, 607-624 (2007).
37. P. Castric, *pilO*, a gene required for glycosylation of *Pseudomonas aeruginosa* 1244 pilin. *Microbiology* **141** (Pt 5), 1247-1254 (1995).
38. M. Virji *et al.*, Pilus-facilitated adherence of *Neisseria meningitidis* to human epithelial and endothelial cells: modulation of adherence phenotype occurs concurrently with changes in primary amino acid sequence and the glycosylation status of pilin. *Mol Microbiol* **10**, 1013-1028 (1993).
39. E. Stimson *et al.*, Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose. *Mol Microbiol* **17**, 1201-1214 (1995).
40. H. Li *et al.*, Understanding protein glycosylation pathways in bacteria. *Future Microbiol* **12**, 59-72 (2017).
41. S. K. Ghosh *et al.*, Pathogenic consequences of *Neisseria gonorrhoeae* pilin glycan variation. *Microbes Infect* **6**, 693-701 (2004).
42. N. M. Young *et al.*, Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. *J Biol Chem* **277**, 42530-42539 (2002).
43. D. Linton *et al.*, Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway. *Mol Microbiol* **55**, 1695-1703 (2005).
44. J. Gross *et al.*, The *Haemophilus influenzae* HMW1 adhesin is a glycoprotein with an unusual N-linked carbohydrate modification. *J Biol Chem* **283**, 26010-26015 (2008).
45. S. Grass *et al.*, The *Haemophilus influenzae* HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. *Mol Microbiol* **48**, 737-751 (2003).
46. S. Grass, C. F. Lichti, R. R. Townsend, J. Gross, J. W. St Geme, 3rd, The *Haemophilus influenzae* HMW1C protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin. *PLoS Pathog* **6**, e1000919 (2010).
47. C. J. Thibodeaux, C. E. Melancon, 3rd, H. W. Liu, Natural-product sugar biosynthesis and enzymatic glycodiversification. *Angew Chem Int Ed Engl* **47**, 9814-9859 (2008).
48. C. Schaffer, P. Messner, Surface-layer glycoproteins: an example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology. *Glycobiology* **14**, 31r-42r (2004).

49. M. Maki, R. Renkonen, Biosynthesis of 6-deoxyhexose glycans in bacteria. *Glycobiology* **14**, 1r-15r (2004).
50. K. Marino, J. Bones, J. J. Kattla, P. M. Rudd, A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol* **6**, 713-723 (2010).
51. F. Schwarz, M. Aebi, Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struct Biol* **21**, 576-582 (2011).
52. J. Eichler, M. Koomey, Sweet New Roles for Protein Glycosylation in Prokaryotes. *Trends Microbiol* **25**, 662-672 (2017).
53. D. Calo, L. Kaminski, J. Eichler, Protein glycosylation in Archaea: sweet and extreme. *Glycobiology* **20**, 1065-1076 (2010).
54. K. Ohtsubo, J. D. Marth, Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855-867 (2006).
55. D. Ribet, P. Cossart, Post-translational modifications in host cells during bacterial infection. *FEBS Lett* **584**, 2748-2758 (2010).
56. M. Schirm *et al.*, Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Mol Microbiol* **48**, 1579-1592 (2003).
57. S. M. Twine *et al.*, Motility and flagellar glycosylation in *Clostridium difficile*. *J Bacteriol* **191**, 7050-7062 (2009).
58. S. M. Logan, Flagellar glycosylation - a new component of the motility repertoire? *Microbiology* **152**, 1249-1262 (2006).
59. P. Lafite, R. Daniellou, Rare and unusual glycosylation of peptides and proteins. *Nat Prod Rep* **29**, 729-738 (2012).
60. P. Drigues *et al.*, Comparative studies of lipopolysaccharide and exopolysaccharide from a virulent strain of *Pseudomonas solanacearum* and from three avirulent mutants. *J Bacteriol* **162**, 504-509 (1985).
61. A. S. Kumar, K. Mody, B. Jha, Bacterial exopolysaccharides--a perception. *J Basic Microbiol* **47**, 103-117 (2007).
62. M. Y. Mistou, I. C. Sutcliffe, N. M. van Sorge, Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in Gram-positive bacteria. *FEMS Microbiol Rev* **40**, 464-479 (2016).
63. M. Schirm *et al.*, Structural and genetic characterization of glycosylation of type a flagellin in *Pseudomonas aeruginosa*. *J Bacteriol* **186**, 2523-2531 (2004).
64. K. Takeuchi *et al.*, Flagellin glycans from two pathovars of *Pseudomonas syringae* contain rhamnose in D and L configurations in different ratios and modified 4-amino-4,6-dideoxyglucose. *J Bacteriol* **189**, 6945-6956 (2007).
65. J. Le Roy, B. Huss, A. Creach, S. Hawkins, G. Neutelings, Glycosylation is a major regulator of phenylpropanoid availability and biological activity in plants. *Front Plant Sci* **7**, 735 (2016).
66. A. M. Saffer, V. F. Irish, Flavonol rhamnosylation indirectly modifies the cell wall defects of rhamnose biosynthesis mutants by altering rhamnose flux. *Plant J* **94**, 649-660 (2018).
67. M. F. Giraud, J. H. Naismith, The rhamnose pathway. *Curr Opin Struct Biol* **10**, 687-696 (2000).
68. Y. Yamashita *et al.*, Recombination between gtfB and gtfC is required for survival of a dTDP-rhamnose synthesis-deficient mutant of *Streptococcus mutans* in the presence of sucrose. *Infect Immun* **67**, 3693-3697 (1999).
69. R. Rahim, L. L. Burrows, M. A. Monteiro, M. B. Perry, J. S. Lam, Involvement of the rml locus in core oligosaccharide and O polysaccharide assembly in *Pseudomonas aeruginosa*. *Microbiology* **146** (Pt 11), 2803-2814 (2000).
70. Y. Xu, K. V. Singh, X. Qin, B. E. Murray, G. M. Weinstock, Analysis of a gene cluster of *Enterococcus faecalis* involved in polysaccharide biosynthesis. *Infect Immun* **68**, 815-823 (2000).
71. S. L. van der Beek *et al.*, Streptococcal dTDP-L-rhamnose biosynthesis enzymes: functional characterization and lead compound identification. *Mol Microbiol* **111**, 951-964 (2019).

