Exceptionally sweet -
Studies on the bacterial arginine
rhamnosyltransferase EarP

Dissertation

Zur Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

vorgelegt von

Ralph Krafczyk

aus Augsburg

München, Oktober 2018

Gutachter:

1. Prof. Dr. Kirsten Jung
2. Prof. Dr. Dario Leister

Datum der Abgabe: 31.10.2018

Datum der mündlichen Prüfung: 18.12.2018

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.

Ralph Krafczyk

München, 31.10.2018

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.

Ralph Krafczyk

Munich, 31.10.2018

Contents

Nomenclature

Gene products are numbered in a way that the first methionine of the wild type protein is designated "1" in the amino acid sequence (if present: independently of the N-terminal affinity tag). Amino acid identifiers are given in single letter code followed by the position within the primary sequence (Example: D13)

Abbreviations

diNAcBac	*N,N'*-Diacetylbacillosamine
ER	Endoplasmic reticulum
EM	Electron microscopy
Gal	Galactose
GalNAc	*N*-Acetylgalactosamin
Glc	Glucose
GlcNAc	*N*-Acetylglucosamin
GT	Glycosyltransferase
LLO	Lipid-linked oligosaccharide
LPS	Lipopolysaccharide
OST	Oligosaccharyltransferase
PTM	Post-translational modification
Rha	Rhamnose
TDP-Rha	TDP-β-L-rhamnose

Publications and Manuscripts Originating from this Thesis

Chapter 2:

Li X*, Krafczyk R*, Macošek J, Li Y-L, Zou Y, Simon B, Pan X, Wu Q-Y, Yan F, Li S, Hennig J, Jung K, Lassak J, Hu H-G. 2016. 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.

Chapter 3:

Krafczyk R*, Macošek J*, Jagtap PKA, Gast D, Wunder S, Mitra P, Jha AK, Rohr J, Hoffmann-Röder A, Jung K, Hennig J, Lassak J. 2017. Structural basis for earP-mediated arginine glycosylation of translation elongation factor EF-P. mBio 8:e01412-01417.

Chapter 4:

Volkwein W*, Krafczyk R*, Jagtap PKA, Parr M, Mankina E, Macošek J, Guo Z, Fürst MJLJ, Pfab M, Frishman D, Hennig J, Jung K & Lassak J. 2018. Switching the post-translational modification of elongation factor P. Manuscript.

Chapter 5:

Volkwein W, Maier C, Krafczyk R, Jung K, Lassak J. 2017. A versatile toolbox for the control of protein levels using N^{ε}-acetyl-l-lysine dependent amber suppression. ACS Synth Biol 6:1892-1902.

*Authors contributed equally

Contributions to Publications and Manuscripts presented in this Thesis

Chapter 2:

Xiang Li performed the organic synthesis, analysis of NMR and single crystal structures of small molecules, generated the *anti*-Arg^{Rha} antibody and wrote the manuscript. Ralph Krafczyk performed the confirmation of antibody specificity, produced proteins for the NMR-analysis of rhamnosylated EF-P and wrote parts of the manuscript. Jakub Macošek, Bernd Simon and Janosch Hennig performed the NMR determination of the anomeric carbon configuration of attached rhamnose and wrote parts of the manuscript. Yu-Lei Li assisted in the organic synthesis. Yan Zou, Qiu-Ye Wu and Fang Yan assisted in the organic synthesis and antibody generation. Xing Pan and Shan Li assisted in the conformation of antibody specificity. Jürgen Lassak performed the conformation of antibody specificity, contributed to study design and wrote the manuscript. Kirsten Jung and Hong-Gang Hu contributed to study design and assisted in modification of the manuscript.

Chapter 3:

Anja Hoffmann-Röder, Swetlana Wunder, and Daniel Gast performed the organic synthesis and NMR analysis of small molecules and wrote the corresponding section of Materials and Methods. Ralph Krafczyk performed the confirmation of antibody specificity raised against the rhamnosyl-arginine-comprising peptide. Additionally, Ralph Krafczyk constructed the EarP$_{Ppu}$ - and EF-P$_{Ppu}$ -encoding plasmids and purified all proteins used for biochemical analyses, NMR studies, and X-ray crystallography. Ralph Krafczyk also performed the biochemical *in vivo*/*in vitro* characterization of EarP$_{Ppu}$ and determined concentrations of TDP-β-L-rhamnose in *E. coli*, *P. putida*, and *P. aeruginosa*. TDP-β-L-rhamnose was synthesized by Jürgen Rohr, Prithiba Mitra and Amit Kumar Jha. Janosch Hennig, Jakub Macošek, and Pravin Kumar Ankush Jagtap performed and analyzed all protein NMR experiments. The crystallization screen was set up by Jakub Macošek. Janosch Hennig, Jakub Macošek, and Pravin Kumar Ankush Jagtap solved the crystal structure of EarP$_{Ppu}$. Jürgen Lassak., Janosch Hennig and Kirsten Jung designed the study. The manuscript was written by Ralph Krafczyk, Jakub Macošek, Pravin Kumar Ankush Jagtap, Kirsten Jung, Janosch Hennig and Jürgen Lassak.

Chapter 4:

Dmitrij Frishman, Marina Parr, Elena Mankina and Jürgen Lassak performed bioinformatic analyses. NMR studies were performed by Janosch Henning, Pravin Kumar Ankush Jagtap and Jakub Macošek. The corresponding proteins were produced and purified by Wolfram Volkwein. Miriam Pfab performed isoelectric focusing experiments. Ralph Krafczyk and Zhenghuan Guo performed *in vitro* rhamnosylation assays. All other biochemical and genetic analyses of β3Ωβ4 substitution variants of *E. coli* and *P. putida* were conducted by Wolfram Volkwein and Ralph Krafczyk. Wolfram Volkwein and Maximilian Josef Ludwig Johannes Fürst performed the biochemical analysis with EarP from *S. oneidensis* with contributions from Jürgen Lassak. Jürgen Lassak, Janosch Hennig, Kirsten Jung and Dmitrij Frishman designed the study. The manuscript was written by Wolfram Volkwein, Ralph Krafczyk, Kirsten Jung, Elena Mankina, Pravin Kumar Ankush Jagtap, Janosch Hennig and Jürgen Lassak.

Chapter 5:

The study was designed by Wolfram Volkwein, Kirsten Jung and Jürgen Lassak and directed as well as coordinated by Kirsten Jung and Jürgen Lassak. Wolfram Volkwein and Cristopher Maier constructed all strains and performed the enzyme assays. Ralph Krafczyk performed the parameter optimization for the Lux reporters. All authors contributed to the writing of the manuscript.

We hereby confirm the above statements:

Ralph Krafczyk   Kirsten Jung

Chapter 2:

Ralph Krafczyk supplied affinity purified elongation factor P (EF-P) of *Shewanella oneidensis* that was used for NMR analyses. Based on the resulting information that EF-P exhibits an α-rhamnosyl-arginine linkage, Xiang Li performed organic synthesis of small molecules and analysed them by NMR and X-ray crystallography. He then generated mono-rhamnosyl-arginine containing glycopeptides and used them to raise and purify the anti-Arg^{Rha} antibody. Ralph Krafczyk confirmed the specificity of this antibody and showed that it is sensitive enough to detect the epitope both in purified protein fractions and bacterial culture. Together, Xiang Li and Ralph Krafczyk demonstrated the possibility of generating sensitive and specific antibodies against mono-rhamnosyl-arginine that can be utilized in biochemical and diagnostic applications.

We hereby confirm the above statements:

Ralph Krafczyk   Xiang Li

Chapter 3:

During the studies on the structural and biochemical characterization of the rhamnosyltransferase EarP, Jakub Macošek performed extensive analyses of NMR and X-ray crystallography data of elongation factor P and the corresponding glycosyltransferase EarP from *Pseudomonas putida* respectively. The protein components for these experiments were supplied by Ralph Krafczyk. This data allowed the characterization of the overall protein structure and unveiled both the localization of the donor substrate and the architecture of the corresponding binding site. The biochemical analyses performed by Ralph Krafczyk, confirmed the structural data, and led to the identification of two amino acids that are likely to be involved in catalyzing the transfer reaction. Thus, Jakub Macošek and Ralph Krafczyk both contributed significantly to the successful characterization of a novel mechanism for arginine linked glycosylation.

We hereby confirm the above statements:

Ralph Krafczyk   Jakub Macošek

Janosch Hennig   Jürgen Lassak

Chapter 4:

In order to study the possibility of unnatural elongation factor P (EF-P) activation, Ralph Krafczyk generated and applied a novel reporter strain for the sensitive *in vivo* detection of protein-protein interactions. This analysis showed that *Pseudomonas putida* EarP interacts both with its cognate EF-P and the sequentially unrelated homologue from *Escherichia coli*. Based on this information, Wolfram Volkwein performed extensive *in vivo* functionality analyses of single substitution variants. These experiments identified several amino acids that are crucial for activity of *P. putida* EF-P and subsequently enabled *in vivo* rhamnosylation - and thus unnatural activation - of *E. coli* EF-P. Complemented by *in vitro* rhamnosylation assays that were established and performed by Ralph Krafczyk, the minimal sequential requirements for modification and activation of elongation factor P by rhamnosylation were determined. Together, Wolfram Volkwein and Ralph Krafczyk managed to demonstrate, that the post-translational modification of *E. coli* EF-P can be switched from lysylation to rhamnosylation and that this unnatural modification can activate the elongation factor.

We hereby confirm the above statements:

Ralph Krafczyk   Wolfram Volkwein

Summary

Bacterial protein glycosylation affects numerous cellular properties, including physiology and pathogenicity. The transfer of carbohydrates to a nitrogen atom is known as N-glycosylation and almost exclusively occurs on asparagine side chains. In contrast, EarP represents a novel type of arginine-modifying glycosyltransferases. This enzyme uses TDP-β-L-rhamnose as a donor substrate to activate the specialized translation elongation factor P (EF-P) in about 10 % of sequenced bacteria, including the clinically relevant species *Pseudomonas aeruginosa* and *Neisseria meningitidis.* The post-translational modification of EF-P is crucial for bacterial fitness and also constitutes a prerequisite for virulence. As the amido group of asparagine and the arginine guanidinium are chemically distinct, the activation of the latter might be based on a so far unsolved molecular mechanism. Consequently, the structural characterization of EarP and its products is of clinical and functional importance.

In this regard, NMR analyses unambiguously identified the product of the glycosylation reaction as α-rhamnosyl-arginine. Thus, EarP inverts the anomeric configuration of rhamnose during the reaction. Anomer-specific mono-rhamnosyl-arginine-containing peptides were synthetized and used to raise antibodies against the modified side chain. These immunoglobulins were characterized with respect to their sensitivity and specificity towards the target epitope and used to determine enzyme kinetics of EarP.

X-ray crystallography identified EarP as a member of the inverting GT-B superfamily and revealed the site for donor binding. Bioinformatic and mutant analyses elucidated the functional significance of several amino acids in orienting the nucleotide sugar and demonstrated the importance of two highly conserved aspartates for catalysis.

Additionally, NMR titration experiments revealed that EarP mainly binds the N-terminal β-barrel domain of its acceptor substrate EF-P. This information was utilized to generate the first synthetic target for EarP-mediated protein modification. The structurally but not sequentially related EF-P homologue from *E. coli* is naturally activated by lysylation of a lysine side chain. Successive mutation not only allowed modification but also activation of *E. coli* EF-P by the non-cognate and EarP-mediated rhamnosylation.

This thesis provides new insights into the structure-function relationship of inverting arginine glycosylation. Additionally, it lays the groundwork for the application of EarP in synthetic biology and clinical research.

Zusammenfassung

Die Glykosylierung bakterieller Proteine beeinflusst zahlreiche zelluläre Eigenschaften wie Physiologie und Pathogenität. Die Übertragung von Kohlenhydraten auf ein Stickstoffatom wird als N-Glykosylierung bezeichnet und erfolgt fast ausschließlich an Asparagin-Seitenketten. Im Gegensatz dazu gehört EarP zu einer neuen Klasse von Arginin-modifizierenden Glykosyltransferasen. In etwa 10 % der sequenzierten Bakterien, einschließlich der klinisch relevanten Spezies *Pseudomonas aeruginosa* und *Neisseria meningitidis,* verwendet dieses Enzym TDP-β-L-rhamnose als Donorsubstrat zur Aktivierung des spezialisierten Translationselongationsfaktors P (EF-P). Die post-translationale Modifikation von EF-P ist von entscheidender Bedeutung für die bakterielle Fitness und eine Voraussetzung für Virulenz. Da die Amidogruppe von Asparagin und die Guanidinogruppe von Arginin chemisch unterschiedlich sind, erfolgt die Aktivierung der letzteren durch einen bisher unerforschten molekularen Mechanismus. Folglich ist die strukturelle Charakterisierung von EarP und seinen Katalyseprodukten sowohl von medizinischer als auch funktioneller Bedeutung.

Mittels NMR wurde zunächst das Produkt der Glykosylierungsreaktion von EarP eindeutig als α-Rhamnosyl-Arginin identifiziert. Somit invertiert EarP die anomere Konfiguration von Rhamnose während der Reaktion. Anomer-spezifische mono-Rhamnosyl-Arginin enthaltende Peptide wurden synthetisiert und zur Generierung von Antikörpern verwendet. Diese Immunglobuline wurden hinsichtlich Sensitivität und Spezifität gegenüber dem Epitop charakterisiert und zur Bestimmung der Enzymkinetik von EarP verwendet.

Die Kristallstrukturanalyse von EarP ermöglichte nicht nur eine Zuordnung des Enzyms zur Superfamilie der invertierenden GT-B-Glykosyltransferasen, sondern zeigte auch die Position der Donorbindestelle auf. Weitere bioinformatische und Mutagenese-basierte Studien führten zur Identifizierung von zwei für die Katalyse wichtigen Aspartaten sowie von mehreren Aminosäuren, die für die Orientierung des Nukleotidzuckers von Bedeutung sind. NMR-Titrationen ergaben, dass EarP hauptsächlich die N-terminale β-Barreldomäne des Akzeptorsubstrates EF-P bindet. Diese Information wurde verwendet, um den ersten synthetischen Akzeptor für eine EarP-vermittelte Proteinmodifikation zu generieren. Das strukturell, aber nicht sequentiell verwandte EF-P-Homolog von *E. coli* wird natürlicherweise durch Lysylierung einer Lysin-Seitenkette aktiviert. Infolge sukzessiver Aminsoäureaustausche wurde nicht nur die Modifikation von *E. coli* EF-P durch eine EarP-vermittelte Rhamnosylierung erreicht, sondern auch die Aktivierung dieses Elongationsfaktors.

Diese Arbeit liefert somit neue Erkenntnisse über die Struktur-Funktionsbeziehung der invertierenden Arginin-Glykosylierung. Darüber hinaus legt sie den Grundstein für die Anwendung von EarP in der Synthetischen Biologie und der klinischen Forschung.

1 Introduction

1.1 Post-translational modification of proteins

Proteins are the functional units of living systems and exert most of the biological tasks in- and outside of the cell. The process of protein biosynthesis follows the transcription of genetic information from DNA to mRNA by the RNA polymerase and is carried out at the ribosome. The ribosome is a ribozyme that utilizes the sequential information on the mRNA to translate it into a chemically distinct chain of molecules - the peptides and proteins. This process is consequently referred to as translation. The building blocks used to assemble proteins are amino acids (1, 2). The universal genetic code encodes for 20 proteinogenic amino acids, therefore generating an immense pool of possible polypeptide combinations. However, only a fraction of protein sequences form functional units with specialized biological capabilities (3). This might be due to the limited inventory of usable side chains provided by the 20 canonical amino acids (4).

Selenocysteine and pyrrolysine were identified as the 21st and 22nd proteinogenic amino acids and allow an increase in proteome complexity (5, 6) (Figure 1). Specialized tRNAs are used to incorporate selenocysteine and pyrrolysine into the nascent peptide chain at the stop codons UGA and UAG, respectively (7). While this naturally occurring expansion of the genetic code does convey specialized function, it occurs only on a selection of species specific proteins (7). The pyrrolysine tRNA and the corresponding tRNA synthetase of the methanogens *Methanosarcina barkeri* and *Methanosarcina mazei* are orthogonal in bacterial cells and eukaryotes (8). The heterologous co-expression of these components mediates suppression of the amber stop codon UAG and therefore the site-specific, cotranslational incorporation of pyrrolysine (9). Over the years, selective engineering of tRNA synthetase - tRNA pairs has greatly increased the number of modified amino acids that can be integrated using this system (8-11). The resulting unnatural expansion of the genetic code has allowed the study of the structural and biochemical implications of various side chain modifications (12).

A naturally occurring and even more extensive increase in the arsenal of functional side chains is provided by the process of post-translational modification (13) (Figure 1). As the name suggests, this form of protein processing occurs after biosynthesis at the ribosome and can be divided into two groups. The hydrolytic cleavage of proteins is often associated with protein localization and modulation of enzyme activity without altering the chemical nature of amino acid side chains (4). Instead, specific proteases cleave stretches of amino acids that are inter alia used as signaling sequences and removed during secretion (7).
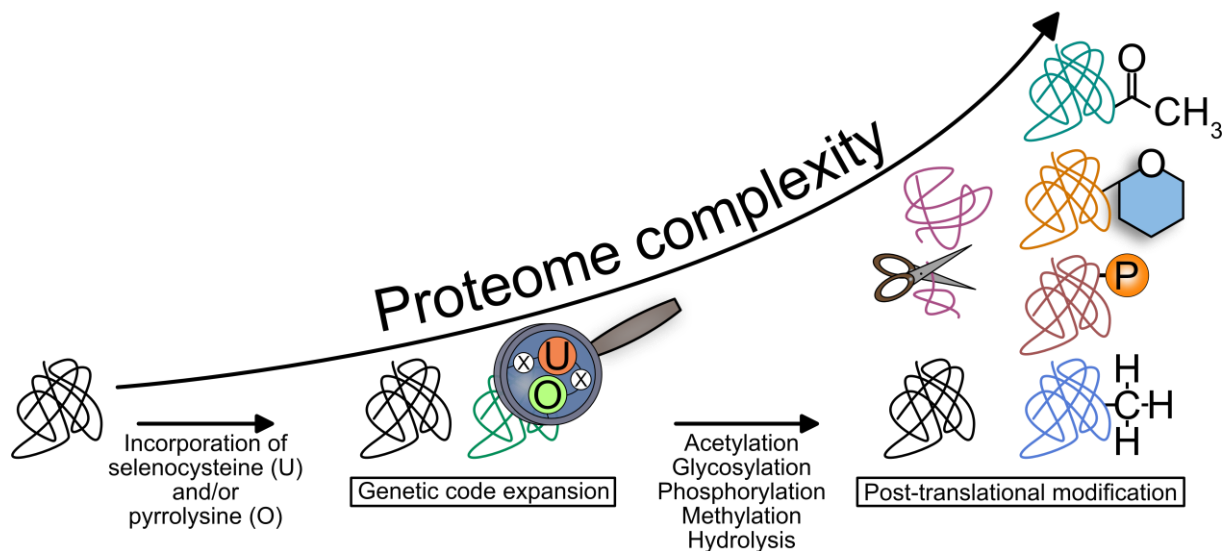
Figure 1. Overview of natural strategies to increase proteome complexity. Bacteria employ in principle two mechanisms to diversify their proteomes. Incorporation of the non-universally conserved amino acids selenocysteine (depicted as 'U' in orange circle) and pyrrolysine (depicted as 'O' in green circle) at the stop-codons UGA and UAG by specialized tRNAs represents an expansion of the universal genetic code. Post-translational modification of proteins involves, but is not limited to: acetylation (depicted as structural formula of $C(O)CH_3$), glycosylation (depicted as stylized hexose with blue filling), phosphorylation (depicted as 'P' in orange sphere), methylation (depicted as structural formula of CH_3) and hydrolytic cleavage (depicted as stylized scissors).

Covalent addition of specific groups on the other hand can dramatically alter the chemical properties of an amino acid side chain and thus control protein maturation and function (4, 14-16). Today, about half of all proteins are proven or predicted to receive at least one of the more than 200 known covalent post-translational modifications (13, 17-19). This diversity and the associated demand for specialized modification systems corroborate the immense functional potential that lies within this enzymatic extension of the genetic code (4). Protein phosphorylation represents one of the most thoroughly studied naturally occurring modifications (20). Since the initial identification of phosphorylated amino acid side chains in 1906 (21), numerous studies on its biosynthesis and its diverse physiological roles have been published (22, 23). The UniProt Knowledgebase contains annotated data on more than 550,000 non-redundant protein sequences (24). Over 55,000 of these proteins are experimentally verified to receive a phosphate modification and more than 80,000 phosphoproteins are predicted (17, 24) (Figure 2). Next to phosphorylation, glycosylation is the second most reported protein modification in this database (~6,800 reports). The number of experimentally verified glycoproteins is therefore an order of magnitude below that of phosphoproteins. However, almost 100,000 putative targets for protein glycosylation have been identified (17) (Figure 2). This discrepancy indicates an immense functional potential that remains yet to be uncovered. The post-translational linkage of carbohydrates to proteins has already been shown to affect numerous cellular functions (16). The study of glycosylated proteins in basic research, disease and glycoengineering is therefore an important focus in current glycobiology (15, 25).
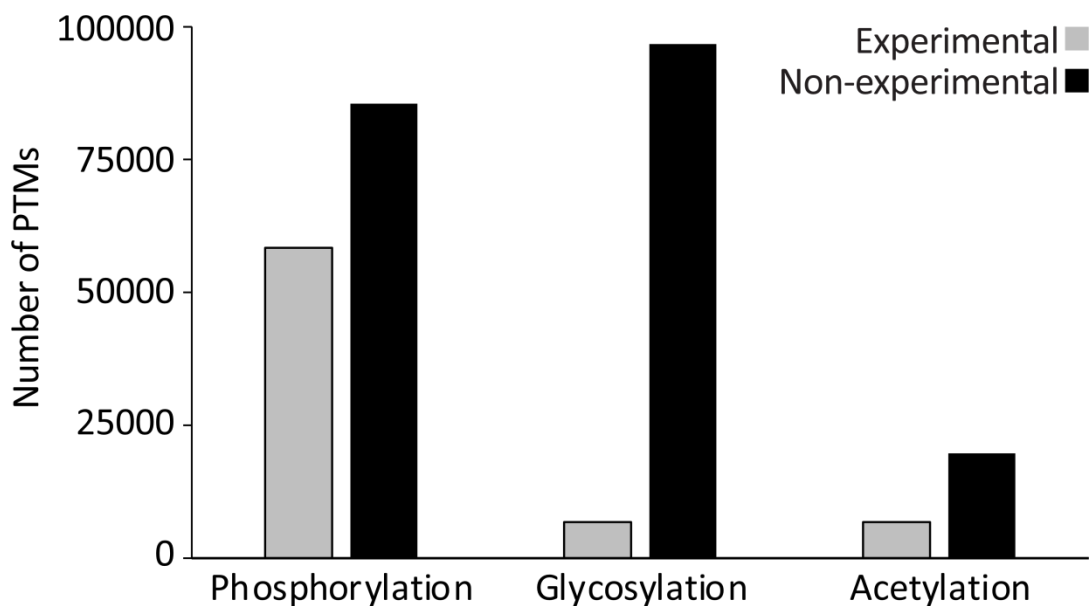
Figure 2. Number of experimentally and non-experimentally reported post-translational modifications in the UniProt Knowledgebase. Presented are the numbers of the top three experimentally verified (grey) and non-experimentally predicted (black) post-translational modifications phosphorylation, glycosylation and acetylation. Raw data was downloaded from the *Proteome-Wide PTM Statistics Curator* (17) (status October 2018).

1.2 Glycosylation and glycosyltransferases

Glycosylation describes the enzymatic transfer of a carbohydrate moiety from an activated donor to an acceptor substrate (26). This process leads to the formation of a glycoconjugate and is carried out by a group of enzymes (EC 2.4) that are referred to as glycosyltransferases (GTs) (27). The targets of glycosylation are highly diverse and comprise virtually all important biomolecules including not only proteins but also carbohydrates, lipids and nucleic acids (28, 29). Glycosylation is a central process in the biosynthesis of the bacterial cell wall (30-32) and the formation of structural and storage components such as cellulose and glycogen, respectively (33, 34). The glycosylation of viral DNA has been shown to be used as a mechanism by which bacteriophages manage to evade the host defence (35-37). Apart from that, bacteria use glycosylation to synthesize important secondary metabolites such as aminoglycosides and tetracyclines (38). This substrate and product diversity is reflected in the multitude of glycosyltransferase sequences (39). In 1998 the <u>c</u>arbohydrate <u>a</u>ctive en<u>zy</u>mes database (CAZy, http://www.cazy.org/) was established by Professor Dr. Bernard Henrissat, Professor Dr. Pedro Coutinho and colleagues. Presently, it provides genomic, structural and biochemical information on more than 450,000 glycosyltransferases. According to sequence similarity, these enzymes have been classified into 106 families (referred to as GTX, with X indicating the family number) (27, 40, 41).