72. A. P. Bischer, C. J. Kovacs, R. C. Faustoferri, R. G. Quivey, Jr., Disruption of I-Rhamnose Biosynthesis Results in Severe Growth Defects in *Streptococcus mutans*. *J Bacteriol* **202**, (2020).
73. S. L. Chiang, J. J. Mekalanos, rfb mutations in *Vibrio cholerae* do not affect surface production of toxin-coregulated pili but still inhibit intestinal colonization. *Infect Immun* **67**, 976-980 (1999).
74. Y. Ma, F. Pan, M. McNeil, Formation of dTDP-rhamnose is essential for growth of mycobacteria. *Journal of bacteriology* **184**, 3392-3395 (2002).
75. Y. Ma *et al.*, Drug targeting *Mycobacterium tuberculosis* cell wall synthesis: genetics of dTDP-rhamnose synthetic enzymes and development of a microtiter plate-based screen for inhibitors of conversion of dTDP-glucose to dTDP-rhamnose. *Antimicrobial agents and chemotherapy* **45**, 1407-1416 (2001).
76. J. Lassak *et al.*, Arginine-rhamnosylation as new strategy to activate translation elongation factor P. *Nature chemical biology* **11**, 266-270 (2015).
77. A. Rajkovic *et al.*, Cyclic Rhamnosylated Elongation Factor P Establishes Antibiotic Resistance in *Pseudomonas aeruginosa*. *mBio* **6**, e00823 (2015).
78. S. Ude *et al.*, Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science (New York, N.Y.)* **339**, 82-85 (2013).
79. M. Bailly, V. de Crecy-Lagard, Predicting the pathway involved in post-translational modification of elongation factor P in a subset of bacterial species. *Biol Direct* **5**, 3 (2010).
80. N. C. Kyrpides, C. R. Woese, Universally conserved translation initiation factors. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 224-228 (1998).
81. A. L. Starosta *et al.*, Translational stalling at polyproline stretches is modulated by the sequence context upstream of the stall site. *Nucleic acids research* **42**, 10711-10719 (2014).
82. L. Peil *et al.*, Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 15265-15270 (2013).
83. T. Yanagisawa, T. Sumida, R. Ishii, C. Takemoto, S. Yokoyama, A paralog of lysyl-tRNA synthetase aminoacylates a conserved lysine residue in translation elongation factor P. *Nature structural & molecular biology* **17**, 1136-1143 (2010).
84. J. Lassak, F. Koller, R. Krafczyk, W. Volkwein, Exceptionally versatile - arginine in bacterial post-translational protein modifications. *Biological chemistry* **400**, 1397-1427 (2019).
85. D. G. Singh *et al.*, beta-Glucosylarginine: a new glucose-protein bond in a self-glucosylating protein from sweet corn. *FEBS letters* **376**, 61-64 (1995).
86. T. Konishi *et al.*, An arginyl residue in rice UDP-arabinopyranose mutase is required for catalytic activity and autoglycosylation. *Carbohydrate research* **345**, 787-791 (2010).
87. S. Li *et al.*, Pathogen blocks host death receptor signalling by arginine GlcNAcylation of death domains. *Nature* **501**, 242-246 (2013).
88. J. S. Pearson *et al.*, A type III effector antagonizes death receptor signalling during bacterial gut infection. *Nature* **501**, 247-251 (2013).
89. J. B. Park *et al.*, Structural basis for arginine glycosylation of host substrates by bacterial effector proteins. *Nature communications* **9**, 4283 (2018).
90. D. Esposito *et al.*, Structural basis for the glycosyltransferase activity of the Salmonella effector SseK3. *The Journal of biological chemistry* **293**, 5064-5078 (2018).
91. R. Krafczyk *et al.*, Structural Basis for EarP-Mediated Arginine Glycosylation of Translation Elongation Factor EF-P. *mBio* **8**, (2017).
92. X. Li *et al.*, Resolving the alpha-glycosidic linkage of arginine-rhamnosylated translation elongation factor P triggers generation of the first Arg(Rha) specific antibody. *Chemical science* **7**, 6995-7001 (2016).

93. C. He *et al.*, Complex structure of *Pseudomonas aeruginosa* arginine rhamnosyltransferase EarP with its acceptor elongation factor P. *Journal of bacteriology* **201**, (2019).
94. T. Sengoku *et al.*, Structural basis of protein arginine rhamnosylation by glycosyltransferase EarP. *Nature chemical biology* **14**, 368-374 (2018).
95. V. M. Monnier, A. Cerami, Nonenzymatic browning in vivo: possible process for aging of long-lived proteins. *Science (New York, N.Y.)* **211**, 491-493 (1981).
96. T. M., *Glycoscience*. Glycation of Proteins (Springer, Tokyo, 2014).
97. R. Bucala, Z. Makita, T. Koschinsky, A. Cerami, H. Vlassara, Lipid advanced glycosylation: pathway for lipid oxidation in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 6434-6438 (1993).
98. T. Miyazawa, K. Nakagawa, S. Shimasaki, R. Nagai, Lipid glycation and protein glycation in diabetes and atherosclerosis. *Amino acids* **42**, 1163-1170 (2012).
99. A. T. Lee, A. Cerami, Nonenzymatic glycosylation of DNA by reducing sugars. *Progress in clinical and biological research* **304**, 291-299 (1989).
100. M. Fournet, F. Bonté, A. Desmoulière, Glycation Damage: A Possible Hub for Major Pathophysiological Disorders and Aging. *Aging and disease* **9**, 880-900 (2018).
101. M. Hellwig, T. Henle, Baking, ageing, diabetes: a short history of the Maillard reaction. *Angewandte Chemie (International ed. in English)* **53**, 10316-10329 (2014).
102. L. C. Maillard, The action of amino acids on sugar; The formation of melanoidin by a methodic route. *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences* **154**, 66 - 68 (1912).
103. L. C. Maillard, General reaction of amino acids on sugars: its biological consequences. *Comptes Rendus des Seances de la Societe de Biologie et de Ses Filiales* **72**, 599-601 (1912).
104. J. E. Hodge, Dehydrated Foods, Chemistry of Browning Reactions in Model Systems. *Journal of agricultural and food chemistry* **15**, 928-943 (1953).
105. M. Amadori, Products of condensation between glucose and p-phenetidine I. *Atti Accad Lincei* **2**, 337-342 (1925).
106. K. Heyns, H. Noack, Die Umsetzung Von D-Fructose mit L-Lysin und L-Arginin und deren Beziehung zu nichtenzymatischen Bräunungsreaktionen. *EurJIC* **95**, 720-727 (1962).
107. P. Ulrich, A. Cerami, Protein glycation, diabetes, and aging. *Recent progress in hormone research* **56**, 1-21 (2001).
108. C. Henning, M. A. Glomb, Pathways of the Maillard reaction under physiological conditions. *Glycoconjugate journal* **33**, 499-512 (2016).
109. V. P. Singh, A. Bali, N. Singh, A. S. Jaggi, Advanced glycation end products and diabetic complications. *The Korean journal of physiology & pharmacology : official journal of the Korean Physiological Society and the Korean Society of Pharmacology* **18**, 1-14 (2014).
110. H. F. Bunn, P. J. Higgins, Reaction of monosaccharides with proteins: possible evolutionary significance. *Science (New York, N.Y.)* **213**, 222-224 (1981).
111. F. Franks, Physical chemistry of small carbohydrates - equilibrium solution properties. *Pure and Appl Chem* **59**, 1189-1202 (1987).
112. A. Simm *et al.*, Protein glycation - Between tissue aging and protection. *Experimental gerontology* **68**, 71-75 (2015).
113. M. N. Lund, C. A. Ray, Control of Maillard Reactions in Foods: Strategies and Chemical Mechanisms. *Journal of agricultural and food chemistry* **65**, 4537-4552 (2017).
114. H. F. Bunn *et al.*, Structural heterogeneity of human hemoglobin A due to nonenzymatic glycosylation. *The Journal of biological chemistry* **254**, 3892-3898 (1979).
115. N. Iberg, R. Flückiger, Nonenzymatic glycosylation of albumin in vivo. Identification of multiple glycosylated sites. *The Journal of biological chemistry* **261**, 13542-13545 (1986).
116. J. W. Baynes *et al.*, The Amadori product on protein: structure and reactions. *Progress in clinical and biological research* **304**, 43-67 (1989).