In light of this enormous substrate and sequence diversity, the structural diversity of glycosyltransferases appears to be surprisingly limited (28). Four different GT folds

(GT-A, GT-B, GT-C, and lysozyme-type) have been identified so far (42) (Figure 3). The vast majority of structurally characterized glycosyltransferases exhibit either the GT-A or GT-B fold (43). Both of these structural families use nucleotide sugars as activated donor substrates and are referred to as Leloir-type glycosyltransferases (44).

A central open twisted β-sheet that is surrounded by several α-helices defines the overall structure of the GT-A fold (Figure 3A). A second structural element is formed by a smaller β-sheet and completes the active site of these enzymes. Despite being reminiscent of a single domain fold, members of the GT-A family exhibit two functional domains (27, 42). The N-terminal domain is dedicated with binding of the sugar donor substrate, while the acceptor is bound in the C-terminal domain. GT-A glycosyltransferases contain a conserved and functionally important DXD motif. This motif has been shown to facilitate the transfer reaction by coordinating a divalent cation that in turn neutralizes the developing negative charge on the phosphoryl group (39, 45, 46).

The GT-B fold (Figure 3B) is characterized by two opposing Rossmann domains that are connected by a flexible linker (47). The active site is localized within the pronounced interdomain cleft. The activated nucleotide sugar donor is oriented within a binding pocket formed by the C-terminal domain (28). While this domain is structurally well conserved, the acceptor binding N-terminal domain is prone to structural alterations. This diversity reflects the potential of GT-B glycosyltransferases to adapt to a multitude of different acceptor molecules (26). Although certain recurring peptide motifs have been observed (48), there is no evidence for strictly conserved residues that convey general function of GT-B enzymes (39). While some members of this structural family do bind divalent cations, these metal ions are most likely involved in product release (37, 49). The stabilization of the negative charge on the phosphate is instead taken over by positively charged amino acid side chains (42).

Non-nucleotide sugar utilizing glycosyltransferases (non-Leloir GTs) use lipid-linked donor substrates and belong to the less common GT-C and lysozyme-like structural families (50-52). Enzymes of the GT-C family exhibit both a α-helical transmembrane domain and an alternating α/β periplasmic domain. Among the tertiary sequences that have so far been determined for this group is the bacterial oligosaccharyltransferase (OST) PglB of *Campylobacter lari* (52) (Figure 3C). Bacterial penicillin binding proteins such as Pbp2 were the first to be crystallized and serve as examples for GTs with a lysozyme-like fold (50, 51). These membrane associated glycosyltransferases are involved in cell wall biosynthesis and often connected to a transpeptidase domain (50, 53) (Figure 3D).
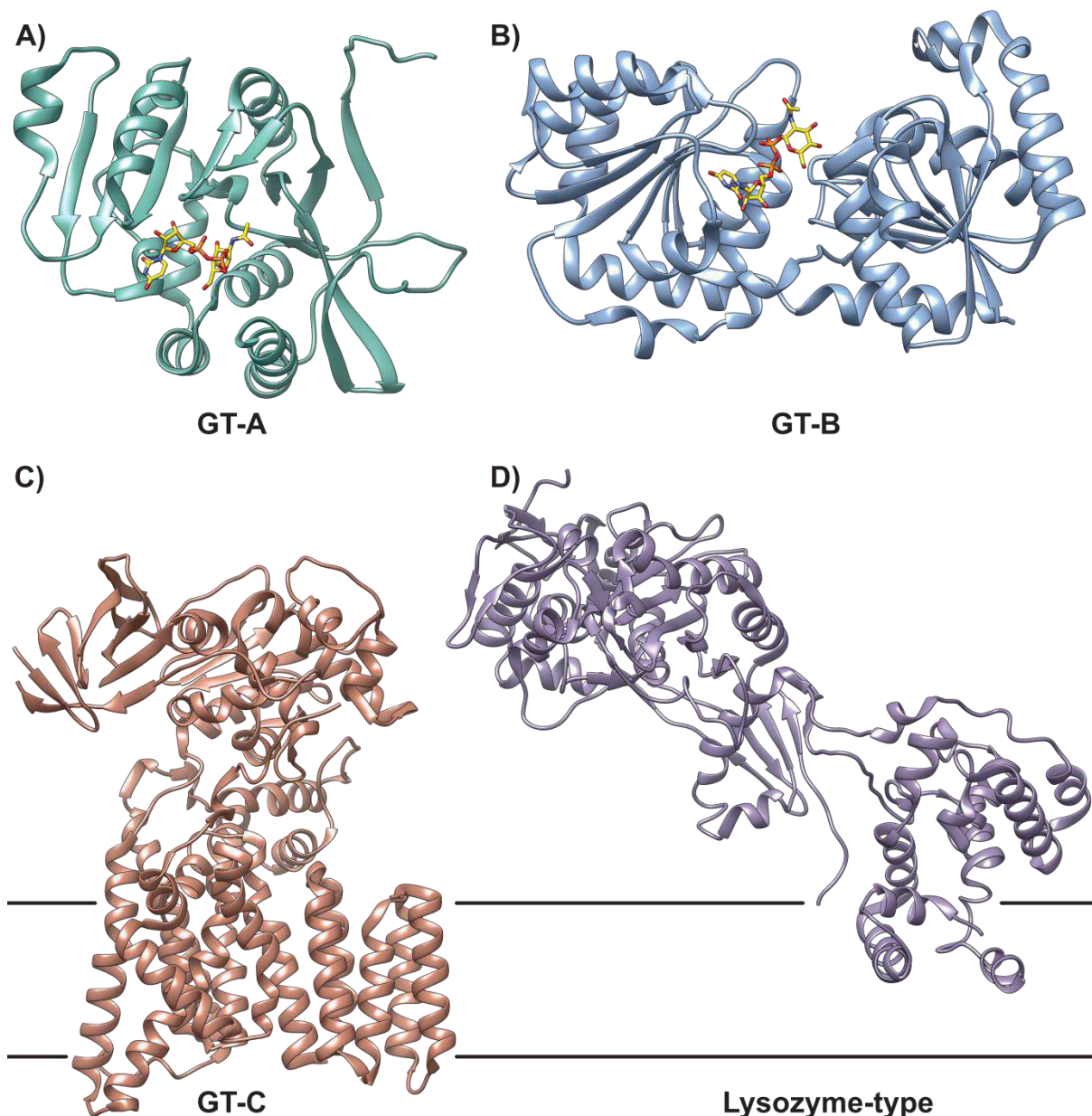
Figure 3. Structural overview of described glycosyltransferase folds. Ribbon diagrams of representative **A)** GT-A (green), **B)** GT-B (cyan), **C)** GT-C (brown) and **D)** lysozyme-type (purple) fold glycosyltransferases. The bound nucleotide sugars in the GT-A structure of mouse alpha-1,4-N-acetylhexosaminyltransferase (EXTL2) (PDB: 1OMZ) and in the GT-B structure of *E. coli* MurG (PDB: 1NLM) are depicted as sticks (carbon: yellow, phosphate: orange, oxygen: red, nitrogen: blue). The GT-C structure of *C. lari* PglB (3RCE) and the lysozyme-type structure of *Staphylococcus aureus* Pbp2 (PDB: 2OLV) are depicted according to their approximate orientation within the cytoplasmic membrane (black lines).

During the glycosyl transfer reaction, the configuration at the anomeric carbon atom of the sugar is either retained or inverted (Figure 4A). The specific stereochemical outcome of the reaction is a consequence of the structural characteristics of a given GT and can therefore be used to classify glycosyltransferases into inverting and retaining enzymes (27). As both GT-A and GT-B comprise members employing either of the two reaction mechanisms, these structural prerequisites are not limited to either of the folds and the stereochemical outcome is not dictated by overall GT fold (27, 42).
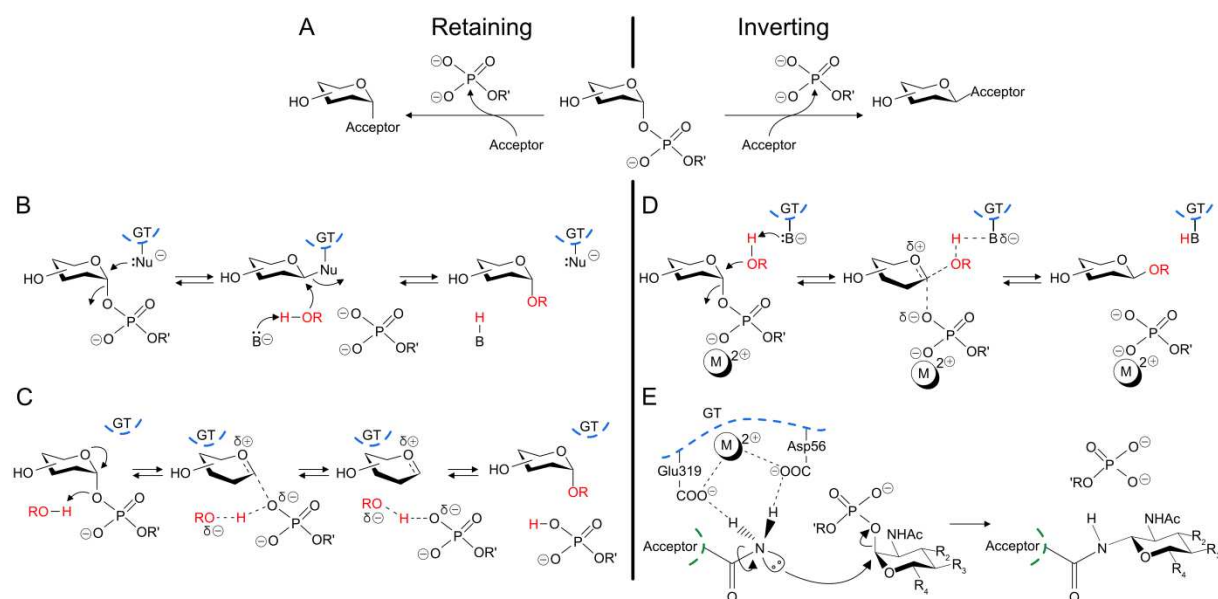
5

Figure 4. Overview of mechanisms for retaining and inverting glycosyl transfer reactions. A) Conceptual difference between retaining (left) and inverting (right) glycosylation. **B)** Double displacement mechanism for retaining GTs. **C)** S_Ni-like mechanism for retaining GTs. **D)** S_N2 mechanism for inverting GTs. **E)** Twisted amide mechanism for catalytic activation of asparagine by inverting GTs. The amino acid backbones of involved glycosyltransferases and acceptors are indicated by blue and green dashed lines respectively. In B), C) and D), the acceptor substrates are indicated by a red 'R'. This figure was adapted and extended from Breton *et al. (29)*, Lizak *et al.* (54), and Ardevol *et al. (55).*

The molecular mechanism of retaining glycosyltransferases has long been and still is a matter of debate (56). Early suggestions include a double displacement mechanism (Figure 4B) in which a nucleophilic GT side chain (glutamate or aspartate) attacks the anomeric carbon, thereby forming a glycosyl-enzyme intermediate. In a second step, the nucleophile of the acceptor substrate attacks the now vacant position of the leaving group, resulting in retention of the anomeric configuration (57, 58). This mode of transfer requires the presence of a properly localized nucleophilic side chain within the active site of the GT. Glycosyltransferases lacking a corresponding nucleophile are suggested to employ an uncommon S_Ni (substitution nucleophilic internal)-like mechanism (Figure 4C). Here, the nucleophile that is used to activate the acceptor substrate develops upon leaving group departure and formation of the phosphate ion (59, 60). As both the double displacement and the S_Ni-like carbohydrate transfer involve the formation of an oxocarbenium ion-like transition state, they can be considered as variants of a mechanistically similar reaction. The molecular differences most likely reflect the necessity to account for specific properties of the donor and acceptor substrates of a given retaining glycosyltransferase (55).

Inversion of the anomeric configuration during the transfer reaction is accomplished by a S_N2 (substitution nucleophilic bimolecular)-like reaction mechanism in which an acceptor nucleophile attacks the anomeric carbon of the donor substrate (Figure 4D) (29, 42). In general, the negatively charged side chain of a GT aspartate or glutamate is suggested to act as base catalyst that partially deprotonates the acceptor to increase its nucleophilicity

(42). By simultaneously promoting leaving group departure, either mediated by coordinated metal ions or positively charged amino acid side chains, inverting glycosyltransferases catalyse the nucleophilic attack of the acceptor onto the donor substrate and thus the direct displacement reaction (26, 42). Beside this general mechanism for inverting glycosyl transfer, alternative pathways have been described. The fucosyltransferase POFUT1 does not possess a suitable negatively charged amino acid for activation of the nucleophile in the reaction centre and was thus proposed to utilize an unusual S_N1 (substitution nuclear monomolecular)-like reaction mechanism instead (61).

Due to the poor nucleophilicity and the partial double bond character of the asparagine amido group, it was long unclear how this amino acid might be activated for a nucleophilic attack (4, 62). The landmark structure of the OST PglB from *C. lari* was the first to provide insight into the catalysis of inverting N-glycosylation of asparagine (52) (Figure 4E). A negatively charged dyad - namely glutamate 319 and aspartate 56 - of PglB forms hydrogen bond interactions with the amido group of the acceptor. These interactions induce rotation of the N-C bond and abolish conjugation of the π-electrons which thereby facilitating the nucleophilic attack onto the anomeric carbon (52, 54). In 2018, the structure of the catalytic STT3 subunit of the *Saccharomyces cerevisiae* OST complex was determined by cryo-electron microscopy analysis. This data suggests that PglB and STT3 employ a common molecular mechanism for activation of their acceptor substrates (63).

1.3 Post-translational protein glycosylation

Glycosylation of proteins represents one of the most diverse forms of post-translational modification (64). The most common target sites of protein glycosyltransferases are the hydroxyl groups of serine and threonine (O-glycosylation) as well as the amido group of asparagine (N-glycosylation) (65). In eukaryotes, the modification state of a newly N-glycosylated protein is essential to determine how it will be processed in the endoplasmic reticulum (ER) (66). During transfer of peptides into the ER lumen, a generic polysaccharide ($Man_5GlcNAc_2$) is transferred *en bloc* from a dolichol-linker to the accepting asparagine amido group by an oligosaccharyltransferase (67). The glycan-binding (lectin-)chaperones calnexin and calreticulin are part of the ER quality control system and retain these N-linked glycoproteins in the folding environment until they have attained their native conformation or misfolded proteins have been marked for degradation (67, 68). Due to its highly hydrophilic nature, the generic glycan intrinsically assists in folding of some glycoproteins (67, 69, 70). Correctly folded proteins are further processed in the ER and more importantly in the Golgi apparatus (71). The immense diversity of N-glycan structures that are obtained after these processing steps make them ideal components to play a role in multiple molecular interaction and recognition pathways (72).

Along the same line, O-glycosylation in the ER and Golgi apparatus produces a plethora of differently decorated peptides including cell-surface proteins. The resulting glycans are involved in various different recognition events and targeting to specific subcellular locations (73, 74). The modifications are sequentially assembled by soluble glycosyltransferases utilizing activated nucleotide sugar donor substrates (61, 75). Considering the multitude of biological processes in which glycans are involved, it is no surprise that aberrant modification patterns lead to formation of diseased states (76-78).

While protein glycosylation has been recognized as an important molecular principle in eukaryotes for 80 years, it has long been thought to be restricted to this domain of life (79, 80). The first O-linked bacterial glycoprotein was described in 1975 almost 40 years after the process of protein glycosylation had first been reported (79, 81, 82). Today, two pathways for bacterial O-glycosylation are known (80). The first way of biosynthesis has been identified both in Gram-positive and Gram-negative bacteria and follows the same principles that were described for eukaryotic O-glycosylation. Accordingly, soluble cytoplasmic glycosyltransferases utilize nucleotide sugar donor substrates to sequentially mono- or polyglycosylate their corresponding protein targets. This pathway is, for instance, important during assembly and stabilization of bacterial flagella (83, 84).

A functionally unrelated pathway for O-glycosylation was identified in Gram-negative bacteria, where a periplasmic OST modifies its target proteins using a preassembled undecaprenyl sugar donor (85). The abundant protein pilin was the first identified target of this modification system that has since been shown to exhibit relaxed acceptor substrate specificity (86-89).

Another 20 years after it had been established that O-glycosylation is a common theme in prokaryotic cells, the first pathway for bacterial N-glycosylation was discovered in the Gram-negative ε-proteobacterium *Campylobacter jejuni* (90-92) (Figure 5A). Analogously to eukaryotic nitrogen-linked carbohydrate transfer, a preassembled lipid-linked oligosaccharide (LLO) is used as donor substrate for this reaction. An undecaprenyl phosphate lipid carrier is initially glycosylated by PglC and the heptasaccharide (GalNAc$_2$[Glc]GalNAc$_3$-diNAcBac) is subsequently elongated by the concerted action of PglA, PglJ and PglH (93-96). The maturated LLO is translocated across the cytoplasmic membrane by the ATP-dependent flippase PglK (97, 98). PglB is structurally similar to the STT3 subunit of the eukaryotic OST complex and mediates the *en bloc* transfer of the heptasaccharide onto more than 60 different protein targets (99, 100). The targeting sequence of PglB has been identified to be D/E-X$_1$-N-X$_2$-S/T (where X represents any amino acid except proline (101)) and thus constitutes an extended version of the eukaryotic OST sequon N-X-S/T (102, 103).

Alternatively, the *Haemophilus influenzae* high-molecular-weight adhesin (HMW1) has been shown to be sequentially N-glycosylated in the cytoplasm by the corresponding

glycosyltransferase HMW1C (104) (Figure 5B). This mode of action resembles the pathway of O-glycosylation in the Golgi apparatus and represents an atypical mechanism of N-linked glycosylation. HMW1C is capable of mono- and diglycosylating HMW1 at 31 distinct sites, with the latter requiring initial modification with glucose (105). The modified adhesin is subsequently shuttled into the periplasm via the general secretory (Sec) pathway (106, 107). Only upon modification is the glycoprotein HMW1 capable of resisting degradation and maintaining interaction with the outer membrane translocator HMW1B that anchors the glycoprotein on the cell surface (105).
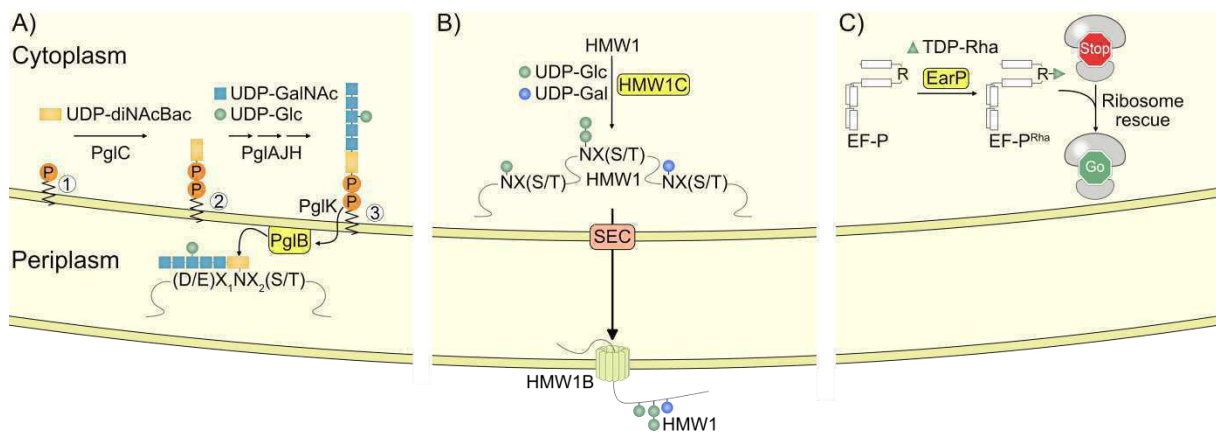

Figure 5. Overview of bacterial pathways for N-glycosylation of asparagine and arginine. A) Block transfer of the heptasaccharide (GalNAc$_2$[Glc]GalNAc$_3$-diNAcBac) from undecaprenyl pyrophosphate to the target sequon (D/E)X$_1$NX$_2$(S/T) in *C. lari*. Undecaprenyl phosphate [1] (depicted as black zigzag line and orange sphere) is initially glycosylated with diNAcBac by PglC. The resulting glycolipid [2] is successively elongated by PglA, PglJ and PglH. The lipid-linked oligosaccharide [3] is translocated across the cytoplasmic membrane by PglK and subsequently used as donor substrate by the membrane bound OST PglB (yellow) that mediates the transfer to the acceptor asparagine. **B)** Sequential transfer of monosaccharide moieties in *H. influenzae*. The high molecular weight adhesin glycosyltransferase (HMW1C, yellow) utilizes the nucleotide sugar donors UDP-glucose and UDP-galactose to modify its target protein at 31 distinct positions. 30 of these modification sites exhibit the eukaryote-like NX(S/T) sequon. The modified glycoprotein is subsequently translocated into the periplasm by the Sec (red) secretory pathway. From there it is excreted and concomitantly tethered to the outer membrane by HMW1B. **C)** Activation of elongation factor P in *Pseudomonas aeruginosa*. The EF-P arginine specific rhamnosyltransferase EarP (yellow) uses the nucleotide sugar TDP-β-ʟ-rhamnose to modify EF-P at a highly conserved arginine (R) in an unstructured acceptor loop region. This covalent carbohydrate linkage activates the elongation factor, which in turn facilitates translation of polyproline containing peptides at the ribosome. This is so far the only known bacterial pathway that involves glycosylation of arginine in the cytosol and is not associated with glycoprotein translocation. This figure was adapted and extended from Nothaft and Szymanski (80).

As many bacterial glycoconjugates are secreted or presented on the cell surface, it might appear that the major role of prokaryotic glycosylation is in mediating pathogenicity (108, 109). Indeed, several virulence associated glycans have been identified so far. For instance, glycosylated bacterial adhesins are known to facilitate pathogenicity by targeting the surface of host epithelial cells (110) and the glycosylation of flagella has been shown to be important for motility and host colonization in *Helicobacter pylori* and other pathogenic bacteria (80). Interestingly though, both commensal and non-virulent bacteria have been found to encode for the same systems as their pathogenic relatives (71). As protein glycosylation modulates a multitude of peptide properties including solubility, subcellular localization, antigenicity and activity it seems likely that the purpose of bacterial protein glycosylation goes beyond the mediation of virulence and is yet to be fully understood (71, 100).

1.4 Arginine glycosylation

An additional level of proteome complexity emerged with the discovery that arginine might be an additional target for protein N-glycosylation. The amylogenin of sweet corn was the first enzyme suggested to perform autoglucosylation on arginine. While this modification appears to stabilize the protein, the exact function of the glucose addition remains elusive (111). A similar mechanism of self-glucosylation and concomitant protein activation was described for the UDP-arabinopyranose mutase of rice (112). While both of these studies managed to determine arginine as the accepting amino acid of the glycosyl transfer reaction, the underlying catalytic mechanism remains unclear.

In 2013 two groups independently reported the first enzyme directly involved in N-glycosylation of arginine *(113, 114)*. The effector glycosyltransferase NleB1 is a pathogenicity factor of enteropathogenic *E. coli* and injected into the host via type III secretion. After translocation, this enzyme blocks death receptor signaling by inactivating several downstream targets of the tumor necrosis factor receptor (TNFR). Specifically, the addition of GlcNAc to a conserved arginine inactivates the TNFR1-associated death domain protein (TRADD) (113, 114). Structural analysis of the NleB1 orthologue SseK3 from *Salmonella enterica*, revealed that the glycosyltransferase exhibits a GT-A fold and contains a conserved DXD motif for binding of divalent cations (115). SseK3 hydrolyzes the nucleotide sugar donor in absence of the acceptor substrate (115). While analysis of the resulting sugar product suggests a retaining mode of glycosyl transfer, this finding is still a matter of debate as structural data on the modified acceptor is missing.