117. M. B. Johansen, L. Kiemer, S. Brunak, Analysis and prediction of mammalian protein glycation. *Glycobiology* **16**, 844-853 (2006).
118. H. M. Reddy *et al.*, GlyStruct: glycation prediction using structural properties of amino acid residues. *BMC bioinformatics* **19**, 547 (2019).
119. T. Koschinsky *et al.*, Orally absorbed reactive glycation products (glycotoxins): an environmental risk factor in diabetic nephropathy. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 6474-6479 (1997).
120. T. Henle, AGEs in foods: do they play a role in uremia? *Kidney international. Supplement*, S145-147 (2003).
121. R. J. Koenig *et al.*, Correlation of glucose regulation and hemoglobin A1c in diabetes mellitus. *The New England journal of medicine* **295**, 417-420 (1976).
122. H. F. Bunn, K. H. Gabbay, P. M. Gallop, The glycosylation of hemoglobin: relevance to diabetes mellitus. *Science (New York, N.Y.)* **200**, 21-27 (1978).
123. R. Bucala, H. Vlassara, A. Cerami, Advanced glycosylation endproducts. *CRC Press Inc., Boca Raton* **2**, (1992).
124. S. Rahbar, An abnormal hemoglobin in red cells of diabetics. *Clinica chimica acta; international journal of clinical chemistry* **22**, 296-298 (1968).
125. V. M. Deppe, J. Bongaerts, T. O'Connell, K. H. Maurer, F. Meinhardt, Enzymatic deglycation of Amadori products in bacteria: mechanisms, occurrence and physiological functions. *Applied microbiology and biotechnology* **90**, 399-406 (2011).
126. Z. Lin, J. Zheng, Occurrence, characteristics, and applications of fructosyl amine oxidases (amadoriases). *Applied microbiology and biotechnology* **86**, 1613-1619 (2010).
127. E. Van Schaftingen, F. Collard, E. Wiame, M. Veiga-da-Cunha, Enzymatic repair of Amadori products. *Amino acids* **42**, 1143-1150 (2012).
128. V. M. Monnier, X. Wu, Enzymatic deglycation with amadoriase enzymes from *Aspergillus sp.* as a potential strategy against the complications of diabetes and aging. *Biochemical Society transactions* **31**, 1349-1353 (2003).
129. G. Delpierre *et al.*, Identification, cloning, and heterologous expression of a mammalian fructosamine-3-kinase. *Diabetes* **49**, 1627-1634 (2000).
130. B. S. Szwegold, S. Howell, P. J. Beisswenger, Human fructosamine-3-kinase: purification, sequencing, substrate specificity, and evidence of activity in vivo. *Diabetes* **50**, 2139-2147 (2001).
131. E. Wiame, G. Delpierre, F. Collard, E. Van Schaftingen, Identification of a pathway for the utilization of the Amadori product fructoselysine in *Escherichia coli*. *The Journal of biological chemistry* **277**, 42523-42529 (2002).
132. E. Wiame, A. Duquenne, G. Delpierre, E. Van Schaftingen, Identification of enzymes acting on alpha-glycated amino acids in *Bacillus subtilis*. *FEBS letters* **577**, 469-472 (2004).
133. E. Wiame, P. Lamosa, H. Santos, E. Van Schaftingen, Identification of glucoselysine-6-phosphate deglycase, an enzyme involved in the metabolism of the fructation product glucoselysine. *The Biochemical journal* **392**, 263-269 (2005).
134. M. M. Ali *et al.*, Fructose-asparagine is a primary nutrient during growth of Salmonella in the inflamed intestine. *PLoS pathogens* **10**, e1004209 (2014).
135. A. Sabag-Daigle *et al.*, Identification of Bacterial Species That Can Utilize Fructose-Asparagine. *Applied and environmental microbiology* **84**, (2018).
136. M. Hellwig, R. Matthes, A. Peto, J. Löbner, T. Henle, N-ε-fructosyllysine and N-ε-carboxymethyllysine, but not lysinoalanine, are available for absorption after simulated gastrointestinal digestion. *Amino acids* **46**, 289-299 (2014).
137. E. Wiame, E. Van Schaftingen, Fructoselysine 3-epimerase, an enzyme involved in the metabolism of the unusual Amadori compound psicoselysine in *Escherichia coli*. *The Biochemical journal* **378**, 1047-1052 (2004).
138. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10**, 845-858 (2015).

139. E. V. Koonin, M. Y. Galperin, in *Evolutionary Concept in Genetics and Genomics*. (Boston: Kluwer Academic, 2003), chap. 2.
140. J. Hallin, C. R. Landry, Regulation plays a multifaceted role in the retention of gene duplicates. *PLoS biology* **17**, e3000519 (2019).
141. M. Graninger, B. Nidetzky, D. E. Heinrichs, C. Whitfield, P. Messner, Characterization of dTDP-4-dehydrorhamnose 3,5-epimerase and dTDP-4-dehydrorhamnose reductase, required for dTDP-L-rhamnose biosynthesis in *Salmonella enterica* serovar Typhimurium LT2. *The Journal of biological chemistry* **274**, 25069-25077 (1999).
142. R. J. Stern *et al.*, Conversion of dTDP-4-keto-6-deoxyglucose to free dTDP-4-keto-rhamnose by the rmlC gene products of *Escherichia coli* and *Mycobacterium tuberculosis*. *Microbiology (Reading, England)* **145 (Pt 3)**, 663-671 (1999).
143. C. Dong *et al.*, High-resolution structures of RmlC from *Streptococcus suis* in complex with substrate analogs locate the active site of this class of enzyme. *Structure (London, England : 1993)* **11**, 715-723 (2003).
144. A. Freitag *et al.*, Metabolic engineering of the heterologous production of clorobiocin derivatives and elloramycin in *Streptomyces coelicolor* M512. *Metabolic engineering* **8**, 653-661 (2006).
145. N. Kochanowski *et al.*, Influence of intracellular nucleotide and nucleotide sugar contents on recombinant interferon-gamma glycosylation during batch and fed-batch cultures of CHO cells. *Biotechnology and bioengineering* **100**, 721-733 (2008).
146. T. Mishra, P. V. Balaji, Highly clade-specific biosynthesis of rhamnose: present in all plants and in only 42% of prokaryotes. Only *Pseudomonas* uses both D- and L-rhamnose. *bioRxiv*, 854612 (2019).
147. P. Messner, U. B. Sleytr, Asparaginyl-rhamnose: a novel type of protein-carbohydrate linkage in a eubacterial surface-layer glycoprotein. *FEBS letters* **228**, 317-320 (1988).
148. T. Yanagisawa *et al.*, *Neisseria meningitidis* Translation Elongation Factor P and Its Active-Site Arginine Residue Are Essential for Cell Viability. *PloS one* **11**, e0147907 (2016).
149. C. M. Sassetti, D. H. Boyd, E. J. Rubin, Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**, 77-84 (2003).
150. J. Ma, G. W. Hart, Analysis of Protein O-GlcNAcylation by Mass Spectrometry. *Curr Protoc Protein Sci* **87**, 24 10 21-24 10 16 (2017).
151. J. Ma, G. W. Hart, O-GlcNAc profiling: from proteins to proteomes. *Clin Proteomics* **11**, 8 (2014).
152. S. R. Modi, J. J. Collins, D. A. Relman, Antibiotics and the gut microbiota. *J Clin Invest* **124**, 4212-4218 (2014).
153. C. N. Spaulding, R. D. Klein, H. L. t. Schreiber, J. W. Janetka, S. J. Hultgren, Precision antimicrobial therapeutics: the path of least resistance? *NPJ Biofilms Microbiomes* **4**, 4 (2018).
154. M. A. Riley *et al.*, Resistance is futile: the bacteriocin model for addressing the antibiotic resistance challenge. *Biochem Soc Trans* **40**, 1438-1442 (2012).
155. V. N. Tra, D. H. Dube, Glycans in pathogenic bacteria--potential for targeted covalent therapeutics and imaging agents. *Chem Commun (Camb)* **50**, 4659-4673 (2014).
156. T. K. Fu *et al.*, Rhamnose Binding Protein as an Anti-Bacterial Agent-Targeting Biofilm of *Pseudomonas aeruginosa*. *Mar Drugs* **17**, (2019).
157. S. K. Ng *et al.*, A recombinant horseshoe crab plasma lectin recognizes specific pathogen-associated molecular patterns of bacteria through rhamnose. *PLoS One* **9**, e115296 (2014).
158. H. Tateno *et al.*, A novel rhamnose-binding lectin family from eggs of steelhead trout (*Oncorhynchus mykiss*) with different structures and tissue distribution. *Biosci Biotechnol Biochem* **65**, 1328-1338 (2001).
159. S. L. van der Beek *et al.*, Streptococcal dTDP-L-rhamnose biosynthesis enzymes: functional characterization and lead compound identification. *Mol Microbiol* **111**, 951-964 (2019).
160. S. Wang *et al.*, Synthesis of rhamnosylated arginine glycopeptides and determination of the glycosidic linkage in bacterial elongation factor P. *Chem Sci* **8**, 2296-2302 (2017).