In 2015, another case of arginine glycosylation was initially reported in the facultative γ-proteobacterium *Shewanella oneidensis* (116) and subsequently confirmed in *Pseudomonas aeruginosa* (117). The specialized translation elongation factor P (EF-P) is a universal bacterial orthologue of the eukaryotic and archaeal initiation factor 5 (e/aIF5A) and

involved in the rescue of polyproline stalled ribosomes (118-120). In about 25 % of sequenced bacteria, including *E. coli*, EF-P (lysine-type) is modified by β-lysylation of a conserved lysine via the EF-P-lysine lysyltransferase EpmA (119, 121-123). Conversely, γ-proteobacteria and 10 % of sequenced bacteria in total encode for an EF-P isoform (arginine-type) that harbors an invariant arginine at the position corresponding to the modification site (116). Mass-spectrometry and mutational analysis revealed that a deoxyhexose moiety is covalently linked to this amino acid side chain. The presence of this modification is dependent on a domain of unknown function that is genetically coupled with *efp* and has hence been renamed to *EF-P arginine 32 rhamnosyltransferase essential for post-translational activation* (EarP). Additional mutational analyses identified TDP-β-L-rhamnose (TDP-Rha) as the activated donor substrate for this reaction. *In vitro* experiments with purified protein variants and biochemically synthesized TDP-Rha unambiguously demonstrated that EarP is a glycosyltransferase that mono-rhamnosylates its protein acceptor EF-P (116).

This modification shows marked differences to previously described mechanisms of N-glycosylation (52, 116, 124) (Figure 5). Considering the cellular localization and the molecular players involved (direct transfer of single sugar moieties onto the acceptor from an activated nucleotide sugar donor), the mode of action is in principle reminiscent of the HMW1C pathway (80, 105, 124) (Figure 5B). However, the product of the rhamnosylation reaction, EF-P, is not exported from the cytoplasm after modification but instead activated for function within the bacterial cell (116) (Figure 5C). Thus the elongation factor is to-date the only known example of a bacterial cytoplasmic glycoprotein. Additionally, the modification of arginine by a dedicated glycosyltransferase represents a new principle for N-linked glycosylation that is so far poorly understood both from a functional and a structural point of view. As the arginine guanidinium group is chemically distinct from the asparagine amido group, a so far unknown mode of acceptor activation might be involved in this reaction (116, 125). Additionally, deletion of *earP* abolishes pathogenicity in *P. aeruginosa,* indicating that the glycosyltransferase could also be an effective target for antibiotics (116). Taken together, EarP is of both mechanistic novelty and clinical significance. The aim of the present thesis is therefore the structural and functional characterization of this rhamnosyltransferase.

2 Resolving the α-glycosidic linkage of arginine-rhamnosylated translation elongation factor P triggers generation of the first Arg[Rha] specific antibody

Li X*, Krafczyk R*, Macošek J, Li Y-L, Zou Y, Simon B, Pan X, Wu Q-Y, Yan F, Li S, Hennig J, Jung K, Lassak J, Hu H-G. 2016. 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.

*Authors contributed equally

Resolving the α-glycosidic linkage of arginine-rhamnosylated translation elongation factor P triggers generation of the first Arg[Rha] specific antibody†

Xiang Li,‡[a] Ralph Krafczyk,‡[bc] Jakub Macošek,[d] Yu-Lei Li,[ae] Yan Zou,[a] Bernd Simon,[d] Xing Pan,[f] Qiu-Ye Wu,[a] Fang Yan,[e] Shan Li,[f] Janosch Hennig,[d] Kirsten Jung,[bc] Jürgen Lassak*[bc] and Hong-Gang Hu*[a]

A previously discovered posttranslational modification strategy – arginine rhamnosylation – is essential for elongation factor P (EF-P) dependent rescue of polyproline stalled ribosomes in clinically relevant species such as *Pseudomonas aeruginosa* and *Neisseria meningitidis*. However, almost nothing is known about this new type of *N*-linked glycosylation. In the present study we used NMR spectroscopy to show for the first time that the α anomer of rhamnose is attached to Arg32 of EF-P, demonstrating that the corresponding glycosyltransferase EarP inverts the sugar of its cognate substrate dTDP-β-L-rhamnose. Based on this finding we describe the synthesis of an α-rhamnosylated arginine containing peptide antigen in order to raise the first *anti*-rhamnosyl arginine specific antibody (*anti*-Arg[Rha]). Using ELISA and Western Blot analyses we demonstrated both its high affinity and specificity without any cross-reactivity to other *N*-glycosylated proteins. Having the *anti*-Arg[Rha] at hand we were able to visualize endogenously produced rhamnosylated EF-P. Thus, we expect the antibody to be not only important to monitor EF-P rhamnosylation in diverse bacteria but also to identify further rhamnosyl arginine containing proteins. As EF-P rhamnosylation is essential for pathogenicity, our antibody might also be a powerful tool in drug discovery.

Introduction

Glycosylation is one of the most important posttranslational modifications (PTMs) of proteins in biological systems[1,2] and is associated with numerous biological processes including viral and bacterial infection, cancer metastasis, inflammatory response, innate and adaptive immunity, as well as many signaling pathways.[3,4] For a long time, protein glycosylation was

[a]*Department of Organic Chemistry, School of Pharmacy, Second Military Medical University, Shanghai 200433, China. E-mail: huhonggang_fox@msn.com*

[b]*Department of Biology I, Microbiology, Ludwig Maximilians-Universität München, Munich, Germany*

[c]*Center for Integrated Protein Science Munich, Ludwig-Maximilians-Universität München, Munich, Germany. E-mail: juergen.lassak@lmu.de*

[d]*Structural and Computational Biology Unit, EMBL Heidelberg, Heidelberg 69117, Germany*

[e]*School of Pharmacy, Wei Fang Medical University, Shandong 261053, China*

[f]*Institute of Infection and Immunity, Taihe Hospital, Hubei University of Medicine, Shiyan, Hubei 442000, China*

‡ These authors contributed equally to this work.

considered to be restricted to eukaryotes. Today it is well accepted that also bacteria including important human pathogens contain a large number of *O*- and *N*-linked glycoproteins.[5,6] However until 2013 only one case of a sugar being added to arginine was reported.[7] At that time, two research groups discovered independently that the type III secretion system effector NleB, of enteropathogenic *Escherichia coli* (EPEC) acts as arginine-*N*-acetylglucosamine (Arg[GlcNAc]) transferase on human death receptor domains, thereby interfering with the host defense.[8,9] We elucidated that another type of arginine glycosylation plays an important role in the activation of the specialized translation elongation factor EF-P, which alleviates ribosome stalling at polyproline sequences (Fig. 1).[10–13]

For effective ribosome rescue certain bacteria, including not only the versatile γ-proteobacterium *Shewanella oneidensis* MR-1 but also the important human pathogens *Pseudomonas aeruginosa* and *Neisseria meningitidis*, post-translationally rhamnosylate a conserved Arg32.[14–16] When EF-P is bound to the ribosome the rhamnosylated arginine protrudes towards the peptidyltransferase center thereby contributing to the favorable positioning of the peptidyl-Pro-tRNA and stabilization of the CCA-end of the prolyl-tRNA.[14,17,18] Loss of the rhamnose modification abolishes the pathogenicity of *P. aeruginosa*[14] and
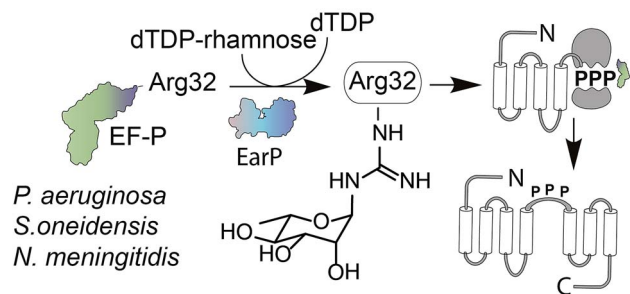
Fig. 1 EF-P arginine rhamnosylation and mode of action. Certain bacteria including *P. aeruginosa*, *S. oneidensis*, and *N. meningitidis* encode an EF-P variant with an invariant arginine at position 32. The glycosyltransferase EarP activates EF-P by rhamnosylation of Arg32 using dTDP-β-L-rhamnose as substrate. EF-P and its rhamnose modification stimulate proline–proline peptide bond formation thereby alleviating ribosome stalling at polyproline stretches. EF-P = translation elongation factor P; EarP = EF-P specific arginine rhamnosyl transferase for posttranslational activation.

increases its susceptibility to certain antibiotics.[15] Thus inhibition of EF-P rhamnosylation might be a novel strategy to selectively suppress virulence. However, little is known about the corresponding glycosyltransferase EarP or arginine rhamnosylation itself. Here we used NMR spectroscopy and found the rhamnosyl moiety on the protein acceptor EF-P in the α-configuration, unambiguously demonstrating that EarP is an inverting glycosyltransferase. Based on this result, we report the generation of the first high-affinity *anti*-Arg[Rha]-specific antibody that allowed us to detect rhamnosylated EF-P even from crude cell lysates. With this molecular tool in hand, we will not only be able to improve our understanding of EarP mediated EF-P rhamnosylation, but the antibody might also help to develop new potent targeted antibiotics and to unveil other arginine rhamnosylated proteins.

Results and discussion

EarP mediates an inverting glycosyl transfer reaction

Previously we and others demonstrated that mono-rhamnosylated EF-P-Arg32 in the EarP-arginine phylogenetic subfamily is essential to efficiently alleviate ribosome stalling at polyproline stretches.[14–16] However, nothing is known about the anomeric configuration of the attached sugar. Knowledge about steric configuration is important to understand how the modification contributes to the stabilization of the CCA-end of the P-site prolyl-tRNA and to classify EarP either as a retaining or inverting glycosyltransferase.[19,20] Notably, the activated sugar substrate is dTDP-β-L-rhamnose. In order to determine the configuration after glycosylation, we employed ^{13}C-edited NOESY-HSQC to assign the sugar resonances (Fig. 2a and b). $^1J_{CH}$ couplings can inform about the configuration of the anomeric carbon. An equatorial position of H1′ (α-anomer) would result in a coupling of around 170 Hz, whereas ≤160 Hz would indicate an axial position (β-anomer).[21,22] An undecoupled ^{13}C-HSQC gave rise to a coupling of 167 Hz, clearly indicating an α-configuration on the protein acceptor EF-P

(Fig. 2c and d). This was confirmed by the absence of an observable NOE between H1′ and H5′ (Fig. 2b). If H1′ was in the axial position a strong NOE should be visible. The change of configuration at the anomeric center from dTDP-β-L-rhamnose to Arg-α-L-rhamnose during the glycosylation reaction identifies EarP to be an inverting glycosyltransferase.

Synthesis of the α-rhamnosylated arginine containing hapten

Having solved the configuration of the rhamnose moiety attached to EF-P-Arg32 we were encouraged to raise specific antibodies against the modification employing an α-rhamnosylated arginine containing peptide, for immunization (Fig. 3a). Such a modification-specific antibody would be a useful tool to investigate EarP mediated rhamnosylation *in vivo* and *in vitro* but might also help in the identification of further arginine rhamnosylated proteins from diverse organisms.

Based on previous work,[23,24] we chose a strategy for glycopeptide synthesis that involves direct silver-promoted glycosylation between an *S*-alkyl-isothiourea and the amine of the amino acid side chain on the solid phase. First, we synthesized the key building block, *N*-glycosyl-*S*-alkyl-isothiourea **6**, starting from L-rhamnose **2** (Fig. 3b): glycosyl chloride **3** in the desired configuration was obtained using well established procedures (85% yield).[25,26] Subsequently, **3** was treated with potassium thiocyanate (KSCN) and tetrabutylammonium hydrogen iodide (TBAI) in anhydrous acetonitrile to get glycosyl isothiocyanate **4** (70% yield).[27] Next, glycosyl thiourea **5** was prepared *via* ammoniation of **4** in tetrahydrofuran (99% yield).[28] Finally, a two-step, one-pot procedure converted **5** into **6** in the presence of ethyl iodide and *tert*-butoxycarbonyl anhydride (75% yield).[29] Taken together from **2** to **6** we ended up with an efficiency of about 44%. The configuration of the attached rhamnose in the hapten depends on the stereochemistry of the key intermediate compounds **5** or **6**. NMR-HSQC showed that the $^1J_{CH}$ coupling underwent a change from 174 Hz to 154 Hz (Fig. 3c) when compound **4** was converted into **5**. Thus we determined the anomeric carbon configuration of compound **5**. A single crystal was obtained *via* slow evaporation of a dichloromethane/*n*-hexane solution at room temperature (Fig. 3d). With this (*N*-(2,3,4-tri-*O*-acetyl-6-deoxy-α-L-manno-pyranos-1-yl)thiourea) in hand, we could show unambiguously that the rhamnose moiety is attached in an α configuration, being consistent with rhamnosylated EF-P. All of the intermediates were fully characterized using ^{1}H-NMR, ^{13}C-NMR, and HR-Q-TOF-MS (Fig. 3c and ESI†).

To synthesize the hapten glycopeptide **1** with the rhamnose moiety in the α-configuration from building block **6**, we chose an on-resin glycosylation strategy (Fig. 3e): to obtain the linear peptide we used 9-fluorenylmethyloxycarbonyl (Fmoc) SPPS standard procedures with Fmoc-Orn(Alloc)-OH as the precursor for the Arg[Rha] residue. A 2-chlorotrityl resin acted as the solid support. Subsequent to the peptide assembly, the Alloc group was removed in the presence of tetrakis(triphenylphosphine) palladium (0) to get compound **7** on-resin.[30–34] Then the on-resin glycopeptide **8** was synthesized with a silver-promoted solid-phase glycosylation between the free amino group of **7** and the key building block **6**. Next, the rhamnose moiety was
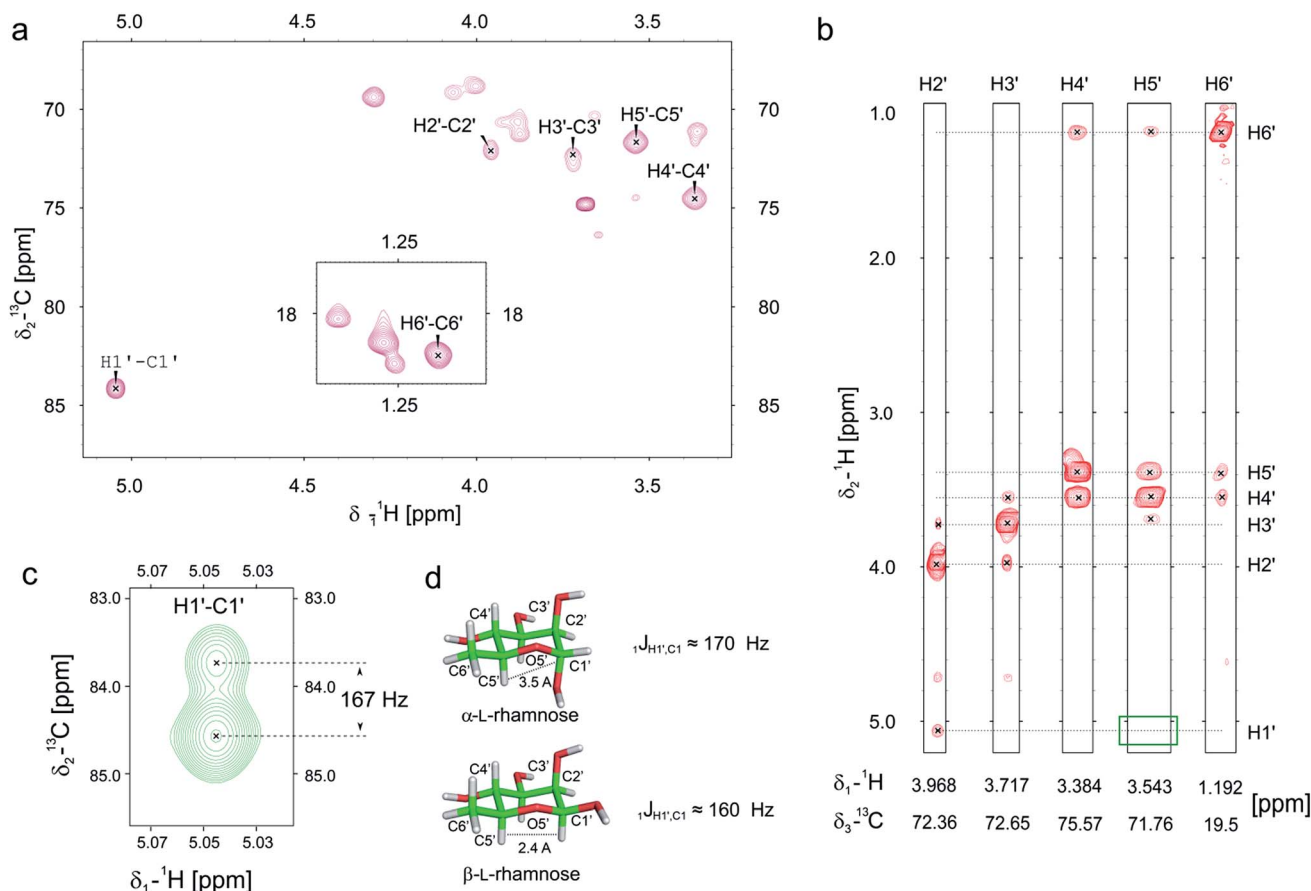
Fig. 2 Determination of the EarP rhamnosylation mechanism *via* NMR. (a) Zoomed in view of the sugar resonance region of the ^{13}C-HSQC of rhamnosylated EF-P. The assignment is based on a ^{13}C-edited NOESY-HSQC (exemplary strips are shown in panel b). Unassigned peaks at around 70 ppm and 18 ppm are the resonances of EF-P's threonine H$_\beta$/C$_\beta$ and methyl groups, respectively. (b) Strips of the ^{13}C-edited NOESY-HSQC to illustrate the lack of an observable NOE between H1′ and H5′ (green rectangle), which confirms that the rhamnose adopts an α-configuration, when bound to EF-P. (c) H1′–C1′ resonance of EF-P rhamnose from an undecoupled ^{13}C-HSQC to derive the $^{1}J_{CH}$ coupling. The resulting coupling of 167 Hz indicates an α-configuration of the sugar.[21,22] (d) Stick representations of α-L-, and β-L-rhamnose.

deacetylated with 5% NH$_2$NH$_2$ in dimethylformamide (DMF).[35] Subsequently the resin was treated with 5% triisopropylsilane (TIPS) in trifluoroacetic acid (TFA) to release the glycopeptide **1** which was further purified *via* preparative reverse-phase HPLC. We calculated from resin loading that the total yield of isolated **1** was 28%, manifesting a good efficiency for the on-resin glycosylation process.[36–38] All of the key intermediates were monitored using analytical HPLC and characterized using HR-Q-TOF-MS (Fig. S1†). The final peptide – Cys–Gly–Arg(Rha)–Gly–Leu – was characterized using 1D-NMR, 2D-NMR, and HR-Q-TOF-MS.

Generation and purification of a rhamnosyl arginine specific primary antibody

To raise the high affinity Arg^{Rha} specific antibody (*anti*-Arg^{Rha}), the hapten was conjugated to BSA as carrier protein *via* the free N-terminal sulfhydryl group distal from the arginine rhamnosyl side chain (Fig. 3a). The resulting BSA-glycoconjugate was injected into rabbits to raise polyclonal antibodies targeting the Arg^{Rha} moiety.[39,40] After the third immunization, the crude *anti*-sera were collected and their specificity was monitored by

employing an enzyme-linked immunosorbent assay (ELISA) analysis with horseradish peroxidase linked *anti*-rabbit IgG. The antibodies from two batches of crude *anti*-sera from two immunized rabbits were found to bind robustly and specifically to the BSA-glycoconjugate with high titers, showing strong immune reactivity even after 128 000 fold dilution (Fig. 3f and S3†).

To purify *anti*-Arg^{Rha} from the crude rabbit *anti* sera, in a first step we used a Protein A Sepharose 4 column (Amersham Biosciences). In a second purification step two agarose columns coupled with BSA or BSA carrying the non-glycosylated "naked" pentapeptide (H-CGRGL-OH) were used to exclude cross-reactivity. Taken together, these two steps resulted in a 95% pure *anti*-Arg^{Rha} antibody (Fig. S4 and S5†) showing a significantly improved specificity against the glycoconjugated BSA compared to the crude *anti*-sera (Fig. 3g and S6†).

anti-Arg^{Rha} allows sensitive and specific detection of endogenous EF-P^{Rha}

Having the *anti*-Arg^{Rha} at hand, we tested whether we can detect the EF-P rhamnose modification. Therefore we used 0.5 μg of
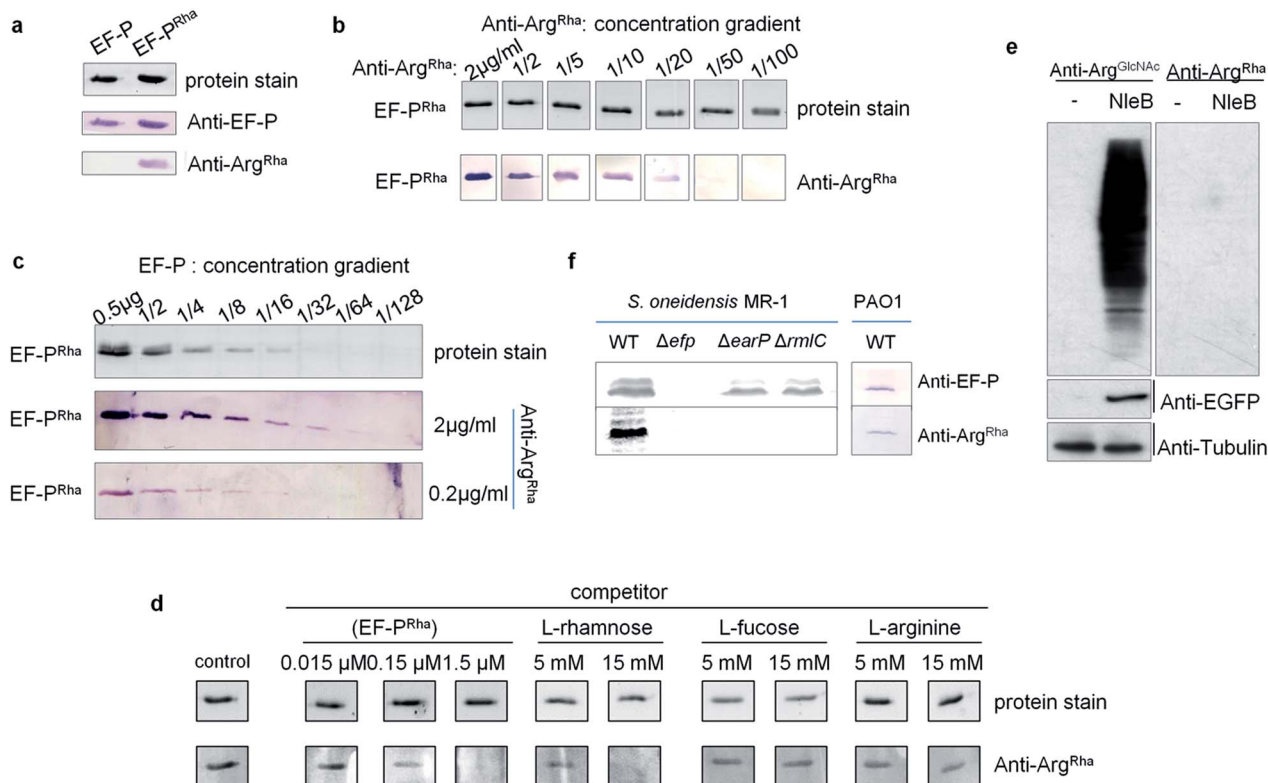
Fig. 3 Synthesis of mono-Arg^{Rha} peptide and antibody generation. (A) Work-flow of antibody generation: in the first step an Arg^{Rha} containing glycopeptide was synthesized *via* guanidyl formation, cleavage and subsequent coupling to bovine serum albumin (BSA). The resulting glyco-conjugate was used to immunize rabbits and accordingly to collect crude sera containing polyclonal antibodies against Arg^{Rha}. Using a two-step affinity chromatography technique we finally purified a highly sensitive and specific polyclonal *anti*-Arg^{Rha} antibody. Trt = trityl; Boc = *tert*-butoxycarbonyl. (B) Synthesis of building block **6**. Reagents and conditions: (a) acetyl chloride, r.t., 2 days, 85%; (b) KSCN, TBAI, and CH$_3$CN, reflux, 3 h, 70%; (c) NH$_3$, and THF, 1 h, 99%; (d) EtI, and MeOH, reflux, 3 h; then Boc$_2$O, Et$_3$N, and CH$_2$Cl$_2$, 75%. (C) NMR spectroscopic characterization of compounds **4**, **5**, **6** and **1**. (D) Single crystal structure of compound **5**. (E) Solid-phase synthesis of mono-Arg^{Rha} peptide **1**. Reagents and conditions: (a) TEA, DMF, AgNO$_3$, and **6** (3 eq.), r.t.; (b) 5% NH$_2$NH$_2$ in DMF; (c) 5% TIPS in TFA. (F) ELISA analysis of two batches of crude *anti*-sera. The crude *anti*-sera immunized by the BSA-glycoconjugate can recognize Arg^{Rha} with high affinity. *anti*-Serum 1# and *anti*-serum 2# were successively diluted up to 128 000 fold and subjected to indirect ELISA experiments against the BSA-glycoconjugate. (G) ELISA analysis of purified *anti*-Arg^{Rha}. Purified *anti*-Arg^{Rha} can recognize Arg^{Rha} with high specificity. The purified antibody was successively diluted up to 32 000 fold and subjected to indirect ELISA experiments against the BSA-glycoconjugate (BSA-Arg^{Rha}) and BSA carrying the non-glycosylated peptide (BSA-Arg).

purified EF-P which was modified *in vivo* (EF-P^{Rha}) employing the enzymatic activity of EarP. Unmodified EF-P served as a negative control. As expected, an EF-P specific antibody (*anti*-EF-P) detected both protein variants with no difference in signal intensity. By contrast *anti*-Arg^{Rha} specifically targeted only the EF-P Arg^{Rha} modification and no signal occurred in the lane with unmodified EF-P (Fig. 4a). The amino acid context of the arginine rhamnosylation site in EF-P is Ser$_{30}$–Gly$_{31}$–Arg$_{32}$(Rha)–Asn$_{33}$–Ala$_{34}$ and thus significantly differs from the peptide Cys–Gly–Arg(Rha)–Gly–Leu, previously used to raise *anti*-Arg^{Rha}. Thus, we conclude that our antibody recognizes Arg^{Rha} irrespective of the adjacent amino acid residues.

Next we assessed the detection limits of *anti*-Arg^{Rha} by using varying concentrations of either EF-P^{Rha} or *anti*-Arg^{Rha} (Fig. 4b and c). When keeping the EF-P^{Rha} concentration constant at 25 μg ml^{-1} (1.25 μM), the signal intensity progressively decreased starting from 2 μg ml^{-1} *anti*-Arg^{Rha} until no further detection was possible at an antibody concentration of 0.04 μg ml^{-1}. When keeping the *anti*-Arg^{Rha} concentration constant at 2 μg ml^{-1}, 15 ng of EF-P^{Rha} were efficiently detected.

To further prove the specificity against Arg^{Rha} we performed a Western Blot based competition assay in which our antibody was preincubated with various concentrations of ʟ-rhamnose, ʟ-fucose, or ʟ-arginine. Pre-added EF-P^{Rha} served as specific

Fig. 4 Sensitivity and specificity analysis of *anti*-Arg^{Rha} against EF-P^{Rha}. (a) The *anti*-Arg^{Rha} antibody specifically recognizes EF-P^{Rha}. Immuno-detection of purified EF-P both unmodified (EF-P) and rhamnosylated (EF-P^{Rha}) using *anti*-EF-P and *anti*-Arg^{Rha}. 0.5 µg of purified EF-P was subjected to SDS-PAGE and subsequent Western Blot analysis with 0.2 µg ml^{-1} *anti*-EF-P or *anti*-Arg^{Rha} respectively. (b) Immunodetection of EF-P^{Rha} when *anti*-Arg^{Rha} was successively diluted. (c) Immunodetection of EF-P^{Rha} when EF-P^{Rha} was successively diluted and *anti*-Arg^{Rha} was used in concentrations of 2 µg ml^{-1} or 0.2 µg ml^{-1}. (d) Cross-reactivity analysis of *anti*-Arg^{Rha} against L-rhamnose, L-fucose and L-arginine. 0.5 µg of purified EF-P^{Rha} were subjected to SDS-PAGE and subsequent Western Blot analysis using 0.2 µg ml^{-1} *anti*-Arg^{Rha}. *anti*-Arg^{Rha} was preincubated with varying concentrations of EF-P^{Rha}, L-rhamnose, L-fucose and L-arginine. Buffer only served as a control. (e) *anti*-Arg^{Rha} cannot detect Arg^{GlcNAc}. 293T cells were transfected with mock vector or pCS2–EGFP–NleB plasmids. Western Blot analysis of total cell lysates using either *anti*-Arg^{GlcNAc} or *anti*-Arg^{Rha}. *anti*-EGFP and *anti*-tubulin served as a control. (f) Detection of EF-P^{Rha} from *S. oneidensis* MR-1 lysates of wildtype (WT) and different mutant strains lacking *efp* (Δ*efp*) the glycosyltransferase EarP (Δ*earP*) or interfering with dTDP-β-L-rhamnose biosynthesis (Δ*rmlC*). *P. aeruginosa* PAO1 WT crude lysates served as an additional *in vivo* control. Approximately, 10^8 cells were used per lane.

competitor and completely prevented detection of EF-P already at concentrations of 1.5 µM (Fig. 4d). On the contrary even 15 mM of L-arginine or L-fucose could not decrease the signal intensity. At this concentration only L-rhamnose abolished the EF-P^{Rha} signal and therefore constitutes a competitor that is around 10 000 times less effective than EF-P^{Rha} (Fig. 4d and S7†). To examine whether the *anti*-Arg^{Rha} antibody shows cross reactivity towards other types of arginine *N*-glycosylation, we prepared lysates from 293T cells ectopically expressing NleB. As expected arginine GlcNAcylation could be detected using *anti*-Arg^{GlcNAc9,24} but no signal occurred when using *anti*-Arg^{Rha} (Fig. 4e). Taken together, our antibody can be regarded as highly sensitive and specific against arginine rhamnosylation.

We ultimately asked whether we can visualize endogenous arginine rhamnosylated proteins from crude cell lysates. From our sensitivity analysis we calculated that it was possible to detect single Arg^{Rha} with about 100 copies per cell when subjecting 10^8 cells and 2 µg ml^{-1} *anti*-Arg^{Rha} to Western Blot analysis. Rich media exponentially growing *E. coli* EF-P carry about 10 000 copies of EF-P per cell[41] and therefore it should be

possible to detect the modified protein. As *Enterobacteriales* modify EF-P with (*R*)-β-lysine[42–44] we used *S. oneidensis* which naturally employs EarP mediated rhamnosylation. Whereas we could readily identify EF-P in wildtype cells, mutants lacking either *efp* or *earP* gave no signal (Fig. 4f). Similarly, we could not detect EF-P rhamnosylation in a strain Δ*rmlC* that cannot produce the EarP substrate for glycosylation – dTDP-β-L-rham-nose. We used *P. aeruginosa* PAO1 crude cell lysates to test the activity of the *anti*-Arg^{Rha} antibody in another species and detected a single band (Fig. 4f). The band was verified to be EF-P in a parallel Western Blot, yielding a signal at the same height, by use of a *S. oneidensis anti*-EF-P antibody. Thus our *anti*-Arg^{Rha} represents a potent tool to detect EF-P rhamnosylation in diverse species.

Conclusion

We recently demonstrated the use of a high affinity *anti*-*N*-acetyl glucosaminyl arginine antibody (*anti*-Arg^{GlcNAc}) to monitor the glycosylation of human death receptor domains mediated by

NleB during EPEC infection.[9,24] Similarly, *anti*-Arg[Rha] represents a novel tool to diagnose infections caused by pathogens such as *P. aeruginosa*[14,15] or *N. meningitidis*.[16] Ultimately, our *anti*-Arg[Rha] might allow us to identify further arginine-rhamnosylated proteins from diverse species. This in turn might help to unveil novel antimicrobial targets and contribute to the task of overcoming the increasing problem of multi resistance. In this regard, it is indispensable to understand the mode of action of arginine dependent glycosyltransferases as they appear to be involved in pathogenicity development. However, our knowledge of *N*-linked glycosylation is so far mainly restricted to asparagine. The stereospecific outcome of the glycosylation reaction is a major characteristic of its molecular mechanism. By determining the α-anomeric nature of the rhamnosyl moiety on EF-P and with this the inverting mode of glycosyl transfer mediated by EarP, we made the first step to elucidating the catalysis of this novel type of glycosyltransferase. Our finding might also help to further understand how the sugar participates in stabilizing the CCA-end of the P-site prolyl-tRNA and thus contributes to the rescue of polyproline dependent ribosome arrest situations.

Author contributions

X. L. performed the organic synthesis, analysis of NMR and single crystal structures of small molecules, generated the *anti*-Arg[Rha] antibody and wrote the manuscript. R. K. performed the confirmation of antibody specificity, produced proteins for the NMR-analysis of rhamnosylated EF-P and wrote parts of the manuscript. J. M., B. S. and J. H. performed the NMR determination of the anomeric carbon configuration of attached rhamnose and wrote parts of the manuscript. Y.-L. L. assisted in the organic synthesis. Y. Z., Q.-Y. W. and F. Y. assisted in the organic synthesis and antibody generation. X. P. and S. L. assisted in the confirmation of antibody specificity. J. L. performed the confirmation of antibody specificity, contributed to study design and wrote the manuscript. K. J. and H.-G. H. contributed to study design and assisted in modification of the manuscript.

Acknowledgements

Notes and references

1 W. Yi, P. M. Clark, D. E. Mason, M. C. Keenan, C. Hill, W. A. Goddard III, E. C. Peters, E. M. Driggers and L. C. Hsieh-Wilson, *Science*, 2012, **337**, 975–980.

2 C. Xu, S. Wang, G. Thibault and D. T. Ng, *Science*, 2013, **340**, 978–981.

3 K. R. Mattaini and M. G. Vander Heiden, *Science*, 2012, **337**, 925–926.

4 E. K. Culyba, J. L. Price, S. R. Hanson, A. Dhar, C. H. Wong, M. Gruebele, E. T. Powers and J. W. Kelly, *Science*, 2011, **331**, 571–575.

5 Q. Lu, S. Li and F. Shao, *Trends Microbiol.*, 2015, **23**, 630–641.

6 H. Nothaft and C. M. Szymanski, *J. Biol. Chem.*, 2013, **288**, 6912–6920.

7 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.

8 J. S. Pearson, C. Giogha, S. Y. Ong, C. L. Kennedy, M. Kelly, K. S. Robinson, T. W. Lung, A. Mansell, P. Riedmaier, C. V. Oates, A. Zaid, S. Muhlen, 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.

9 S. Li, L. Zhang, Q. Yao, L. Li, N. Dong, J. Rong, W. Gao, X. Ding, L. Sun, X. Chen and F. Shao, *Nature*, 2013, **501**, 242–246.

10 S. Ude, J. Lassak, A. L. Starosta, T. Kraxenberger, D. N. Wilson and K. Jung, *Science*, 2013, **339**, 82–85.

11 L. K. Doerfel, I. Wohlgemuth, C. Kothe, F. Peske, H. Urlaub and M. V. Rodnina, *Science*, 2013, **339**, 85–88.

12 A. L. Starosta, J. Lassak, L. Peil, G. C. Atkinson, K. Virumae, T. Tenson, J. Remme, K. Jung and D. N. Wilson, *Nucleic Acids Res.*, 2014, **42**, 10711–10719.

13 L. Peil, A. L. Starosta, J. Lassak, G. C. Atkinson, K. Virumae, M. Spitzer, T. Tenson, K. Jung, J. Remme and D. N. Wilson, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 15265–15270.

14 J. Lassak, E. C. Keilhauer, M. Fürst, K. Wuichet, J. Gödeke, A. L. Starosta, J. Chen, L. Søgaard-Andersen, J. Rohr, D. N. Wilson, S. Häussler, M. Mann and K. Jung, *Nat. Chem. Biol.*, 2015, **11**, 266–270.

15 S. J. Hersch, M. Wang, S. B. Zou, K. M. Moon, L. J. Foster, M. Ibba and W. W. Navarre, *mBio*, 2013, **4**, e00180-13.

16 T. Yanagisawa, H. Takahashi, T. Suzuki, A. Masuda, N. Dohmae and S. Yokoyama, *PLoS One*, 2016, **11**, e0147907.

17 G. Blaha, R. E. Stanley and T. A. Steitz, *Science*, 2009, **325**, 966–970.

18 L. K. Doerfel, I. Wohlgemuth, V. Kubyshkin, A. L. Starosta, D. N. Wilson, N. Budisa and M. V. Rodnina, *J. Am. Chem. Soc.*, 2015, **137**, 12997–13006.

19 P. M. Coutinho, E. Deleury, G. J. Davies and B. Henrissat, *J. Mol. Biol.*, 2003, **328**, 307–317.

20 C. Breton, S. Fournel-Gigleux and M. M. Palcic, *Curr. Opin. Struct. Biol.*, 2012, **22**, 540–549.

21 I. Tvaroska and F. R. Taravel, *Adv. Carbohydr. Chem. Biochem.*, 1995, **51**, 15–61.

22 W. A. Bubb, *Concepts Magn. Reson.*, 2003, **19A**, 1–19.

23 X. Li, Y. Li, Y. Chen, Y. Zou, X. Zhuo, Q. Wu, Q. Zhao and H. Hu, *RSC Adv.*, 2015, **5**, 94654–94657.

24 M. Pan, S. Li, X. Li, F. Shao, L. Liu and H. Hu, *Angew. Chem., Int. Ed.*, 2014, **53**, 14517–14520.

25 A. Singhamahapatra, L. Sahoo, K. J. V. Paul, B. Varghese and D. Loganathan, *Tetrahedron Lett.*, 2013, **54**, 6121–6124.

26 C. Chang, S. Chang, C. Chao and K. T. Mong, *Tetrahedron Lett.*, 2009, **50**, 4536–4540.

27 T. K. Lindhorst and C. Kieburg, *Synthesis*, 1995, 1228–1230.

28 A. Kovalová, M. Ledvina, D. Saman, D. Zyka, M. Kubíčková, L. Zídek, V. Sklenár, P. Pompach, D. Kavan, J. Bílý, V. Ondřej, K. Zuzana, L. Martina, L. Ljubina, A. Mária, M. Hynek, R. Daniel, H. Kateřina, K. Vladimír and B. Karel, *J. Med. Chem.*, 2010, **53**, 4050–4065.

29 D. Ma, C. Xia, J. Jiang, J. Zhang and W. Tang, *J. Org. Chem.*, 2003, **68**, 442–451.

30 J. W. Trauger, R. M. Kohli, H. D. Mootz, M. A. Marahiel and C. T. Walsh, *Nature*, 2000, **407**, 215–218.

31 Q. Xiao and D. Pei, *J. Med. Chem.*, 2007, **50**, 3132–3137.

32 C. Qin, X. Bu, X. Zhong, N. L. J. Ng and Z. Guo, *J. Comb. Chem.*, 2004, **6**, 398–406.

33 C. Qin, X. Bu, X. Wu and Z. Guo, *J. Comb. Chem.*, 2003, **5**, 353–355.

34 Y. C. Huang, Y. M. Li, Y. Chen, M. Pan, Y. T. Li, L. Yu, Q. X. Guo and L. Liu, *Angew. Chem., Int. Ed.*, 2013, **52**, 4858–4862.

35 C. S. Bennett, S. M. Dean, R. J. Payne, S. Ficht, A. Brik and C. H. Wong, *J. Am. Chem. Soc.*, 2008, **130**, 11945–11952.

36 P. Wang, S. Dong, J. H. H. Shieh, E. Peguero, R. Hendrickson, M. A. Moore and S. J. Danishefsky, *Science*, 2013, **342**, 1357–1360.

37 M. N. Amin, J. S. McLellan, W. Huang, J. Orwenyo, D. R. Burton, W. C. Koff, P. D. Kwong and L. X. Wang, *Nat. Chem. Biol.*, 2013, **9**, 521–526.

38 J. Hudak, H. Yu and C. Bertozzi, *J. Am. Chem. Soc.*, 2011, **133**, 16127–16135.

39 U. Westerlind, H. Schröder, A. Hobel, N. Gaidzik, A. Kaiser, C. Niemeyer, E. Schmitt, H. Waldmann and H. Kunz, *Angew. Chem., Int. Ed.*, 2009, **48**, 8263–8267.

40 H. Cai, M. S. Chen, Z. Y. Sun, Y. F. Zhao, H. Kunz and Y. M. Li, *Angew. Chem., Int. Ed.*, 2013, **52**, 6106–6110.

41 G. An, B. R. Glick, J. D. Friesen and M. C. Ganoza, *Can. J. Biochem.*, 1980, **58**, 1312–1314.

42 T. Yanagisawa, T. Sumida, R. Ishii, C. Takemoto and S. Yokoyama, *Nat. Struct. Mol. Biol.*, 2010, **17**, 1136–1143.

43 W. W. Navarre, S. B. Zou, H. Roy, J. L. Xie, A. Savchenko, A. Singer, E. Edvokimova, L. R. Prost, R. Kumar, M. Ibba and F. C. Fang, *Mol. Cell*, 2010, **39**, 209–221.

44 M. Bailly and V. de Crecy-Lagard, *Biol. Direct*, 2010, **5**, 3.

3 Structural basis for EarP-mediated arginine glycosylation of translation elongation factor EF-P

Krafczyk R*, Macošek J*, Jagtap PKA, Gast D, Wunder S, Mitra P, Jha AK, Rohr J, Hoffmann-Röder A, Jung K, Hennig J, Lassak J. 2017. Structural basis for earP-mediated arginine glycosylation of translation elongation factor EF-P. mBio 8:e01412-01417.

*Authors contributed equally

Structural Basis for EarP-Mediated Arginine Glycosylation of Translation Elongation Factor EF-P

Ralph Krafczyk,[a] Jakub Macošek,[b] Pravin Kumar Ankush Jagtap,[b] Daniel Gast,[c] Swetlana Wunder,[c] Prithiba Mitra,[d] Amit Kumar Jha,[d] Jürgen Rohr,[d] Anja Hoffmann-Röder,[c] Kirsten Jung,[a] Janosch Hennig,[b] Jürgen Lassak[a]

Center for integrated Protein Science Munich (CiPSM), Department of Biology I, Microbiology, Ludwig-Maximilians-Universität München, Munich, Germany[a]; Structural and Computational Biology Unit, EMBL Heidelberg, Heidelberg, Germany[b]; Center for integrated Protein Science Munich (CiPSM), Department of Chemistry, Ludwig-Maximilians-Universität München, Munich, Germany[c]; University of Kentucky College of Pharmacy, Lexington, Kentucky, USA[d]

ABSTRACT Glycosylation is a universal strategy to posttranslationally modify proteins. The recently discovered arginine rhamnosylation activates the polyproline-specific bacterial translation elongation factor EF-P. EF-P is rhamnosylated on arginine 32 by the glycosyltransferase EarP. However, the enzymatic mechanism remains elusive. In the present study, we solved the crystal structure of EarP from *Pseudomonas putida*. The enzyme is composed of two opposing domains with Rossmann folds, thus constituting a B pattern-type glycosyltransferase (GT-B). While dTDP-β-L-rhamnose is located within a highly conserved pocket of the C-domain, EarP recognizes the KOW-like N-domain of EF-P. Based on our data, we propose a structural model for arginine glycosylation by EarP. As EarP is essential for pathogenicity in *P. aeruginosa*, our study provides the basis for targeted inhibitor design.

IMPORTANCE The structural and biochemical characterization of the EF-P-specific rhamnosyltransferase EarP not only provides the first molecular insights into arginine glycosylation but also lays the basis for targeted-inhibitor design against *Pseudomonas aeruginosa* infection.

KEYWORDS *Pseudomonas aeruginosa*, *Pseudomonas putida*, TDP-rhamnose, glycosylation, glycosyltransferase, nucleotide sugar, posttranslational modification, ribosomes, translation

Translation elongation is a nonuniform process and directly depends on the amino acids (aa) to be incorporated into the growing polypeptide chain (1). Due to its chemical and physical properties, proline delays the peptidyl transfer reaction (2), and ribosomes can even stall upon translation of distinct diprolyl-containing sequence motifs (Fig. 1) (3, 4). Such ribosome stalling is alleviated by the eukaryotic and archaeal elongation factor 5A (e/aEF-5A) (5–7) and its prokaryotic orthologue the bacterial translation elongation factor P (EF-P) (8–14). The L-shaped EF-P is composed of three β-barrel domains and structurally resembles tRNA in both size and shape (15). EF-P binds to the polyproline-stalled ribosomes between the binding sites of peptidyl-tRNA (P-site) and the exiting tRNA (E-site) (16) and stimulates peptide bond formation by stabilization of the CCA end of the P-site prolyl-tRNA (Fig. 1) (17, 18). A conserved positively charged residue—located at the tip of the EF-P KOW-like N-domain—is essential for function (11, 17). However, for full EF-P activity, this residue is posttranslationally elongated (19). Certain bacteria—including *Escherichia coli* and *Salmonella enterica*—ß-lysinylate a conserved lysine, K34^{EF-P}, by EpmA. This EF-P-specific ligase
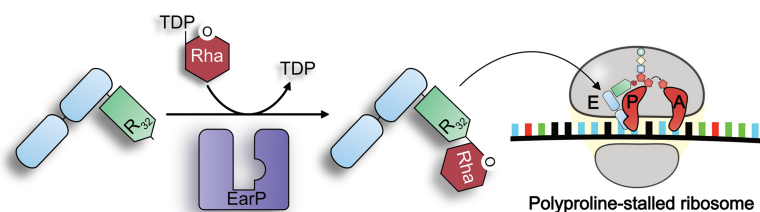
FIG 1 Activation and molecular function of EarP-arginine-type translation elongation factor EF-P. (Left) The bacterial translation elongation factor EF-P is composed of two OB-Fold domains (light blue) and one KOW-like N-domain (light green). In about 10% of all bacteria, EF-P is posttranslationally activated by α-glycosylation of a strictly conserved arginine (R32) (17, 26). The glycosylation reaction is catalyzed by the EF-P–arginine rhamnosyltransferase EarP (purple) using dTDP-β-L-rhamnose (TDP-Rha [red]) as a substrate. (Right) Activated EF-P is recruited to polyproline-stalled ribosomes and binds between the E and P sites. Thereby, R32^{EF-P} and the attached rhamnose moiety presumably stabilize the CCA end of the P-site prolyl-tRNA, which in turn stimulates Pro–Pro peptide bond formation and thus alleviates the translational arrest.

uses β-(R)-lysine as the substrate, which is generated by isomerization of α-(S)-lysine by employing the activity of the amino mutase EpmB (20–23). In contrast, activation of a phylogenetically distinct group of EF-Ps encoded in species such as *Pseudomonas aeruginosa* and *Neisseria meningitidis* depends on rhamnosylation of an arginine, R32^{EF-P}, in the equivalent position (17, 24, 25). Rhamnosylation is mediated by the recently discovered inverting glycosyltransferase EarP, which utilizes dTDP-beta-L-rhamnose (TDP-Rha) as donor substrate, resulting in α-rhamnosyl-arginine on the acceptor EF-P (26, 27). Unlike with the common and relatively well understood glycosylation of asparagine, sugar modifications on the guanidino group of arginine appear to be rare, and almost nothing is known about the molecular mechanism (28, 29). To date, there are only two reported cases of arginine glycosylation other than EF-P rhamnosylation. The first one described self-β-glycosylation of sweet corn amylogenin (30). In the second case, an effector glycosyltransferase termed NleB of enteropathogenic *E. coli* (EPEC) was shown to inactivate human cell death domain-containing proteins by *N*-acetylglucosaminylation of arginine, with this being a major pathogenicity determinant during infection (31). Similarly, a lack of *earP* abolishes the pathogenicity of *P. aeruginosa* (17). Accordingly, solving the molecular mechanism of arginine rhamnosylation might pave the way to ultimately design and develop targeted inhibitors against EarP.

Here we present the X-ray crystal structure of EarP from *Pseudomonas putida* KT2440 (EarP$_{Ppu}$) bound to its cognate nucleotide-sugar donor substrate TDP-Rha at a 2.3-Å resolution (PDB accession number 5NV8). Together with reporting the results of nuclear magnetic resonance (NMR) spectroscopy analyses and an *in vitro*/*in vivo* biochemical enzyme characterization, we lay the foundation for understanding arginine glycosylation.