161. D. M. Liang *et al.*, Glycosyltransferases: mechanisms and applications in natural product development. *Chem Soc Rev* **44**, 8350-8374 (2015).
162. P. A. Videira, F. Marcelo, R. K. Grewal, in *Carbohydrate Chemistry: Chemical and Biological Approaches Volume 43*. (The Royal Society of Chemistry, 2018), vol. 43, pp. 135-158.
163. W. Volkwein *et al.*, Switching the Post-translational Modification of Translation Elongation Factor EF-P. *Front Microbiol* **10**, 1148 (2019).
164. M. Charbonneau *et al.*, A structural motif is the recognition site for a new family of bacterial protein O-glycosyltransferases. *Mol Microbiol* **83**, 894-907 (2012).
165. M. E. Charbonneau *et al.*, O-linked glycosylation ensures the normal conformation of the autotransporter adhesin involved in diffuse adherence. *J Bacteriol* **189**, 8880-8889 (2007).
166. H. L. P. Tytgat *et al.*, Cytoplasmic glycoengineering enables biosynthesis of nanoscale glycoprotein assemblies. *Nat Commun* **10**, 5403 (2019).
167. J. Härle, A. Bechthold, Chapter 12. The power of glycosyltransferases to generate bioactive natural compounds. *Methods Enzymol* **458**, 309-333 (2009).
168. M. Dicker, R. Strasser, Using glyco-engineering to produce therapeutic proteins. *Expert Opin Biol Ther* **15**, 1501-1516 (2015).
169. C. Krauth, M. Fedoryshyn, C. Schleberger, A. Luzhetskyy, A. Bechthold, Engineering a function into a glycosyltransferase. *Chem Biol* **16**, 28-35 (2009).
170. N. Sethuraman, T. A. Stadheim, Challenges in therapeutic glycoprotein production. *Curr Opin Biotechnol* **17**, 341-346 (2006).
171. H. Clausen, H. H. Wandall, C. Steentoft, P. Stanley, R. L. Schnaar, in *Essentials of Glycobiology*, A. Varki *et al.*, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2015), pp. 713-728.
172. A. Naegeli, M. Aebi, Current Approaches to Engineering N-Linked Protein Glycosylation in Bacteria. *Methods Mol Biol* **1321**, 3-16 (2015).
173. T. G. Keys, M. Aebi, Engineering protein glycosylation in prokaryotes. *Curr opin sys biol* **5**, 23-31 (2017).
174. L. E. Yates, D. C. Mills, M. P. DeLisa, Bacterial Glycoengineering as a Biosynthetic Route to Customized Glycomolecules. *Adv Biochem Eng Biotechnol*, (2018).
175. J. B. McArthur, X. Chen, Glycosyltransferase engineering for carbohydrate synthesis. *Biochem Soc Trans* **44**, 129-142 (2016).
176. C. Li, L. X. Wang, Chemoenzymatic methods for the synthesis of glycoproteins. *Chem Rev* **118**, 8359-8413 (2018).
177. A. W. Truman *et al.*, Chimeric glycosyltransferases for the generation of hybrid glycopeptides. *Chem Biol* **16**, 676-685 (2009).
178. S. H. Park *et al.*, Expanding substrate specificity of GT-B fold glycosyltransferase via domain swapping and high-throughput screening. *Biotechnol Bioeng* **102**, 988-994 (2009).
179. E. B. Irvine, G. Alter, Understanding the role of antibody glycosylation through the lens of severe viral and bacterial diseases. *Glycobiology* **30**, 241-253 (2020).
180. R. Jefferis, Glycosylation as a strategy to improve antibody-based therapeutics. *Nat Rev Drug Discov* **8**, 226-234 (2009).
181. Y. Mimura *et al.*, Glycosylation engineering of therapeutic IgG antibodies: challenges for the safety, functionality and efficacy. *Protein Cell* **9**, 47-62 (2018).
182. U. Galili, The alpha-gal epitope and the anti-Gal antibody in xenotransplantation and in cancer immunotherapy. *Immunol Cell Biol* **83**, 674-686 (2005).
183. U. M. Abdel-Motal, K. Wigglesworth, U. Galili, Mechanism for increased immunogenicity of vaccines that form in vivo immune complexes with the natural anti-Gal antibody. *Vaccine* **27**, 3072-3082 (2009).
184. O. Oyelaran, L. M. McShane, L. Dodd, J. C. Gildersleeve, Profiling human serum antibodies with a carbohydrate antigen microarray. *J Proteome Res* **8**, 4301-4310 (2009).
185. M. E. Huflejt *et al.*, Anti-carbohydrate antibodies of normal sera: findings, surprises and challenges. *Mol Immunol* **46**, 3037-3049 (2009).