RESULTS

Despite low sequence conservation most nucleotide sugar dependent (Leloir-type) glycosyltransferases adopt one of two major folding patterns, GT-A or GT-B (28). However, so far, there is no available information on the structure and folding properties of EarP. We used SWISS-MODEL (32), Phyre² (33), and the I-TASSER server for protein structure and function predictions (34–36) to generate fold recognition models of EarP from *Pseudomonas putida* (see Fig. S3A in the supplemental material). These predictions suggested the UDP-*N*-acetylglucosamine (UDP-GlcNAc)-dependent glycosyltransferases MurG from *E. coli* (MurG$_{Eco}$) (37) and O-GlcNAc transferase (OGT) from *Xanthomonas campestris* (OGT$_{Xca}$) (38) as structural orthologues. Accordingly, EarP$_{Ppu}$ adopts a clamp-like structure with two opposing Rossmann-like domains that are separated by an interdomain cleft (Fig. S3A). With this, the protein is presumably a GT-B-type glycosyltransferase (28).
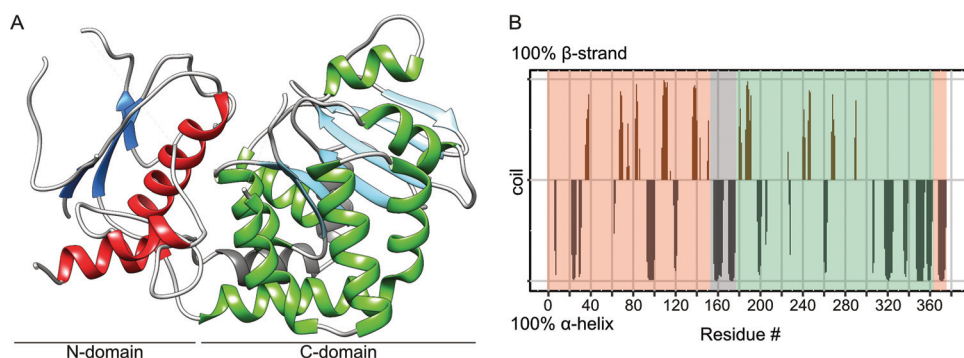
FIG 2 EarP folding pattern and topology. (A) Ribbon representation of the 2.3-Å crystal structure of EarP$_{Ppu}$. The illustration was generated with UCSF Chimera (82). Secondary-structure elements are shown, with α-helices in red and green for the N- and C-domains, respectively, and β-strands correspondingly in blue and cyan. The bipartite helix of the linker domain is grey. β-Strand 3, α-helix 5, and short loops with weak electron density are also displayed here but are missing in the PDB coordinate file to improve statistics, as discontinuity in the electron density did not allow proper modeling. (B) Secondary structure of EarP determined by NMR secondary shifts. The secondary structure of individual amino acids is indicated as propensity to form either an α-helix (grey) or a β-strand (brown). The amino acids with a propensity to adopt a random coil or lacking information about secondary structure were assigned zero values in the plot. The propensity values were obtained from C_α, C_β, NH, and H chemical shifts using TALOS+ (103). The N- and C-terminal EarP domains are boxed in peach and green, respectively. The interconnecting linker is boxed in grey.

Structure of *Pseudomonas putida* EarP. We were able to subsequently confirm the GT-B fold by having solved the crystal structure of EarP$_{Ppu}$ at 2.3-Å resolution (Fig. 2A; Data Set S2). Indeed, the EarP$_{Ppu}$ C-domain includes residues 184 to 361 and follows the Rossmann fold topology, with six β-strands ($\beta8$ to $\beta13$) and seven α-helices ($\alpha8$ to $\alpha14$) (Fig. 2 and see Fig. 4A). On the other hand, the N-domain (aa 1 to 153 and 362 to 377) could only be built in part. Although weak electron density for likely other regions of the N-domain was noticed, it was not sufficient to be unambiguously and reliably interpreted as particular missing parts of the protein chain. It is important to note that there is no indication that the diffraction data are twinned or anisotropic. The poor density of the N-domain is not caused by misinterpretation of noncrystallographic symmetry as crystallographic symmetry, because choosing a space group with lower symmetry does not improve the electron density. Yet, the structure has a higher-than-usual R-free (35.1%) value at this resolution, which cannot be explained by a simple absence of disordered regions. This is likely due to the N-domain adopting different conformations in different unit cells, causing crystal disorder. The potential mobility of the N-domain is further supported by higher average B-factors for this domain than for the C-domain (61 Å^2 versus 46 Å^2) (see Fig. S3B for B-factors mapped onto the protein structure). In addition, our rigorous attempts to obtain crystals in different space groups by screening various crystallization conditions were not successful. In the predicted structure (Fig. S3A, model 2), the N-domain features a central β-sheet of seven β-strands ($\beta1$ to $\beta7$), surrounded by the α-helices ($\alpha1$ to $\alpha5$ and $\alpha15$) (see Fig. 4A and Fig. S3A). In the crystal structure, only β-strands $\beta1$, $\beta2$, and $\beta3$, as well as α-helixes $\alpha1$, $\alpha5$, and $\alpha15$, are modeled (Fig. 2A). However, the missing structural elements in the protein N-domain are not in close vicinity to the active site according to the fold recognition model (Fig. S3A, model 2), and we did not observe any unassigned electron density in the vicinity of the ligand. Thus, despite this disorder, our crystal structure still provides crucial information important for understanding ligand binding and the catalytic mechanism. For structure validity assessment, the EarP crystal structure with electron density is shown in Fig. S4 in the supplemental material.

Furthermore, the presence of the predicted strands and helices and thus the validity of the model and crystal structure could be confirmed by NMR secondary chemical shifts (Fig. 2B). A prerequisite for this analysis is the backbone chemical shift assignment by triple-resonance NMR experiments. The relatively large size of EarP$_{Ppu}$ at 43 kDa exceeds the sensitivity limitations of NMR, demanding deuteration in order to decrease

cross-relaxation effects and to decrease the signal line width. Nonetheless, using transverse relaxation-optimized spectroscopy (TROSY)-based experiments, we were able to assign 62% of the EarP$_{Ppu}$ backbone.

The two domains are interconnected by a bipartite helix (α6, α7) comprising aa 156 to 176. This linker region together with an unstructured segment that positions α15 in the vicinity of the N terminus defines the floor of the cleft that separates the domains (Fig. 2A and see Fig. 4).

Based on these and previous data (17, 24–27), EarP was built in the <u>c</u>arbohydrate-<u>a</u>ctive en<u>zy</u>mes (CAZy) database (39) and now represents the new glycosyltransferase family GT104.

Analysis of the TDP-β-L-rhamnose binding site in the EarP C-domain. In Leloir-type GT-B glycosyltransferases, the nucleotide-sugar binding site is canonically located in the protein C-domain (40). Similarly, TDP-Rha in the EarP$_{Ppu}$ crystal structure is located in a binding pocket that is composed of residues located in the C-domain (Fig. 3A). F191^{EarP}, F252^{EarP}, and F258^{EarP} side chains form an aromatic cage that stacks against the base of the nucleotide moiety of TDP-Rha. The sugar ring of the nucleotide is then specifically recognized by a hydrogen bond between the hydroxyl group on C3′ of the sugar and the side chain of Q255^{EarP}. The diphosphate is recognized by hydrogen bonds formed with the side chain guanidine of R271^{EarP}, the Y193^{EarP} side chain hydroxyl, and backbone amides of E273^{EarP} and D274^{EarP}. The binding pocket is closed by the bulky side chain of Y193^{EarP}, which may sterically ensure proper positioning of the rhamnose sugar (Fig. 3A). The rhamnose sugar itself does not seem to make any contact with the protein and is solvent exposed. We further confirmed this by saturation transfer difference (STD) NMR experiments (41), where we did not observe any difference signal from the rhamnose moiety but did observe one from the TDP moiety of TDP-Rha (Fig. S5A).

In parallel, small-angle X-ray scattering (SAXS) of free EarP$_{Ppu}$ and EarP$_{Ppu}$ bound to TDP-Rha has been performed (Fig. S3D). The overall shape of the molecule could be validated to be the same in solution. Protein backbone conformational changes upon TDP-Rha binding are confirmed by chemical shift perturbations (see Fig. 7B); however, SAXS indicates that there are no large (>10-Å) conformational changes or movements of the two Rossmann fold domains with respect to each other upon binding of TDP-Rha, as the scattering density does not change from that in the free state. To show that TDP-Rha is bound to EarP under SAXS experimental conditions, STD NMR experiments were performed. They confirm again that TDP-Rha binding occurs with the ligand at a 7-fold excess compared to the amount of protein (Fig. S5B).

Database mining identified 432 EarP homologues representing about 10% of sequenced bacteria (Data Set S3) (17). Phylogenetically, EarP originated presumably in the betaproteobacterial subdivision and was horizontally transferred into the gammaproteobacterial orders of *Pseudomonadales*, *Aeromonadales*, and *Alteromonadales* (17). It can also be found in certain *Fusobacteria*, *Planctomycetes*, and *Spirochetes* (17).

In order to identify conserved amino acids, we used Clustal Omega (42) and generated a multiple-sequence alignment (Fig. 4A). We found 49 residues with a sequence conservation of ≥95%. Mapping of these residues onto the crystal structure revealed an accumulation at or near the interdomain cleft (Fig. 4B), including the binding pocket for the nucleotide sugar donor substrate (Fig. 3A), which is highly supportive of the correctness of the solved structure.

To substantiate our structural findings with biochemical data, we prepared EarP$_{Ppu}$ constructs with single-amino-acid substitutions of the individual residues forming the binding pocket and tested the activities of the EarP$_{Ppu}$ variants both *in vivo* and *in vitro* (Fig. 5). This included F191^{EarP}, F252^{EarP}, and F258^{EarP}, which form the aromatic pocket, as well as Y193^{EarP}, Q255^{EarP}, R271^{EarP}, and D274^{EarP}, which are involved in hydrogen bond networking (Fig. 5B).

Previously, we could show that the heterologous expression of *efp* and *earP* from *Shewanella oneidensis* in *E. coli* can fully complement a lack of EF-P (17) with respect to
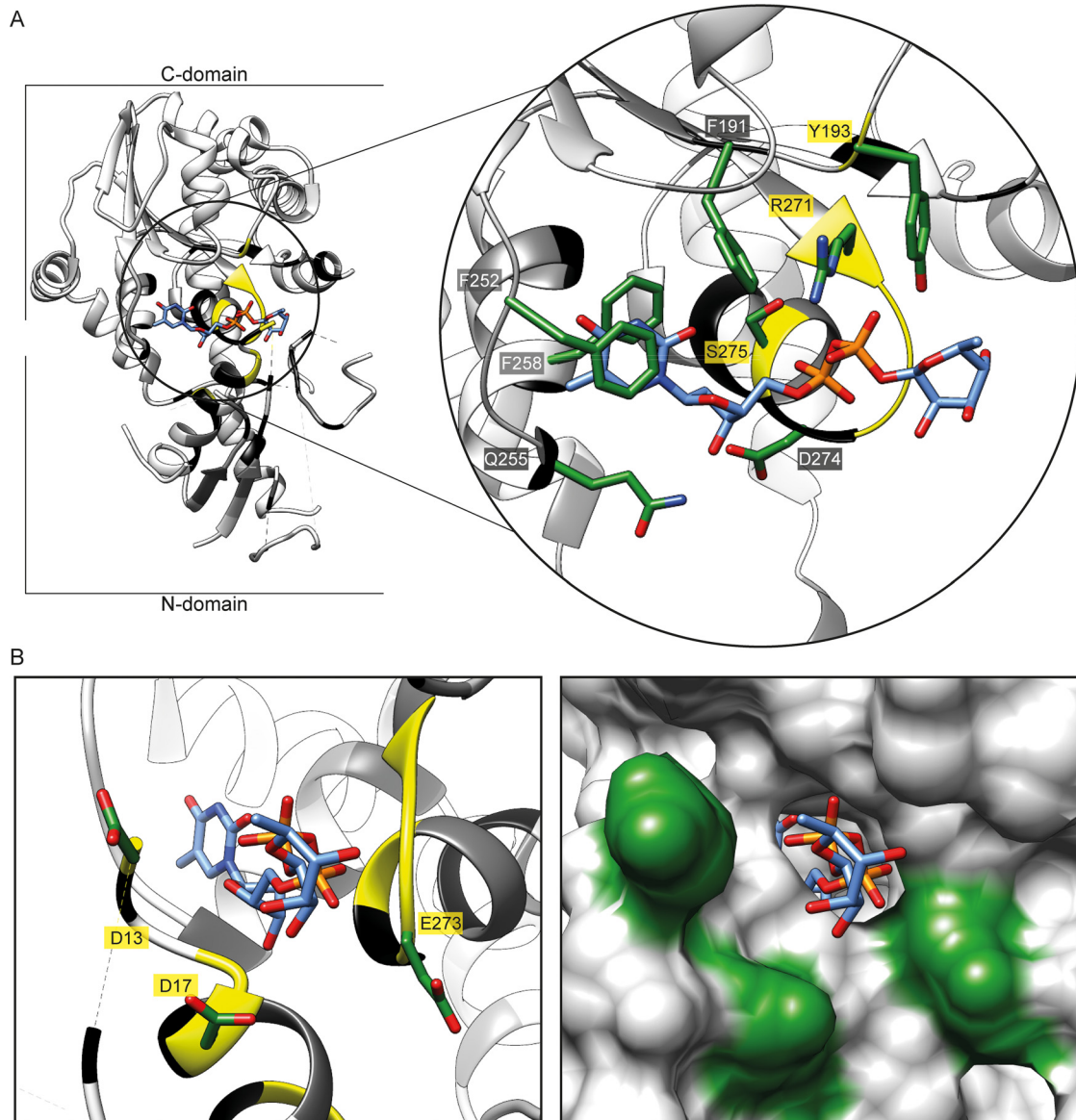
FIG 3 EarP TDP-β-ʟ-rhamnose binding site. (A, left) Three-dimensional structure of EarP$_{Ppu}$ in a ribbon representation. The TDP-Rha binding pocket in the C-domain is circled in black. (Right) Zoom into the nucleotide-sugar binding pocket with bound TDP-Rha (blue sticks). Important residues for TDP-Rha positioning are depicted as green sticks and labeled with single-letter code identifiers. (B, left) Ribbon representation of the nucleotide-sugar binding pocket with stick representation of bound TDP-Rha (blue sticks) as well as the three invariant residues D13, D17, and E273 (green sticks), which are presumably involved in catalysis. (Right) Surface representation of the nucleotide-sugar binding pocket with stick representation of bound TDP-Rha (blue). Surfaces of D13, D17, and E273 are in green. Ribbons are color coded according to their degree of conservation, as follows: yellow, 100%; black, ≥95%; dark grey, ≥90%; light grey, ≥50%; and white, <50% identical residues in all analyzed EarP orthologues. The electron density for TDP-Rha bound to EarP is shown in Fig. S3C. All illustrations were generated with UCSF Chimera (82).

the activation of the lysine-dependent acid stress response by the transcriptional activator CadC (11). Similarly, coproduction of wild-type EF-P$_{Ppu}$ and wild-type EarP$_{Ppu}$ (WT^{EarP}) can restore β-galactosidase activity in an *E. coli* P$_{cadBA}$::*lacZ* Δ*efp* strain (Fig. 5A and S1B). From the nine tested EarP$_{Ppu}$ substitution variants, we measured reduced β-galactosidase activities for the variants F191A^{EarP}, Y193A^{EarP}, R271A^{EarP}, S275A^{EarP}, and Y291A^{EarP}. The variants R271A^{EarP} and Y291A^{EarP} failed to induce β-galactosidase expression at all (Fig. 5B and S1B).

In parallel, the enzymatic activity of EarP$_{Ppu}$ was investigated *in vitro* by employing an anti-Arg^{Rha} antibody. The antibody was raised against a chemically synthesized glycopeptide antigen (SGR^{Rha}NAAIVK) and specifically detects arginine rhamnosylation
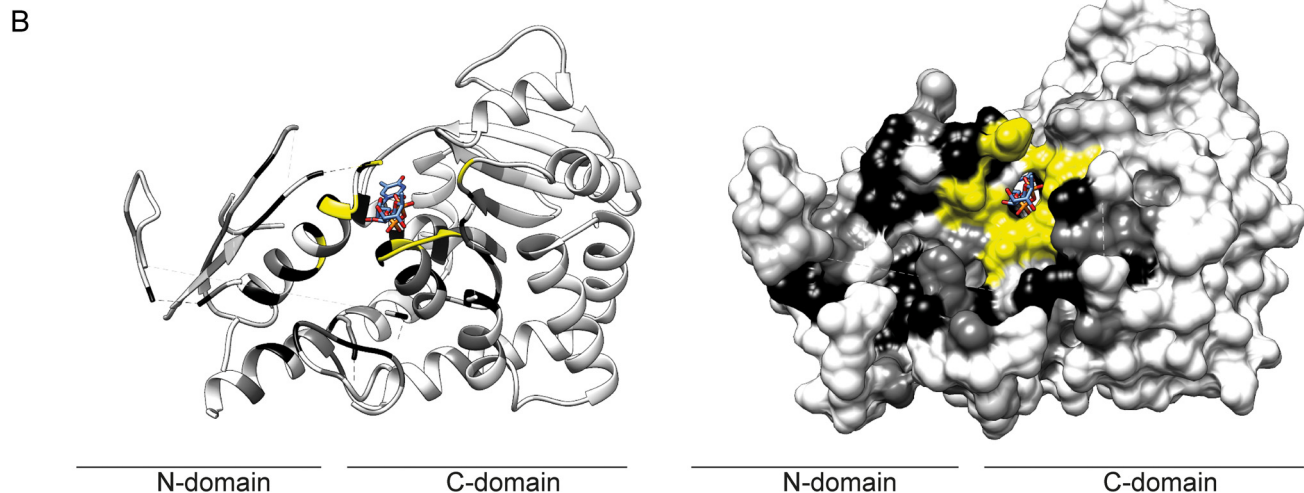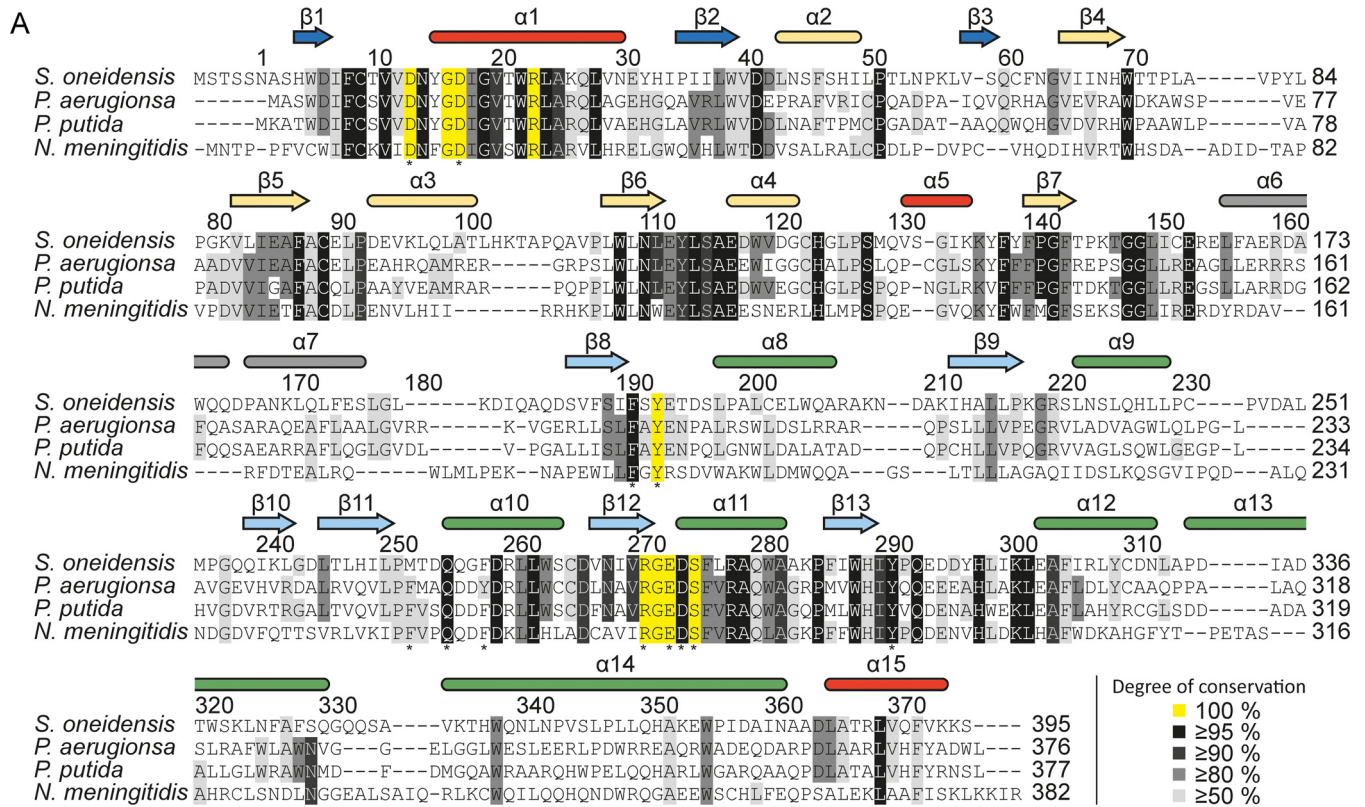
FIG 4 Evolutionary conservation of amino acids in EarP homologues. (A) Multiple-sequence alignment of EarP proteins from *Shewanella oneidensis*, *P. aeruginosa*, *P. putida*, and *Neisseria meningitidis* as a selection from the alignment of 432 protein sequences that were collected from the NCBI database (Data Set S3). The multiple-sequence alignment was generated using Clustal Omega (104). Secondary-structure elements of EarP are shown and based on the EarP$_{Ppu}$ crystal structure, NMR analysis, and predictions by MINNOU (105). α-Helices are in red and green for the N- and C-domains, respectively, and β-strands are in blue and cyan. The bipartite helix of the linker domain is grey. Helices and β-strands not resolved in the crystal structure are yellow. Amino acids selected for mutational analysis are indicated by asterisks. (B) The EarP$_{Ppu}$ crystal structure was colored according to the degree of conservation of the respective amino acids. Ribbon (left) and surface (right) representations of the EarP$_{Ppu}$ crystal structure are shown. Colors indicate the following: yellow, 100%; black, ≥95%; dark grey, ≥90%; light grey, ≥50%; and white, <50% identical residues in all analyzed EarP orthologues. Illustrations were generated with UCSF Chimera (82).

(see Materials and Methods) (Fig. S1A). This in turn enabled the quantification of rhamnosylation rates of EF-P$_{Ppu}$ by Western blot analysis (Fig. 5C and D). In a first step, the K_m and k_{cat} of WT^{EarP} were determined to be 53 µM and 35 min^{-1}, respectively (Fig. 5B, C, and D).

We wondered whether this K_m makes sense physiologically and therefore analyzed the cellular TDP-Rha levels in *P. putida*, *P. aeruginosa*, and *E. coli*, which were 3.5 mM,
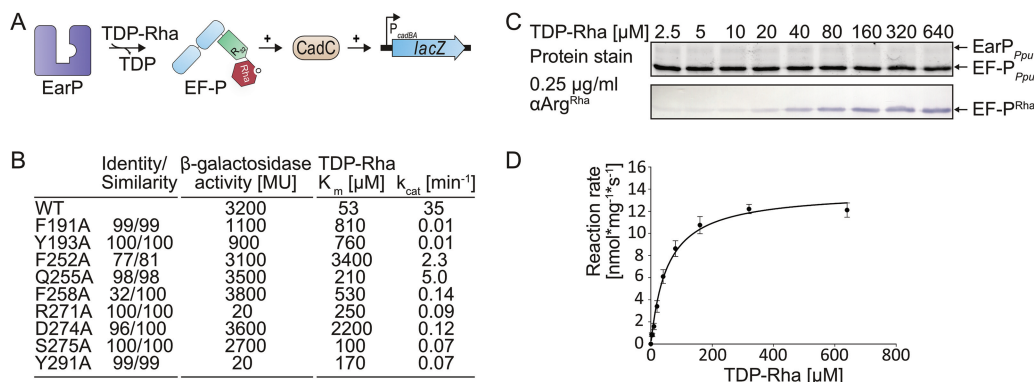
FIG 5 Analysis of kinetic parameters and *in vivo* activities of EarP$_{Ppu}$ variants. (A) Molecular principle of the *in vivo* EF-P$_{Ppu}$ functionality assay. This assay is based on the lysine decarboxylase acid stress response of *E. coli*, the CadABC module (68). At low pH and with the concomitant presence of lysine, the transcriptional activator CadC activates the promoter of its two downstream genes (P$_{cadBA}$) and with this induces expression of *lacZ* in an *E. coli* MG1655 P$_{cadBA}$::*lacZ* strain (11). Proper translation of CadC is dependent on the presence of EF-P and its corresponding modification system, and thus β-galactosidase activity can be taken as a direct readout for EF-P and EarP functionality. (B) Degree of conservation (identity/similarity) in percent, *in vivo* activities, and kinetic parameters of tested single-amino-acid exchange variants of EarP$_{Ppu}$. *In vivo* EarP$_{Ppu}$ activities were determined by measuring the β-galactosidase activities of an *E. coli* MG1655 P$_{cadBA}$::*lacZ* Δ*efp* strain heterologously expressing *efp*$_{Ppu}$ together with wild-type or mutant *earP*$_{Ppu}$ genes from o/n cultures in LB (pH 5.8). Background corrected mean values from three independent measurements are shown. Standard deviations were determined from three independent experiments to be ≤10%; the K_m and k_{cat} of wild-type EarP$_{Ppu}$ (WT^{EarP}) and variants with single-amino-acid substitutions are given in micromolar concentrations and per minute, respectively. Standard errors were determined by SigmaPlot to be <20%. (C, top) 2,2,2-Trichlorethanol (TCE) protein stain (75) of a representative SDS gel used for determination of kinetic parameters. Fixed amounts of EF-P$_{Ppu}$ (2.5 μM) and WT^{EarP} (0.1 μM) were incubated with various concentrations of TDP-Rha for 20 s and subjected to SDS-PAGE. (Bottom) Detection of rhamnosylated EF-P$_{Ppu}$. EF-P$_{Ppu}$ was visualized after Western blotting using 0.25 μg/ml anti-Arg^{Rha}. (D) TDP-Rha saturation curve of WT^{EarP}. Band intensities from panel C were quantified using ImageJ (76). Reaction rates were calculated as means of four independent measurements. Standard deviations are shown as error bars for each concentration.

2.0 mM, and 4.0 mM, respectively (see Materials and Methods and Fig. S6). In good accordance with our measurements, the physiological TDP-Rha concentration in *Lactococcus lactis* was previously determined to be as high as 1 mM (43). Thus, within a bacterial cell, the donor substrate reaches saturating concentrations, according to the WT^{EarP} K_m measurements.

Next, the K_m and k_{cat} of EarP$_{Ppu}$ substitution variants were determined and compared to those of the wild-type protein. Strikingly, all *earP* mutations affected enzymatic activity (Fig. 5B and S2B). Depending on the substituted residue, the K_m increased up to 60-fold for the F252A^{EarP} variant (K_m = 3.4 mM). Conversely, the k_{cat} decreased up to 3,500 times when we measured the kinetics of the F191A^{EarP} and Y193A^{EarP} variants.

To exclude the possibility that decreased enzyme activity was due to fold disruption, selected EarP$_{Ppu}$ variants (F191A^{EarP}, Y193A^{EarP}, F252A^{EarP}, R271A^{EarP}, D274A^{EarP}, and Y291A^{EarP}) were analyzed by NMR ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) experiments (Fig. S7). All tested substitution variants showed no structural alterations from the wild-type protein, except for the D274A^{EarP} variant. The structural instability of this EarP variant might be a result of disrupting a salt bridge that is formed between the side chains of D274^{EarP} in the protein C-domain and an equally conserved arginine at position 23 (R23^{EarP}) in the protein N-domain (Fig. 4). This salt bridge might be of importance in clamping both EarP domains together, and a lack of it might therefore destabilize the protein. Indeed, further purification of the D274A^{EarP} variant by size exclusion chromatography (SEC) revealed an elution pattern with three distinct EarP peaks, indicating a certain degree of protein aggregation. However, the lowest molecular peak in the D274A^{EarP} SEC profile is congruent with the one that we found when subjecting WT^{EarP} to SEC. Accordingly, K_m (TDP-Rha) and k_{cat} values were determined from this protein fraction to be 206 μM and 0.74 min^{-1}, respectively (Fig. 5B).

In parallel, a bacterial two-hybrid analysis (44) was set up to investigate interactions between EF-P$_{Ppu}$ and WT^{EarP} as well as the above-mentioned nine substitution variants
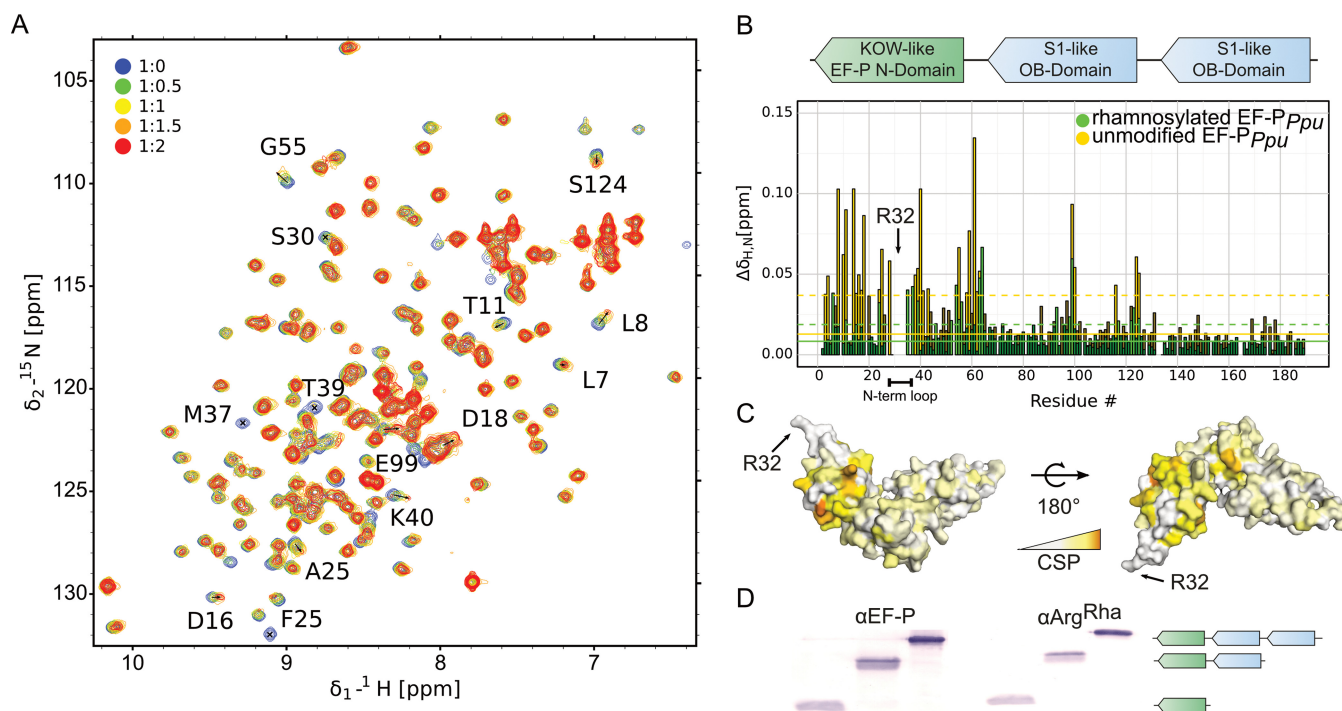
FIG 6 Interaction of EF-P$_{Ppu}$ with EarP$_{Ppu}$. (A) NMR titration of unmodified EF-P$_{Ppu}$ titrated by EarP$_{Ppu}$. Overlay of ^{1}H-^{15}N HSQC spectra of EF-P recorded at different titration steps. EF-P was titrated in a 1:2 EF-P$_{Ppu}$/EarP$_{Ppu}$ molar ratio. Color coding for respective titration steps is indicated in the upper left corner. Examples of peaks with high chemical-shift perturbations (CSPs) or severe line broadening are shown by labels indicating the assignment of given peaks. (B, top) Domain structure of EF-P. EF-P consists of three β-barrel domains. The KOW-like EF-P N-domain harbors the rhamnosylation target R32^{EF-P}. (Bottom) CSPs of EF-P$_{Ppu}$ titrated by EarP$_{Ppu}$ derived from panel A. Unmodified and rhamnosylated EF-P$_{Ppu}$ proteins were titrated by EarP$_{Ppu}$ to a 1:2 EF-P$_{Ppu}$/EarP$_{Ppu}$ molar ratio. To analyze the interaction, CSPs were calculated as described in Materials and Methods and plotted against residue numbers. Color coding is indicated in the upper right corner. Full lines indicate median CSPs, dashed lines indicate median CSPs plus standard deviations, and residues with CSPs higher than the median plus standard deviation are shown in brighter shades of the colors. The N-terminal loop containing rhamnosylation target R32^{EF-P} is indicated. (C) CSPs of unmodified EF-P$_{Ppu}$ titrated by EarP$_{Ppu}$ plotted on the model of EF-P from *P. aeruginosa* (45) (PDB accession number 3OYY) using a white-to-orange gradient, where white represents the weakest CSP and orange depicts the strongest CSP. The position of R32^{EF-P} is indicated. (D) Rhamnosylation experiments using full-length EF-P$_{Ppu}$ and C-terminally truncated variants (EF-P$_{Ppu}$ with aa 1 to 128, EF-P$_{Ppu}$ with aa 1 to 65). EF-P was detected using 0.2 μg/ml anti-EF-P. Rhamnosylation of purified protein was detected using 0.25 μg/ml anti-Arg^{Rha}. The domain structure of the respective protein variants is indicated as in panel B.

(Fig. 5B). Therefore, fusions were generated with two complementary fragments, T25 and T18, encoding segments of the catalytic domain of the *Bordetella pertussis* adenylate cyclase CyaA. If EF-P$_{Ppu}$ and WT^{EarP} do interact, then CyaA is reconstituted, which in turn allows induction of the *lac* promoter and results in *lacZ* expression. Accordingly, β-galactosidase activity is a measure of the interaction strength. When coproducing EF-P$_{Ppu}$ with WT^{EarP}, we determined ca. 250 MU, whereas combinations with solely T25 and T18 resulted in <60 MU, thus defining the threshold of the assay (Fig. S1C). Except for the R271A^{EarP} and Y291A^{EarP} proteins, all other variants were below this threshold, indicating that alterations in the donor binding site might also affect acceptor binding (Fig. S1C).

The KOW-like EF-P N-domain is sufficient for EarP-mediated rhamnosylation.
To test which part of EF-P is involved in the interaction with EarP, NMR chemical shift perturbation experiments were performed by comparing ^{1}H-^{15}N HSQC results between unbound EF-P$_{Ppu}$ and EarP$_{Ppu}$-bound EF-P$_{Ppu}$ (Fig. 6A). Triple-resonance experiments of EF-P$_{Ppu}$ enabled backbone assignment, with a sequence coverage of 97%. Missing assignments are for residues S123, R133, N140, V164, D175, and G185. The assignment also enabled secondary-structure determination from secondary chemical shifts and confirmed the validity of the EF-P model for *P. putida*, based on the crystal structure of *P. aeruginosa* EF-P (Fig. S3E) (45). The titration experiment showed clear chemical shift perturbations in the N-terminal acceptor domain of EF-P$_{Ppu}$ (Fig. 6B and C). However, R32^{EF-P} and residues surrounding the rhamnosylation site (e.g., S30^{EF-P}, G31^{EF-P}, R32^{EF-P},

N33^{EF-P}) are severely line broadened beyond detection. Therefore, chemical shift perturbation values cannot be determined for these and vicinal residues. This line broadening is an indication that they are bound by EarP$_{Ppu}$ and thus have rotational correlation times expected for a complex of that size. Several residues located in the S1-like OB-domain are also slightly affected. However, this is not necessarily due to direct contacts with EarP$_{Ppu}$ but might also be propagating effects. Therefore, we also investigated *in vitro* rhamnosylation of truncated EF-P$_{Ppu}$ variants comprising either amino acids 1 to 128 or amino acids 1 to 65 (Fig. 6D). Both truncations were readily rhamnosylated by EarP$_{Ppu}$, further corroborating that EF-P contact sites are predominantly located in the KOW-like N-domain.

In addition, we compared NMR interactions between EarP$_{Ppu}$ and unmodified EF-P$_{Ppu}$ or rhamnosylated EF-P$_{Ppu}$. This experiment clearly showed that chemical shift perturbations for unmodified EF-P are stronger than for rhamnosylated EF-P (Fig. 6B). Thus, EarP releases EF-P after rhamnosylation due to decreased affinity, while unmodified EF-P binds with higher affinity to enable efficient posttranslational modification.

Mutational analysis of the three invariant EarP residues D13, D17, and E273. We and others previously showed that EarP inverts the anomeric configuration on the sugar moiety from TDP-β-L-rhamnose to α-rhamnosyl arginine (26, 27). Reportedly, inverting glycosyltransferases employ a direct-displacement S$_N$2-like reaction (46). The molecular basis for inverted N-linked glycosylation was elucidated for the oligosaccharyl transferase PglB (47). Here the catalytic site features three acidic side chains (29). As with PglB, three negatively charged residues—aspartates D13^{EarP} and D17^{EarP} and glutamate E273^{EarP}—were identified as potential candidates to catalyze the glycosylation reaction (Fig. 3B). All three residues are invariant in all EarP orthologues (Fig. 4A; Data Set S3). Moreover, the D13^{EarP} and D17^{EarP} variants as well as the E273^{EarP} variant are in the vicinity of the rhamnose moiety and might therefore be proximal to the putative active center and R32 of EF-P (Fig. 3B). The distances of these three residues to rhamnose atoms range from 2.5 to 4.5 Å (the carboxyl group of D13 is the closest, with a distance of 2.5 Å to the methyl group of the rhamnose, followed by the side chains of D17 and E273, with distances of 3.9 and 4.5 Å to the hydroxyl group of C4 and C2, respectively). Consequently, we constructed the corresponding alanine substitution variants D13A^{EarP}, D17A^{EarP}, and E273A^{EarP} and investigated their enzymatic activities *in vitro*. In line with the idea that these residues might be involved in catalysis, EF-P rhamnosylation could not be detected even after 8 h of incubation, and accordingly these EarP variants are inactive (Fig. 7A).

To exclude misfolding being causative for the nonfunctional EarP$_{Ppu}$ protein variants, ^{15}N HSQCs were measured for D13A^{EarP}, D17A^{EarP}, and E273A^{EarP}. The spectra show no structural alterations from WT^{EarP} (Fig. 7B, C, and D and see Fig. S7). Additionally, the variants D13A^{EarP} and D17A^{EarP} were titrated with TDP-Rha being indistinguishable from WT^{EarP} perturbations. Interestingly, although D13^{EarP} and D17^{EarP} resonances could not be assigned, other residues in close proximity (G16^{EarP} and G19^{EarP}) exhibited strong perturbations not only in WT^{EarP} but also in the D13A^{EarP} and D17A^{EarP} variants upon TDP-Rha binding, despite not forming direct ligand contacts (Fig. 7E). Similarly, we could measure TDP-Rha binding for E273A/D/N^{EarP} variants using STD NMR (Fig. S5C). This confirms that these mutations do not affect donor substrate binding.

To investigate interactions between EF-P$_{Ppu}$ and the D13A^{EarP}, D17A^{EarP}, and E273A^{EarP} variants, we again performed a bacterial two-hybrid analysis and were able to show that all substitution variants are capable of acceptor binding, demonstrated by a blue colony on X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside)-containing LB plates (Fig. 7F, S1C).

To further corroborate our findings on the *in vitro*-inactive D13A^{EarP}, D17A^{EarP}, and E273A^{EarP} variants, they were subjected to an *in vivo* experiment in which we investigated their ability to activate EF-P$_{Ppu}$ (Fig. 5A). Additional substitutions—D13N/E^{EarP}, D17N/E^{EarP}, and E273Q/D^{EarP}—were also included in the study. Expectedly, coproduction of the D13A^{EarP}, D17A^{EarP}, and E273A^{EarP} variants with EF-P$_{Ppu}$ phenocopies Δ*efp*
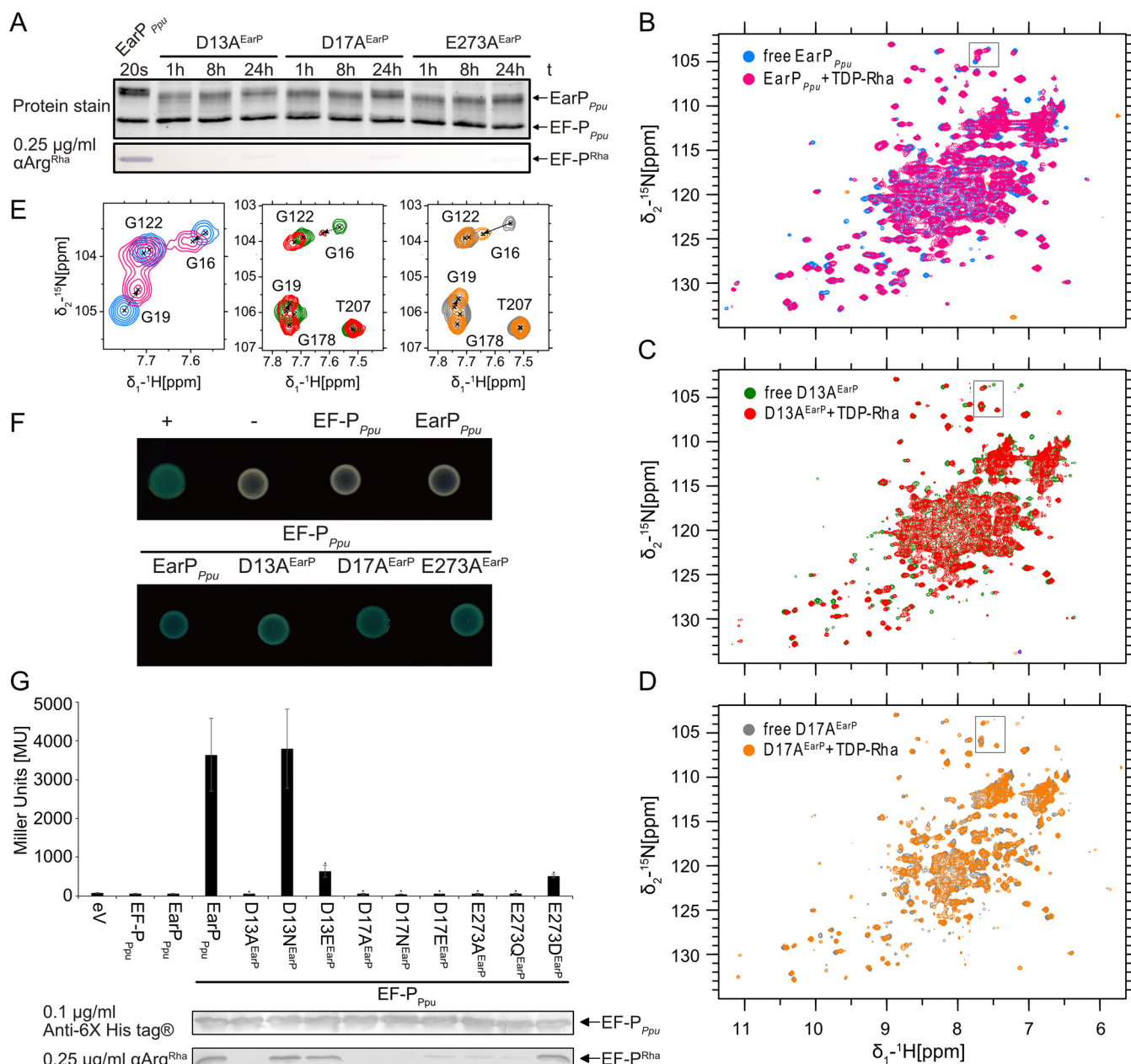
FIG 7 Mutational analysis of the three invariant EarP residues D13, D17, and E273. (A) *In vitro* rhamnosylation of EF-P$_{Ppu}$ by single-amino-acid exchange variants, specifically, D13A^{EarP}, D17A^{EarP}, and E273A^{EarP} variants. EF-P (2.5 μM) and TDP-Rha (1 mM) were incubated together with the EarP$_{Ppu}$ variants (0.5 μM) and sampled at different time points. Rhamnosylated EF-P$_{Ppu}$ (EF-P^{Rha}) was detected after Western blotting using 0.25 μg/ml anti-Arg^{Rha}. (B) Overlay of ^{1}H-^{15}N HSQC spectra of wild-type EarP$_{Ppu}$ that was free and bound to TDP-Rha. (C) Overlay of ^{1}H-^{15}N HSQC spectra of the free and TDP-Rha-bound D13A^{EarP} variant. (D) Overlay of ^{1}H-^{15}N HSQC spectra of the free and TDP-Rha-bound D17A^{EarP} variant. The color coding is indicated in the upper left corner of each spectrum. The titrations are described in detail in Materials and Methods. (E) Zoom into the overlaid spectra shown in panels B, C, and D. The position of the zoom is indicated by a black frame in the corresponding original overlay. Peak assignments are shown. The movement of G16 and G19 upon TDP-Rha titration is indicated by dashed arrows. (F) Bacterial two-hybrid analysis of protein-protein interactions between EarP$_{Ppu}$, the D13A^{EarP}, D17A^{EarP}, and E273A^{EarP} variants, and the protein acceptor EF-P$_{Ppu}$ in *E. coli* BTH101. The blue color of colonies results from cleavage of X-Gal by β-galactosidase and indicates protein-protein interaction between coexpressed hybrids. (G, top) Analysis of *in vivo* activities of EarP$_{Ppu}$, D13A^{EarP}, D17A^{EarP}, and E273A^{EarP}. *In vivo* EarP$_{Ppu}$ activities were determined by measuring the β-galactosidase activities of an *E. coli* MG1655 P$_{cadBA}$::*lacZ* Δefp strain heterologously expressing *efp*$_{Ppu}$ together with the wild-type or mutant *earP*$_{Ppu}$ genes from o/n cultures in LB (pH 5.8). Means of three independent measurements are shown. Standard deviations from three independent experiments were determined. (Bottom) Western blot analysis of o/n cultures of *E. coli* MG1655 P$_{cadBA}$::*lacZ* Δefp heterologously expressing *efp*$_{Ppu}$ together with the wild-type or *earP*$_{Ppu}$ mutants. Rhamnosylated EF-P$_{Ppu}$ (EF-P^{Rha}) was detected using 0.25 μg/ml anti-Arg^{Rha}.

with respect to P$_{cadBA}$ activation and *in vivo* rhamnosylation (Fig. 7G; Fig. S1B). Similar results were obtained with the D17N/E^{EarP} and E273Q^{EarP} variants, whereas the D13E^{EarP} and E273D^{EarP} variants were drastically impaired in function, although they retained some residual activity. Their impairment is indicated by a certain degree of

P*cadBA* activation as well as a band in the *in vivo* rhamnosylation blot (Fig. 7G; Fig. S1B). In contrast, a variant with a change of D13 to asparagine was indistinguishable from WT^{EarP}, implying an importance of the chain length over charge.

Our thorough analysis of these EarP variants suggests that they are promising candidates to be involved in catalysis.