186. W. Chen *et al.*, L-rhamnose antigen: a promising alternative to α -gal for cancer immunotherapies. *ACS Chem Biol* **6**, 185-191 (2011).
187. L. Wang, A. Brock, B. Herberich, P. G. Schultz, Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498-500 (2001).
188. K. Wals, H. Ovaa, Unnatural amino acid incorporation in *E. coli*: current and future applications in the design of therapeutic proteins. *Front Chem* **2**, 15 (2014).
189. R. Brabham, M. A. Fascione, Pyrrolysine Amber Stop-Codon Suppression: Development and Applications. *Chembiochem : a European journal of chemical biology* **18**, 1973-1983 (2017).
190. A. Crnkovic, T. Suzuki, D. Soll, N. M. Reynolds, Pyrrolysyl-tRNA synthetase, an aminoacyl-tRNA synthetase for genetic code expansion. *Croat Chem Actas* **89**, 163-174 (2016).
191. A. Gorelik, D. M. F. van Aalten, Tools for functional dissection of site-specific O-GlcNAcylation. *RSC Chemical Biology* **1**, 98-109 (2020).
192. A. Dumas, L. Lercher, C. D. Spicer, B. G. Davis, Designing logical codon reassignment – Expanding the chemistry in biology. *Chem Sci* **6**, 50-69 (2015).
193. N. E. Fahmi, L. Dedkova, B. Wang, S. Golovine, S. M. Hecht, Site-specific incorporation of glycosylated serine and tyrosine derivatives into proteins. *J Am Chem Soc* **129**, 3586-3597 (2007).
194. T. Matsubara, K. Iijima, T. Watanabe, T. Hohsaka, T. Sato, Incorporation of glycosylated amino acid into protein by an in vitro translation system. *Bioorg Med Chem Lett* **23**, 5634-5636 (2013).
195. J. Deutscher, C. Francke, P. W. Postma, How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol Mol Biol Rev* **70**, 939-1031 (2006).
196. D. R. Griffiths, J. B. Pridham, The metabolism of 1-(ϵ -N-Lysyl)-1-deoxy-D-fructose by *Escherichia coli*. *J Sci Food Agric* **31**, 1214-1220 (1980).

Curriculum Vitae

Personal Information

Name: Franziska Maria Teresa Koller

Date of Birth: 10.02.1992, Munich

Work experience

- since 02/18: **Doctoral student**, LMU Munich, Microbiology, AG Jürgen Lassak
- 08/17 - 01/18: **Biologic-technical assistant**, Klinikum rechts der Isar, III. med, hematological research
- 04/17 - 09/17: **Student assistant**, Klinikum rechts der Isar, III. med, hematological research
- 01/16 - 04/16: **Internship** (3 months) in patent agent's office, Maiwald Patentanwalts-gesellschaft mBH

Higher education

- 10/14- 07/17: **Biology, Master of Science (grade 1,2)**, Technical University, Munich, focus areas: genetics, biochemistry, medical biology
- Master thesis: Klinikum rechts der Isar, III. med, hematological research, „*The significance of the microenvironmental SFRP1 for the development of p185 BCR-ABL- and MLL-AF9- triggered leukaemia*”
- 10/11 - 07/14: **Biology, Bachelor of Science (grade 2,3)**, University Regensburg, Regensburg
- Bachelor thesis: Chair of microbiology, University Regensburg, „*Construction and analysis of point mutations in the A” subunit of the RNA-polymerase*“

Poster presentation

- 07/18: Synthetic glycopeptides for generating specific anti rhamnosyl asparagine antibodies, Synthetic Biology III, Landshut, Germany
- 12/18: A set of rhamnosylation specific antibodies enables detection of novel protein glycosylations, 3rd International Conference on Post-Translational Modifications in Bacteria, Tübingen
- 03/19: Detection of protein rhamnosylation in bacteria using a set of novel amino sugar specific antibodies, Association for General and Applied Microbiology (VAAM), Mainz
- 08/19: Protein mono-rhamnosylation – a kingdom spanning post-translational modification, 2nd International Conference on Molecular Systems Engineering, Nijmegen

Languages

German - first language
English - very good knowledge
French - basic knowledge

Danksagung

Besonders danken möchte ich Prof. Dr. Jürgen Lassak, der mich sehr herzlich in seiner Arbeitsgruppe empfangen hat und während der Jahre immer ein offenes Ohr für alle Anliegen hatte. Vielen Dank für die Betreuung und Unterstützung, die Begeisterung für kleinste Erfolge und die ehrlichen Diskussionen.

Bedanken möchte ich mich auch bei allen Mitgliedern der Prüfungskommission, insbesondere bei Prof. Dr. Osman, der sich als Zweitgutachter zur Verfügung gestellt hat.

Grazyna Wlodarska-Lauer danke ich für die kreativen Pausen und die vielschichtigen Gespräche. Danke, dass du den Überblick über die Anliegen und Probleme der Mitarbeiter hast und immer ein Ansprechpartner bist.

Bei meinem Büro- und Laborkollegen Ralph Krafczyk möchte ich mich für die unermüdliche wissenschaftliche Betreuung bedanken. Die ganzen schönen Gespräche zu jeglichem Thema, das gegenseitige Hoch- und Runterziehen, die Reisen und das Feiern sind unvergesslich. Bei Jannis Brehm möchte ich mich für den schönen Start am Lehrstuhl, das Abliegen und die Fend/Fox Kombination bedanken. Die Zeit im und außerhalb des Labors zusammen mit Urte Tomasiunaite war unglaublich witzig und unkonventionell. Mit euch verging die Zeit wie im Fluge und ich bin sehr froh, so tolle Kollegen und Freunde in euch gefunden zu haben.

Bei Nicola Stein und Wolfram Volkwein möchte ich mich besonders für die wunderschöne gemeinsame Zeit im Büro bedanken. Der Austausch mit Nicola über wissenschaftliche wie auch private Probleme, innovative Produkte und viele andere Themen ist sehr bereichernd gewesen. Wolfram danke ich für seine offene und ehrliche Art und seinen herrlichen Humor.

Bei Judith Mehler, Alina Sieber, Nicola Gericke und Benedikt von Armansperg möchte ich mich für die Zusammenarbeit und die schönen Mittags- und Kaffeepausen bedanken, ihr habt die AG Lassak sehr bereichert. Bei Stephanie Göing, Miriam Pfab, Elisabeth Hoyer und Simone Eckstein bedanke ich mich für die schönen Gespräche in sämtlichen Mittags- und Kaffeepausen. Korinna Burdack und Ingrid Weitzl danke ich für die angenehme Zusammenarbeit.

Nicht nur zu Pandemie-Zeiten hat sich die Eselau-Alm in Wegscheid als wertvoller Inspirationsort erwiesen. Für sinnstiftende Gespräche am Lagerfeuer, aufreibende Volleyballturniere und brüderlich/schwesterlich geteilte „Himbeerbretter“ möchte ich meinen geschätzten Mitreisenden danken. Dank sei auch dem Lucullus-Personal um Giorgios Makris für die stets verständnisvolle und umsichtige Betreuung unserer Stammtische. Besonders zur Entspannung beigetragen haben sämtliche Aufenthalte in Tenno, im Blockhäuschen – sowie alle weiteren Urlaube und Ausflüge. Familie und Freunden möchte ich für die Auswahl (und partielle Finanzierung) der wirklich exquisiten Unterkünfte, erfrischende Isarbäder, gemeinsame Radtouren und lehrreiche Bastelstunden danken.