DISCUSSION

Activation of the proline-specific translation elongation factors EF-P and IF-5A is usually achieved by posttranslational elongation of the ε-amino group of a conserved lysine (20–23, 48, 49). The resultant noncanonical amino acids—β-lysinyl-hydroxylysine, hypusine, and 5-amino-pentanolyl-lysine—appear to be chemically and structurally analogous. We recently showed that in a subset of bacteria, a so-far-unappreciated form of posttranslational modification plays an important role in the activation of EF-P. Here, instead of lysine, the guanidine group of a conserved arginine is modified with a rhamnose moiety by a glycosyltransferase termed EarP (17). This type of modification not only contrasts with the other known EF-P/IF-5A activation strategies but is also one of only two reported cases of enzyme-mediated arginine glycosylation. In canonical N-linked glycosylation, the sugar is attached to the amide nitrogen of an asparagine in an N-X-S/T consensus sequence (X is any amino acid except for a proline) (46, 50). In contrast, the effector glycosyltransferase NleB of enteropathogenic *E. coli* *N*-acetyl-glucosaminylates (GlcNAc) specifically the arginines at positions 117 and 235 in the death domain-containing proteins FADD and TRADD, respectively (31, 51). This in turn antagonizes the apoptosis of infected cells, thereby blocking a major antimicrobial host response. Notably, EarP shows neither sequential nor structural homologies to the GT-A-type glycosyltransferase NleB, and thus the arginine glycosylation of death domains and EF-P are examples of convergent evolution. Instead EarP seems to be structurally related to MurG. Moreover, and despite the lack of a significant overall sequence similarity, certain residues important for function remain the same. According to these facts, one might speculate that EarP is not simply analogous to MurG but a distinct homologue. Note that MurG is essential for cell wall biosynthesis in both Gram-negative and Gram-positive bacteria, and due to its degree of conservation, it is most likely more ancient then EarP. Although there is no real evidence for this, one might hypothesize about the possibility of a duplication of MurG in a betaproteobacterial progenitor, which is the presumed origin of EarP (17). Subsequently, the sequences of both proteins more and more diverged in consequence of distinct donor and acceptor substrates. This assumption is at least also in line with the theory that NleB (GT-A type) and EarP (GT-B type) are phylogenetically nonrelated enzymes. Accordingly, one can also assume that the molecular mechanisms of the glycosyl transfer reactions in both arginine glycosyltransferases differ. In 2016, Wong Fok Lung and coworkers mutated *nleB* and identified certain residues in NleB either interfering with FADD binding or preventing GlcNAcylation (52). They confirmed the importance of two invariant aspartate residues, D221 and D223, from among the nonfunctional NleB protein variants (31). A catalytic Asp-X-Asp motif is featured by various GT-A glycosyltransferases. Here, the two negatively charged aspartate side chains coordinate a divalent cation that facilitates departure of the nucleoside phosphate. Negatively charged amino acids also play important catalytic roles in inverting GT-B glycosyltransferases (46). In the case of the metal-independent fucosyltransferase FucT (53), for example, the side chain carboxyl groups of D13 and E95 may work as base catalysts (46). Also, the activation of the acceptor amide nitrogen by the lipid donor utilizing bacterial oligosaccharyltransferase PglB depends on the two negatively charged amino acids D56 and E319. These residues abolish the conjugation of the nitrogen electrons and allow the positioning of a free electron pair for the nucleophilic attack onto the anomeric center of the donor substrate (29, 47). Analogously, the invariant negatively charged residues D13^{EarP}, D17^{EarP}, and E273^{EarP} in the EarP glycosyltransferase family might play a role in activating the R32 guanidino group of EF-P. Especially D17^{EarP} and E273^{EarP}—both in close proximity to each other—may form a catalytic dyad (Fig. 3B).

While activation of the acceptor substrate might be driven by the essential amino acids D13^{EarP}, D17^{EarP}, and E273^{EarP}, the nucleotide sugar donor TDP-Rha is bound in a highly conserved cavity of the protein C domain. A cocrystal structure of the putative structural EarP analogue MurG$_{Eco}$ with its cognate substrate reveals that aromatic amino acid side chains play important roles in UDP binding (PDB accession number 1NLM) (54). Similar interactions were reported for the protein O-fucosyltransferase POFUT1 (PDB accession number 3ZY6), where F357 is involved in π-stacking with the respective nucleobase (55). Stacking interactions also play a role in EarP, in which the aromatic side chains of F252^{EarP} and F258^{EarP} bind the thymine and ribose moiety of TDP-Rha, respectively. In contrast, contacts with the ribose or the phosphate moieties frequently occur via interactions with side chain amines, hydroxyl groups, and backbone amides (37, 54, 55). Accordingly, this is also the case for EarP.

In GT-B glycosyltransferases, positively charged amino acids are often involved in facilitating leaving group departure. This is achieved by neutralization of evolving negative charges on the phosphate moiety during the glycosyl transfer reaction, as described, e.g., for R261 of MurG$_{Eco}$ (PDB accession number 1F0K) (37). Notably, earP$_{Ppu}$ encodes an invariant R271^{EarP} in the equivalent position and a substitution to alanine (R271A^{EarP}) strongly impairs protein function, all of which suggests that they have similar roles in product stabilization.

In GT-B glycosyltransferases, the two Rossmann folds can generally be divided into one donor and one acceptor substrate binding domain (40). As with other glycosyltransferases, the nucleotide sugar is bound by the protein C-domain of EarP. Accordingly, it is worth assuming important binding sites for EF-P in the protein N-domain. Conversely, EF-P presumably contacts EarP by amino acids that are in close proximity to the glycosylation site R32^{EF-P}. In agreement with this hypothesis, the EF-P β-lysine ligase EpmA, for example, recognizes EF-P via identity elements in a region located around the E. coli EF-P modification site K34 (21, 22, 56). Along the same line, the deoxyhypusine synthase (DHS) can efficiently modify a human eIF-5A fragment comprising only the first 90 amino acids of the protein (57). Similarly, we could show that the KOW-like N-terminal domain of EF-P (Fig. 6B) is sufficient to be glycosylated by EarP (Fig. 6D), being congruent with the NMR titrations of EF-P with EarP (Fig. 6A to C). Upon titration with EarP, the chemical shift perturbations observed were (with a few exceptions) restricted to the first 65 residues.

Taking all of this together, we propose a three-step model for the rhamnosylation of EF-P by its cognate modifier EarP. In the ground state, both the nucleotide sugar binding site in the C-domain and the putative acceptor binding site in the N-domain are unoccupied.

In the donor-bound state, TDP-Rha is coordinated within a highly conserved cavity in the protein C-domain, including an aromatic pocket that surrounds the thymine ring (Fig. 3). Previous studies showed that binding of the donor substrate induces structural alterations in both the N and C-domains of glycosyltransferases (40, 58, 59). In MurG, these rearrangements include rotation of F244, which stacks over the nucleobase to cap the donor binding pocket (37). Notably, in the crystal structure of EarP, a phenylalanine, F252, is in the equivalent position, indicating that this capping interaction is conserved (Fig. 3A) (54).

In the catalytic state, the R32 guanidino group of EF-P might be activated by a mechanism analogous to the one that was reported for the oligosaccharyltransferase PglB (47). Hence, in the EF-P rhamnosylation reaction, R271^{EarP} might stabilize the nucleotide product, thereby facilitating leaving group departure. Upon successful inverting glycosyl transfer from TDP-Rha to R32^{EF-P}, presumably by a single S_N2 displacement reaction, the products are released from the active site of EarP, in turn reverting to the unbound ground state.

We point out that there is most likely no strict sequence of binding events, as NMR measurements demonstrate that EarP can interact with either substrate independently.

Altogether, our structural and biochemical investigation of EarP provides first insights into arginine glycosylation and improves our general understanding of

N-linked glycosyl transfer reactions. Additionally, our research might open up new avenues for the development of antimicrobial drugs in order to fight, e.g., *P. aeruginosa* infections.

MATERIALS AND METHODS

Bacterial strains and growth conditions. Strains and plasmids used in this study are listed in Data Set S1 in the supplemental material. *P. putida* and *E. coli* were routinely grown in lysogeny broth (LB) (60, 61) according to the Miller modification (62) at 30°C (for *P. putida*) and 37°C (for *E. coli*), unless indicated otherwise. When required, media were solidified by using 1.5% (wt/vol) agar. If necessary, media were supplemented with 50 μg/ml chloramphenicol, 100 μg/ml kanamycin sulfate, and/or 100 μg/ml ampicillin sodium salt. For promoter induction from P_{BAD}-containing plasmids (63), L-arabinose was added to a final concentration of 0.2% (wt/vol) in liquid medium. For promoter induction from plasmids comprising the *lac* operator sequences, isopropyl β-D-1-thiogalactopyranoside (IPTG) (Sigma-Aldrich) was added to a final concentration of 1 mM.

Molecular biology methods. Enzymes and kits were used according to the manufacturers' directions. Genomic DNA was obtained according to the protocol of Pospiech and Neumann (64), and 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). Restriction endonucleases were purchased from New England Biolabs (NEB). Sequence amplifications by PCR were performed utilizing the Q5 high-fidelity DNA polymerase (NEB) or the OneTaq DNA polymerase (NEB). Mutations were introduced into the *earP* gene by overlap extension PCR (65, 66). Oligonucleotides used in this study are listed in Data Set S1. All constructs were analyzed by Sanger sequencing (LMU Sequencing Service). Standard methods were performed according to the instructions of Sambrook and Russel (67).

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 68 in biological triplicates and are given in Miller units (MU) (69). The significance of the results was determined by applying a two-sided Student *t* test and stating a result as significantly different if P was <0.05.

Bacterial two-hybrid analysis. Protein-protein interactions were detected using the bacterial adenylate cyclase two-hybrid system kit (Euromedex) according to the product manuals. Chemically competent (70) *E. coli* BTH101 cells were cotransformed with pUT18C-*efp_{Ppu}* and/or the respective pKT25 variants (pKT25-*earP*, pKT25-D13A, pKT25-D17A, pKT25-F191A, pKT25-Y193A, pKT25-F252A, pKT25-Q255A, pKT25-F258A, pKT25-R271A, pKT25-D274A, pKT25-S275A, pKT25-R278A, pKT25-Y291A, pKT25-E273A) and plated on LB screening medium containing 40 μg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) and 0.5 mM IPTG as well as 50 μg/ml kanamycin sulfate and 100 μg/ml ampicillin sodium salt. Transformants containing pUT18-*zip* and pKT25-*zip* were used as positive controls. Transformants carrying pUT18C and pKT25 vector backbones were used as negative controls. Bacteria expressing interacting protein hybrids exhibit a blue phenotype on screening plates due to functional complementation of the CyaA fragments (T18 and T25). After 48 h of incubation at 30°C, plates containing around 100 colonies were evaluated. Representative colonies were transferred to liquid LB cultures containing kanamycin sulfate and ampicillin sodium salt and incubated o/n at 30°C. Subsequently, 2 μl of the o/n culture were spotted on LB X-Gal–IPTG screening plates. Pictures were taken after 48 h of cultivation at 30°C.

For quantification of interaction strength, which corresponds to the β-galactosidase activity, cells were inoculated in 1.5 ml LB medium containing 0.5 mM IPTG as well as 50 μg/ml kanamycin sulfate and 100 μg/ml ampicillin sodium salt. After incubation in 2-ml reaction tubes under microaerobic conditions at 30°C for 42 h, cells were harvested and β-galactosidase activities were determined as described above.

Protein purification. C-terminally His$_6$-tagged EarP$_{Ppu}$ variants (pBAD33-*earP_{Ppu}*) were overproduced in *E. coli* LMG194 by addition of 0.2% arabinose to exponentially growing cells and subsequent cultivation at 18°C o/n. N-terminally His$_6$-tagged EarP (pACYC-DUET-*earP_{Ppu}*) and His$_6$-SUMO-tagged EF-P$_{Ppu}$ (pET-SUMO-*efp_{Ppu}*) were overproduced in *E. coli* BL21(DE3) by addition of 1 mM IPTG to exponentially growing cells. Subsequently, cells were incubated at 18°C overnight. Rhamnosylated EF-P$_{Ppu}$ (EF-P^{Rha}) was produced by cooverproduction with His$_6$-tagged EarP$_{Ppu}$. Cells were lysed by sonication, and His$_6$-tagged proteins were purified using Ni-nitrilotriacetic acid (Ni-NTA; Qiagen) according to the manufacturer's instructions. The His$_6$-SUMO tag was removed by incubation with 1 μ/mg His$_6$-Ulp1 (71) overnight. Subsequently, tag-free EF-P$_{Ppu}$ was collected from the flowthrough after metal chelate affinity chromatography. For biochemical analyses, cells were cultivated in LB. For use in NMR spectroscopy, cells were grown in M9 minimal medium (62). If necessary, ^{15}N-labeled nitrogen (^{15}NH$_4$Cl) and ^{13}C-labeled glucose were used. For NMR backbone assignment of EarP$_{Ppu}$, additionally 99.8%-pure heavy water D$_2$O (Sigma-Aldrich) was used instead of H$_2$O in growth medium to allow partial deuteration of the protein in order to reduce cross-relaxation effects and increase the signal-to-noise ratio. Size exclusion chromatography of EarP$_{Ppu}$ and the D274A^{EarP} variant was performed in 100 mM NaP$_i$ (pH 7.6) 50 mM NaCl using a Superdex 200 Increase 10/300-Gl column with a flow rate of 0.3 ml/min on an Äkta purifier (GE Healthcare). Four milligrams of protein was loaded in a volume of 0.5 ml (8 mg/ml). Eluting protein was detected at 280 nm. Fractions of 0.5 ml were collected.

For the production of selenomethylated EarP$_{Ppu}$, *E. coli* BL21(DE3) cells expressing N-terminally His$_6$-tagged EarP$_{Ppu}$ were cultivated in 1 liter M9 minimal medium at 37°C to an optical density at 600
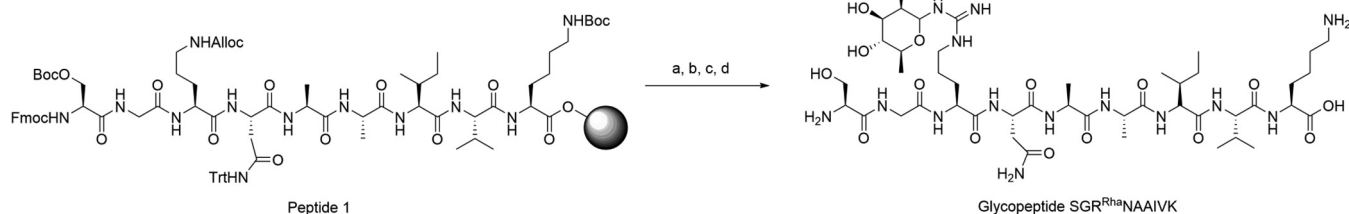
FIG 8 Synthesis of glycopeptide SGR^{Rha}NAAIVK. a, SiPhH₃ (phenylsilane), Pd(PPh₃)₄, CH₂Cl₂; b, 1-(*tert*-butoxycarbonyl)-3-(2,3,4-tri-*O*-acetyl-6-deoxy-L-mannopyranos-1-yl)-2-ethyl-isothiourea (26) AgNO₃ (silver nitrate), NEt₃ (triethylamine), DMF; c, N₂H₄·H₂O (5% solution in DMF); d, TFA-H₂O-phenol-TIPS (88/5/5/2).

nm (OD$_{600}$) of 0.6. One hundred micrograms of threonine, 100 μg lysine, and 50 μg isoleucine were added to feedback inhibit methionine biosynthesis (72). Additionally, 50 μg L-(+)-selenomethionine was added 15 min prior to induction. Protein production was induced by addition of 1 mM IPTG, and cells were incubated at 18°C overnight. Protein concentrations were determined as described by Bradford (73). For biochemical analyses, EarP$_{Ppu}$ and EF-P$_{Ppu}$ were dialyzed against 100 mM NaP$_i$, pH 7.6, 5 mM dithiothreitol (DTT), whereas a buffer composed of 100 mM NaP$_i$, pH 7.6, 50 mM NaCl, and 5 mM DTT was used when the proteins were subjected to NMR analysis.

Synthesis of a single rhamnosyl-arginine containing glycopeptide. Moisture- and air-sensitive reactions were conducted in flame-dried glassware under an argon atmosphere. Commercially available reagents and solvents were used without further purification. CH$_2$Cl$_2$ was distilled from calcium hydride, and tetrahydrofuran (THF) was distilled from sodium benzophenone immediately prior to use. Dimethylformamide (DMF) was stored under argon in a flask containing 4 Å molecular sieves. Reactions were monitored by thin layer chromatography (TLC) with precoated Silica Gel 60 F$_{254}$ aluminum plates (Merck, Darmstadt, Germany) using UV light and methoxyphenol reagent (100 ml 0.2% ethanolic methoxyphenol solution and 100 ml 2 M ethanolic sulfuric acid) as the visualizing agent. Flash chromatography was performed using silica gel (35 to 70 μm) from Acros Organics. Peptide purification by reverse-phase high-performance liquid chromatography (RP-HPLC) was performed on a JASCO purification system with a UV–visible-light detector (model UV-2075Plus) using a Phenomenex Aeris Peptide 5-μm XB-C$_{18}$ column (250 by 21.2 mm). Analytical RP-HPLC was measured on a JASCO system with a Phenomenex Aeris Peptide 5-μm XB-C$_{18}$ column (250 by 4.6 mm). In all cases, mixtures of water (eluent A) and acetonitrile (eluent B) were used as eluents; if required, 0.1% formic acid (FA) or 0.1% trifluoroacetic acid (TFA) was added. High-resolution electrospray ionization (HR-ESI) mass spectra were recorded on a Thermo Finnegan LTQ FT mass spectrometer or on a Bruker maxis apparatus equipped with a Waters ACQUITY ultrahigh-performance liquid chromatograph (UPLC) using a Kinetex C$_{18}$ column (2.6 μm, 100 Å) at 40°C (Fig. 8).

Glycopeptide SGR^{Rha}NAAIVK was synthesized using a Liberty Blue automated microwave peptide synthesizer, followed by on-resin glycosylation and deprotection (Fig. 8). For construction of peptide 1, 0.1 mmol of preloaded H-Lys(Boc)-2-chlorotrityl resin (loading concentration, 0.78 mmol/g) was applied. Cleavage of the Fmoc-protecting group was achieved with 20% piperidine in DMF (75°C, 35 W, 3 min). Fmoc-protected amino acids (5 eq) were activated for peptide coupling using 5 eq of ethyl (hydroxyimino)cyanoacetate (Oxyma Pure), 0.5 eq of *N,N*-diisopropylethylamine (DIPEA), and 5 eq of *N,N′*-diisopropylcarbodiimide. All coupling reactions were conducted at 75°C and 28 W for 5 min. Removal of the allyloxycarbonyl-protecting group and subsequent coupling of the sugar moiety, as well as deprotection of the acetyl groups, were performed according to established procedures (26). Final deprotection gave the desired glycopeptide, SGR^{Rha}NAAIVK, yielding 39% after HPLC purification. The amino acid sequence of the glycopeptide corresponds to the primary structure of the *S. oneidensis* acceptor loop, which is highly similar to the consensus sequence of EarP-arginine-type EF-Ps (17).

High-resolution mass spectrometry (HRMS) (ESI$^+$), calculated for C$_{44}$H$_{82}$N$_{14}$O$_{16}$ [M+2H]$^{2+}$, m/z = 531.3011; found, 531.3016.

HPLC (0.1% TFA, 0 min, 8% B → 45 min, 50% B; flow, 1 ml/min), t_R (retention time) = 9.61 min, λ = 204 nm (Fig. 9).

Antibody generation. Polyclonal antibodies were raised commercially by Eurogentec according to the Rabbit Speedy 28-day (AS superantigen) program. The mono-rhamnosyl-arginine-containing peptide was coupled to bovine serum albumin (BSA) according to an internal protocol (AS-PECO 05). Antibodies capable of binding to rhamnosyl-arginine were purified from rabbit sera by affinity chromatography (AS-PURI−MED) against the glycopeptide SGR^{Rha}NAAIVK. To test the specificity of the purified polyclonal antibodies toward EF-P^{Rha}, 1.5 μg of unmodified and 0.5 μg of modified EF-P were transferred to a nitrocellulose membrane by Western blotting. While polyclonal antibodies that were raised against EF-P from *S. oneidensis* detect both unmodified and modified EF-P$_{Ppu}$, anti-Arg^{Rha} specifically detects the modified protein variant (Fig. S1A).

SDS-PAGE and Western blotting. Electrophoretic separation of proteins was carried out using SDS-PAGE as described by Lämmli (74). Separated proteins were visualized in gel using 0.5% (vol/vol) 2-2-2-trichloroethanol (75) and transferred onto a nitrocellulose membrane by vertical Western blotting. Antigens were detected using 0.1 μg/ml anti-His$_6$ tag (Abcam, Inc.), 0.2 μg/ml anti-EF-P, or 0.25 μg/ml of anti-Arg^{Rha}. Primary antibodies (rabbit) were targeted by 0.2 μg/ml alkaline phosphatase-conjugated anti-rabbit IgG (H&L) (goat) antibody (Rockland). Target proteins were visualized by addition of substrate
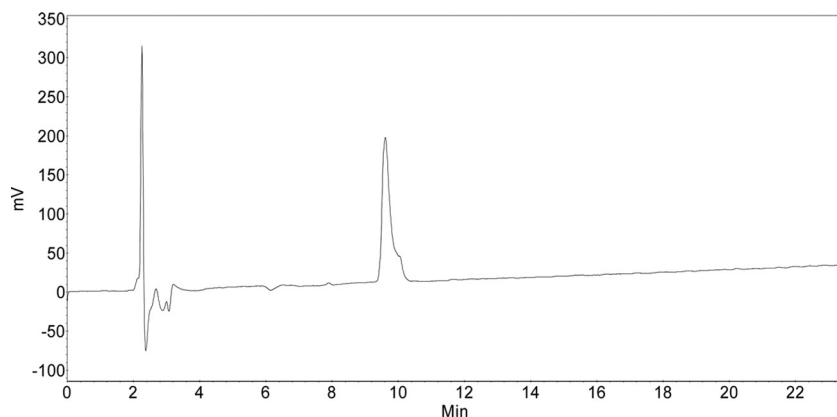
FIG 9 HPLC data.

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).

Determination of kinetic parameters. Kinetic parameters were determined by varying TDP-Rha concentrations while keeping concentrations of $EarP_{Ppu}$ (0.1 μM) and unmodified $EF-P_{Ppu}$ (2.5 μM) constant. A mixture of $EarP_{Ppu}$ and unmodified $EF-P_{Ppu}$ was equilibrated to 30°C in 100 mM NaP_i (pH 7.6). The reaction was started by the addition of TDP-Rha and was stopped after 20 s of incubation at 30°C by the addition of 1 vol of 2× Lämmli buffer (74) and incubation at 95°C for 5 min. Samples were subjected to SDS-PAGE, and rhamnosylated $EF-P_{Ppu}$ was detected as described above. Band intensities were quantified using ImageJ (76). Product formation (in nanomoles per milligram) was calculated relative to fully (*in vivo*) rhamnosylated $EF-P_{Ppu}$. K_m and k_{cat} values were determined by fitting reaction rates (in nanomoles per milligram per second) to the Michaelis-Menten equation using SigmaPlot. Time course experiments conducted at a TDP-Rha concentration of 500 μM show that the rhamnosylation reaction is not saturated after 20 s of incubation (Fig. S2A).

Fold recognition. Fold recognition models were generated using the online user interface of Phyre[2] (33, 77), SWISS-MODEL (78–81), and the I-TASSER server (34–36) as instructed on the websites. Model structures were selected from the array of results according to best confidence, Q mean, and *z* scores, respectively. All images of tertiary protein structures in this work were generated using the UCSF Chimera package developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (82). Protein structures were obtained as .pdb files from http://www.rcsb.org (83) or the respective modeling platforms mentioned above.

Determination of intracellular TDP-Rha concentrations. Cells were grown in 1 liter LB to an OD_{600} of 0.5 (5×10^8 cells/ml), harvested by centrifugation, and resuspended in 25 ml 100 mM NaP_i (pH 7.6) (2×10^{10} cells/ml). After disruption of cells with a Constant Systems Ltd. continuous-flow cabinet at 1.35 kb, cell debris were removed by centrifugation and lysates were sterilized by filtration (Steriflip). A mixture of $EarP_{Ppu}$ (0.1 μM) and unmodified $EF-P_{Ppu}$ (2.5 μM) was equilibrated to 30°C in 10 μl 100 mM NaP_i (pH 7.6). The reaction was started by addition of 10 μl lysate from ~2×10^7 or ~2×10^8 cells and stopped after 20 s of incubation at 30°C by addition of 1 vol 2× Lämmli buffer (74) and incubation at 95°C for 5 min. In parallel, a TDP-Rha calibration series was generated by addition of TDP-Rha at final concentrations ranging from 5 μM to 160 μM, including the linear range of the rhamnosylation reaction rate (Fig. 5D). Samples were subjected to SDS-PAGE, and rhamnosylated $EF-P_{Ppu}$ was detected as described above. Band intensities were quantified using ImageJ (76). TDP-Rha concentrations in samples containing lysate were calculated by dividing the respective relative band intensities by the slope of the corresponding calibration curve (5 μM to 80 μM TDP-Rha). Intracellular TDP-Rha concentrations were calculated from the amount of substance (in moles) per cell, with an assumption of equal distribution of TDP-Rha across all cells as well as an average cell volume of 3.9 μm^3 for *E. coli* (84) and 2.1 μm^3 for *P. putida* and *P. aeruginosa* (85).

NMR spectroscopy and backbone assignment of EF-P and EarP. All NMR experiments were performed at 298 K on Bruker Avance III spectrometers with a magnetic field strength corresponding to a proton Larmor frequency of 600 MHz (equipped with a Bruker TXI cryogenic probe head), 700 MHz (equipped with a Bruker room temperature probe head), or 800 MHz (equipped with a Bruker TXI cryogenic probe head). All data sets were processed using NMRPipe (91).

Before NMR measurements of ^{15}N- and ^{13}C-labeled EF-P (700 μM) in 100 mM NaP_i, 50 mM NaCl, and 5 mM DTT (pH 7.6), 0.02% NaN_3 was added to the sample. Sequential resonance assignment was obtained from two-dimensional (2D) 1H-15N HSQC and three-dimensional (3D) HNCA, CBCACONH, and HNCACB backbone experiments, using a constant time during ^{13}C evolution (86). The assignment process was assisted by CARA (http://cara.nmr.ch) and CcpNmr Analysis (63), and 98% of the backbone resonances could be assigned. Missing assignments for residues other than prolines are S123, R133, N140, V164, D175, and G185. Secondary chemical shift analysis was performed based on the difference between measured $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts and random coil chemical shifts of the same nuclei to assign a secondary structure to the EF-P sequence (Fig. S3E) and confirm the validity of the model shown in Fig. 6 (87, 88).

Due to the size of EarP (43 kDa), backbone resonance assignment was possible only for ^{2}H-, ^{15}N-, and ^{13}C-labeled samples to reduce the number of protons and thus cross-relaxation effects, which also enables efficient acquisition of backbone assignment experiments in TROSY mode (89). TROSY-HNCA, -HNCACB, and -CBCACONH experiments (90), processed by NMRPipe (91) and analyzed using CARA (http://cara.nmr.ch), enabled backbone resonance assignment of 62% of all assignable residues (excluding prolines).

The NMR titrations were always performed by adding an unlabeled interaction partner to the ^{15}N-labeled protein sample and monitoring the progress of the titration by recording ^{1}H-^{15}N HSQC. First, ^{15}N-labeled 150 μM unmodified EF-P was titrated with unlabeled EarP to a 1:2 EF-P/EarP molar ratio with intermediate steps at 1:0, 1:0.5, 1:1, and 1:1.5 EF-P/EarP molar ratios. ^{15}N-labeled 41 μM rhamnosylated EF-P was titrated with unlabeled EarP to a 1:2 EF-P/EarP molar ratio without any intermediate steps. ^{15}N-labeled 540 μM wild-type EarP was titrated with unlabeled TDP-Rha to a 1:5 EarP/TDP-Rha molar ratio with intermediate steps at 1:0, 1:0.2, 1:1, and 1:3 molar ratios. ^{15}N-labeled 186 μM D13A variant or 209 μM D17A EarP variant was titrated by the addition of TDP-Rha to an approximately 1:10 molar ratio with no intermediate steps. To analyze the EF-P/EarP and wild-type EarP/TDP-Rha ratio titration, the chemical-shift perturbations (CSPs) were calculated according to the formula $\mathrm{CSPs} = \sqrt{\Delta\sigma_H^2 + (\Delta\sigma_N \times 0.15)^2}$, where 0.15 is the weighting factor to account for nitrogen resonances generally spanning a broad frequency range.

To check proper folding of EarP variants, ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled EarP variants with the following single-amino-acid substitutions at the indicated concentrations were recorded: 209 μM D13A, 209 μM D17A, 162 μM F191A, 197 μM Y193A, 139 μM D274A, 186 μM R271A, and 162 μM Y291A.

STD NMR experiments were performed with 10 μM WT^{EarP} or mutants and either 70 μM (1:7 ratio of protein to ligand to mimic SAXS conditions) or 1 mM TDP-Rha in 100 mM potassium phosphate buffer, pH 7.5, 150 mM NaCl, 1 mM DTT, and 10% D$_2$O. The experiments were performed on a Bruker Avance III 700-MHz spectrometer equipped with a triple resonance (TXI) room temperature probe head at 277 K. Protein was saturated with 49-ms Gaussian pulses at the resonance frequency of methyl resonances at 0.592 ppm. The experimental results were collected after a total saturation time of 20 s, with 1,596 scans performed for the WT^{EarP} sample with a 100-fold excess of ligand, and after a total saturation time of 5 s, with 4,096 scans performed for the WT^{EarP} sample with a 7-fold excess of ligand. For EarP mutants, the experimental results were collected after a total saturation time of 4 s and with 128 scans.

Small-angle X-ray scattering. Thirty microliters of EarP, EarP plus TDP-rhamnose, and buffer (with and without TDP-rhamnose) were measured at 20°C at BioSAXS beamline BM29 at the European Synchrotron Radiation Facility using a 2D Pilatus detector. For each measurement, 10 frames with a 1-s exposure time per frame were recorded for each EarP and buffer sample, using an X-ray wavelength (λ) of 0.9919 Å. Measurements were performed in flow mode, where samples are pushed through a capillary at a constant flow rate to minimize radiation damage. The protein concentrations measured were 1.0, 2.0, 4.0, and 8.0 mg/ml. TDP-Rha was used in a 7:1 excess (ligand to protein). The buffer measurements were subtracted from each protein sample, and the low Q range of 1.0 mg/ml was merged with the high Q range of the 8.0-mg/ml sample, using PRIMUS (92). The merging was done due to the rising scattering density at low Q ranges for the more highly concentrated samples, indicative of aggregation. CRYSOL (93) was used to fit the back-calculated scattering densities from the crystal structure to the experimental data.

X-ray crystallography. For crystallization, N-terminally His$_6$-tagged EarP$_{Ppu}$ expressed as a seleno-methionine derivative was used. The protein was dialyzed to 50 mM Tris, 100 mM NaCl, 1 mM DTT, pH 7.6, and concentrated to 183 μM. TDP-Rha was added to a final concentration of 10 mM. The crystallization condition was 0.2 M ammonium acetate, 0.1 M bis-Tris (pH 6.0), and 27% (wt/vol) polyethylene glycol 3350. A full data set was collected at the ID29 beamline, ESRF, Grenoble, France, at a wavelength of 0.97 Å (the absorption peak for selenium) and with a 15.05% beam transmission with a 0.15° oscillation range, 0.037-s exposure time, and 2,400 frames. The space group was determined to be I4. The data set was phased using single-wavelength anomalous diffraction (SAD) by the Crank2 (94) automatic pipeline in CCP4 (95), using Afro provided by N. S. Pannu (unpublished) for substructure factor amplitude (FA) estimation, Crunch2 (96) for substructure detection, and Solomon (97) for density modification. The anomalous signal extended to a 3.4-Å resolution (in a data set with a 3-Å resolution). We could successfully find 3 Se-Met signals with an occupancy of 1 located in the C-terminal domain and 2 Se-Met signals with an occupancy of ~0.5 located in the N-terminal domain. The initial structure was built in Phenix Autobuild (98), completed with several rounds of manual model building in Coot (99), and used as the model for molecular replacement (MR) of a native data set extending to 2.3 Å. Despite our rigorous efforts in manual model building, which included extreme density modification, use of homology models to model the N-terminal domain, Rosetta modeling, and refinement strategies with different refinement software (Phenix [98], refmac [100], and CNS [101, 102] [and CNS-DEN-assisted refinement]), the structure displays an R-free of 35% at 2.3 Å, with large parts of the electron density in the N-domain not interpretable. No crystallographic pathology (twinning, anisotropy) could be identified in any of the multiple data sets that we obtained, and trying to interpret crystallographic symmetry as noncrystallographic symmetry by deliberately choosing space groups with lower symmetry (C2, P1) did not improve the density. This indicates intrinsic crystal disorder caused by the N-terminal domain adopting several conformations in different unit cells.

Accession number(s). Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession number 5NV8.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio
.01412-17.

FIG S1, TIF file, 1.8 MB.
FIG S2, TIF file, 3.6 MB.
FIG S3, TIF file, 10.7 MB.
FIG S4, TIF file, 7.9 MB.
FIG S5, TIF file, 3.2 MB.
FIG S6, TIF file, 6.9 MB.
FIG S7, TIF file, 3.1 MB.
DATA SET S1, XLSX file, 0.02 MB.
DATA SET S2, XLSX file, 0.01 MB.
DATA SET S3, XLSX file, 0.1 MB.

REFERENCES

1. Varenne S, Buc J, Lloubes R, Lazdunski C. 1984. Translation is a non-
   uniform process. Effect of tRNA availability on the rate of elongation of
   nascent polypeptide chains. J Mol Biol 180:549–576. https://doi.org/10
   .1016/0022-2836(84)90027-5.
2. Pavlov MY, Watts RE, Tan Z, Cornish VW, Ehrenberg M, Forster AC. 2009.
   Slow peptide bond formation by proline and other N-alkylamino acids
   in translation. Proc Natl Acad Sci U S A 106:50–54. https://doi.org/10
   .1073/pnas.0809211106.
3. Tanner DR, Cariello DA, Woolstenhulme CJ, Broadbent MA, Buskirk AR.
   2009. Genetic identification of nascent peptides that induce ribosome
   stalling. J Biol Chem 284:34809–34818. https://doi.org/10.1074/jbc
   .M109.039040.
4. Woolstenhulme CJ, Parajuli S, Healey DW, Valverde DP, Petersen EN,
   Starosta AL, Guydosh NR, Johnson WE, Wilson DN, Buskirk AR. 2013.
   Nascent peptides that block protein synthesis in bacteria. Proc Natl Acad
   Sci U S A 110:E878–E887. https://doi.org/10.1073/pnas.1219536110.
5. Gutierrez E, Shin BS, Woolstenhulme CJ, Kim JR, Saini P, Buskirk AR,
   Dever TE. 2013. eIF5A promotes translation of polyproline motifs. Mol
   Cell 51:35–45. https://doi.org/10.1016/j.molcel.2013.04.021.
6. Pelechano V, Alepuz P. 2017. eIF5A facilitates translation termination
   globally and promotes the elongation of many non polyproline-specific
   tripeptide sequences. Nucleic Acids Res https://doi.org/10.1093/nar/
   gkx479.
7. Schuller AP, Wu CC, Dever TE, Buskirk AR, Green R. 2017. eIF5A func-

tions globally in translation elongation and termination. Mol Cell 66: 194–205.e5. https://doi.org/10.1016/j.molcel.2017.03.003.

8. Doerfel LK, Wohlgemuth I, Kothe C, Peske F, Urlaub H, Rodnina MV. 2013. EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. Science 339:85–88. https://doi.org/10.1126/science.1229017.

9. Hersch SJ, Wang M, Zou SB, Moon KM, Foster LJ, Ibba M, Navarre WW. 2013. Divergent protein motifs direct elongation factor P-mediated translational regulation in Salmonella enterica and Escherichia coli. mBio 4:e00180-13. https://doi.org/10.1128/mBio.00180-13.

10. Peil L, Starosta AL, Lassak J, Atkinson GC, Virumäe K, Spitzer M, Tenson T, Jung K, Remme J, Wilson DN. 2013. 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. https://doi.org/10.1073/pnas.1310642110.

11. Ude S, Lassak J, Starosta AL, Kraxenberger T, Wilson DN, Jung K. 2013. Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. Science 339:82–85. https://doi.org/10.1126/science.1228985.

12. Elgamal S, Katz A, Hersch SJ, Newsom D, White P, Navarre WW, Ibba M. 2014. EF-P dependent pauses integrate proximal and distal signals during translation. PLoS Genet 10:e1004553. https://doi.org/10.1371/journal.pgen.1004553.

13. Starosta AL, Lassak J, Peil L, Atkinson GC, Virumäe K, Tenson T, Remme J, Jung K, Wilson DN. 2014. Translational stalling at polyproline stretches is modulated by the sequence context upstream of the stall site. Nucleic Acids Res 42:10711–10719. https://doi.org/10.1093/nar/gku768.

14. Woolstenhulme CJ, Guydosh NR, Green R, Buskirk AR. 2015. High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. Cell Rep 11:13–21. https://doi.org/10.1016/j.celrep.2015.03.014.

15. Hanawa-Suetsugu K, Sekine S, Sakai H, Hori-Takemoto C, Terada T, Unzai S, Tame JR, Kuramitsu S, Shirouzu M, Yokoyama S. 2004. Crystal structure of elongation factor P from Thermus thermophilus HB8. Proc Natl Acad Sci U S A 101:9595–9600. https://doi.org/10.1073/pnas.0308667101.

16. Blaha G, Stanley RE, Steitz TA. 2009. Formation of the first peptide bond: the structure of EF-P bound to the 70S ribosome. Science 325:966–970. https://doi.org/10.1126/science.1175800.

17. Lassak J, Keilhauer EC, Fürst M, Wuichet K, Gödeke J, Starosta AL, Chen JM, Søgaard-Andersen L, Rohr J, Wilson DN, Häussler S, Mann M, Jung K. 2015. Arginine-rhamnosylation as new strategy to activate translation elongation factor P. Nat Chem Biol 11:266–270. https://doi.org/10.1038/nchembio.1751.

18. Doerfel LK, Wohlgemuth I, Kubyshkin V, Starosta AL, Wilson DN, Budisa N, Rodnina MV. 2015. Entropic contribution of elongation factor P to proline positioning at the catalytic center of the ribosome. J Am Chem Soc 137:12997–13006. https://doi.org/10.1021/jacs.5b07427.

19. Lassak J, Wilson DN, Jung K. 2016. Stall no more at polyproline stretches with the translation elongation factors EF-P and IF-5A. Mol Microbiol 99:219–235. https://doi.org/10.1111/mmi.13233.

20. Bailly M, de Crécy-Lagard V. 2010. Predicting the pathway involved in post-translational modification of elongation factor P in a subset of bacterial species. Biol Direct 5:3. https://doi.org/10.1186/1745-6150-5-3.

21. Navarre WW, Zou SB, Roy H, Xie JL, Savchenko A, Singer A, Edvokimova E, Prost LR, Kumar R, Ibba M, Fang FC. 2010. PoxA, YjeK, and elongation factor P coordinately modulate virulence and drug resistance in Salmonella enterica. Mol Cell 39:209–221. https://doi.org/10.1016/j.molcel.2010.06.021.

22. Yanagisawa T, Sumida T, Ishii R, Takemoto C, Yokoyama S. 2010. A paralog of lysyl-tRNA synthetase aminoacylates a conserved lysine residue in translation elongation factor P. Nat Struct Mol Biol 17:1136–1143. https://doi.org/10.1038/nsmb.1889.

23. Peil L, Starosta AL, Virumäe K, Atkinson GC, Tenson T, Remme J, Wilson DN. 2012. Lys34 of translation elongation factor EF-P is hydroxylated by YfcM. Nat Chem Biol 8:695–697. https://doi.org/10.1038/nchembio.1001.

24. Rajkovic A, Erickson S, Witzky A, Branson OE, Seo J, Gafken PR, Frietas MA, Whitelegge JP, Faull KF, Navarre W, Darwin AJ, Ibba M. 2015. Cyclic rhamnosylated elongation factor P establishes antibiotic resistance in Pseudomonas aeruginosa. mBio 6:e00823. https://doi.org/10.1128/mBio.00823-15.

25. Yanagisawa T, Takahashi H, Suzuki T, Masuda A, Dohmae N, Yokoyama S. 2016. Neisseria meningitidis translation elongation factor P and its active-site arginine residue are essential for cell viability. PLoS One 11:e0147907. https://doi.org/10.1371/journal.pone.0147907.

26. Li X, Krafczyk R, Macošek J, Li YL, Zou Y, Simon B, Pan X, Wu QY, Yan F, Li S, Hennig J, Jung K, Lassak J, Hu HG. 2016. 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. https://doi.org/10.1039/c6sc02889f.

27. Wang S, Corcilius L, Sharp PP, Rajkovic A, Ibba M, Parker BL, Payne RJ. 2017. Synthesis of rhamnosylated arginine glycopeptides and determination of the glycosidic linkage in bacterial elongation factor P. Chem Sci 8:2296–2302. https://doi.org/10.1039/C6SC03847F.

28. Coutinho PM, Deleury E, Davies GJ, Henrissat B. 2003. An evolving hierarchical family classification for glycosyltransferases. J Mol Biol 328:307–317. https://doi.org/10.1016/S0022-2836(03)00307-3.

29. Breton C, Fournel-Gigleux S, Palcic MM. 2012. Recent structures, evolution and mechanisms of glycosyltransferases. Curr Opin Struct Biol 22:540–549. https://doi.org/10.1016/j.sbi.2012.06.007.

30. Singh DG, Lomako J, Lomako WM, Whelan WJ, Meyer HE, Serwe M, Metzger JW. 1995. β-Glucosylarginine: a new glucose-protein bond in a self-glucosylating protein from sweet corn. FEBS Lett 376:61–64. https://doi.org/10.1016/0014-5793(95)01247-6.

31. Pearson JS, Giogha C, Ong SY, Kennedy CL, Kelly M, Robinson KS, Lung TW, Mansell A, Riedmaier P, Oates CV, Zaid A, Mühlen S, Crepin VF, Marches O, Ang CS, Williamson NA, O'Reilly LA, Bankovacki A, Nachbur U, Infusini G, Webb AI, Silke J, Strasser A, Frankel G, Hartland EL. 2013. A type III effector antagonizes death receptor signalling during bacterial gut infection. Nature 501:247–251. https://doi.org/10.1038/nature12524.

32. Guex N, Peitsch MC. 1997. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723. https://doi.org/10.1002/elps.1150181505.

33. Kelley LA, Sternberg MJ. 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4:363–371. https://doi.org/10.1038/nprot.2009.2.

34. Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9:40. https://doi.org/10.1186/1471-2105-9-40.

35. Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–738. https://doi.org/10.1038/nprot.2010.5.

36. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER suite: protein structure and function prediction. Nat Methods 12:7–8. https://doi.org/10.1038/nmeth.3213.

37. Ha S, Walker D, Shi Y, Walker S. 2000. The 1.9 A crystal structure of Escherichia coli MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. Protein Sci 9:1045–1052. https://doi.org/10.1110/ps.9.6.1045.

38. Martinez-Fleites C, Macauley MS, He Y, Shen DL, Vocadlo DJ, Davies GJ. 2008. Structure of an O-GlcNAc transferase homolog provides insight into intracellular glycosylation. Nat Struct Mol Biol 15:764–765. https://doi.org/10.1038/nsmb.1443.

39. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178.

40. Liang DM, Liu JH, Wu H, Wang BB, Zhu HJ, Qiao JJ. 2015. Glycosyltransferases: mechanisms and applications in natural product development. Chem Soc Rev 44:8350–8374. https://doi.org/10.1039/c5cs00600g.

41. Viegas A, Manso J, Nobrega FL, Cabrita EJ. 2011. Saturation-transfer difference (STD) NMR: a simple and fast method for ligand screening and characterization of protein binding. J Chem Educ 88:990–994. https://doi.org/10.1021/ed101169t.

42. Sievers F, Higgins DG. 2014. Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol Biol 1079:105–116. https://doi.org/10.1007/978-1-62703-646-7_6.

43. Boels IC, Beerthuyzen MM, Kosters MH, Van Kaauwen MP, Kleerebezem M, De Vos WM. 2004. Identification and functional characterization of the Lactococcus lactis rfb operon, required for dTDP-rhamnose biosynthesis. J Bacteriol 186:1239–1248. https://doi.org/10.1128/JB.186.5.1239-1248.2004.

44. Karimova G, Pidoux J, Ullmann A, Ladant D. 1998. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proc Natl Acad Sci U S A 95:5752–5756. https://doi.org/10.1073/pnas.95.10.5752.

45. Choi S, Choe J. 2011. Crystal structure of elongation factor P from *Pseudomonas aeruginosa* at 1.75 A resolution. Proteins 79:1688–1693. https://doi.org/10.1002/prot.22992.

46. Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem 77:521–555. https://doi.org/10.1146/annurev.biochem.76.061005.092322.

47. Lizak C, Gerber S, Numao S, Aebi M, Locher KP. 2011. X-ray structure of a bacterial oligosaccharyltransferase. Nature 474:350–355. https://doi.org/10.1038/nature10151.

48. Park MH, Cooper HL, Folk JE. 1982. The biosynthesis of protein-bound hypusine (N$^\varepsilon$-(4-amino-2-hydroxybutyl)lysine). Lysine as the amino acid precursor and the intermediate role of deoxyhypusine (N$^\varepsilon$-(4-aminobutyl)lysine). J Biol Chem 257:7217–7222.

49. Rajkovic A, Hummels KR, Witzky A, Erickson S, Gafken PR, Whitelegge JP, Faull KF, Kearns DB, Ibba M. 2016. Translation control of swarming proficiency in *Bacillus subtilis* by 5-amino-pentanolylated elongation factor P. J Biol Chem 291:10976–10985. https://doi.org/10.1074/jbc.M115.712091.

50. Spiro RG. 2002. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology 12:43R–56R. https://doi.org/10.1093/glycob/12.4.43R.

51. Li S, Zhang L, Yao Q, Li L, Dong N, Rong J, Gao W, Ding X, Sun L, Chen X, Chen S, Shao F. 2013. Pathogen blocks host death receptor signalling by arginine GlcNAcylation of death domains. Nature 501:242–246. https://doi.org/10.1038/nature12436.

52. Wong Fok Lung T, Giogha C, Creuzburg K, Ong SY, Pollock GL, Zhang Y, Fung KY, Pearson JS, Hartland EL. 2016. Mutagenesis and functional analysis of the bacterial arginine glycosyltransferase effector NleB1 from enteropathogenic *Escherichia coli*. Infect Immun 84:1346–1360. https://doi.org/10.1128/IAI.01523-15.

53. Sun HY, Lin SW, Ko TP, Pan JF, Liu CL, Lin CN, Wang AH, Lin CH. 2007. Structure and mechanism of *Helicobacter pylori* fucosyltransferase. A basis for lipopolysaccharide variation and inhibitor design. J Biol Chem 282:9973–9982. https://doi.org/10.1074/jbc.M610285200.

54. Hu Y, Chen L, Ha S, Gross B, Falcone B, Walker D, Mokhtarzadeh M, Walker S. 2003. Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. Proc Natl Acad Sci U S A 100:845–849. https://doi.org/10.1073/pnas.0235749100.

55. Lira-Navarrete E, Valero-González J, Villanueva R, Martínez-Júlvez M, Tejero T, Merino P, Panjikar S, Hurtado-Guerrero R. 2011. Structural insights into the mechanism of protein O-fucosylation. PLoS One 6:e25365. https://doi.org/10.1371/journal.pone.0025365.

56. Katz A, Solden L, Zou SB, Navarre WW, Ibba M. 2014. Molecular evolution of protein-RNA mimicry as a mechanism for translational control. Nucleic Acids Res 42:3261–3271. https://doi.org/10.1093/nar/gkt1296.

57. Joe YA, Park MH. 1994. Structural features of the eIF-5A precursor required for posttranslational synthesis of deoxyhypusine. J Biol Chem 269:25916–25921.

58. Qasba PK, Ramakrishnan B, Boeggeman E. 2005. Substrate-induced conformational changes in glycosyltransferases. Trends Biochem Sci 30:53–62. https://doi.org/10.1016/j.tibs.2004.11.005.

59. Ni L, Sun M, Yu H, Chokhawala H, Chen X, Fisher AJ. 2006. Cytidine 5′-monophosphate (CMP)-induced structural changes in a multifunctional sialyltransferase from *Pasteurella multocida*. Biochemistry 45:2139–2148. https://doi.org/10.1021/bi0524013.

60. Bertani G. 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. J Bacteriol 62:293–300.

61. Bertani G. 2004. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. J Bacteriol 186:595–600. https://doi.org/10.1128/JB.186.3.595-600.2004.

62. Miller JH. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

63. Guzman LM, Belin D, Carson MJ, Beckwith J. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P$_{BAD}$ promoter. J Bacteriol 177:4121–4130. https://doi.org/10.1128/jb.177.14.4121-4130.1995.

64. Pospiech A, Neumann B. 1995. A versatile quick-prep of genomic DNA from gram-positive bacteria. Trends Genet 11:217–218. https://doi.org/10.1016/S0168-9525(00)89052-6.

65. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77:51–59. https://doi.org/10.1016/0378-1119(89)90358-2.

66. Lassak J, Henche AL, Binnenkade L, Thormann KM. 2010. ArcS, the cognate sensor kinase in an atypical Arc system of *Shewanella oneidensis* MR-1. Appl Environ Microbiol 76:3263–3274. https://doi.org/10.1128/AEM.00512-10.

67. Sambrook J, Russell DW. 2001. Molecular cloning. A laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

68. Tetsch L, Koller C, Haneburger I, Jung K. 2008. 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. https://doi.org/10.1111/j.1365-2958.2007.06070.x.

69. Miller JH. 1992. A short course in bacterial genetics: a laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

70. Inoue H, Nojima H, Okayama H. 1990. High efficiency transformation of *Escherichia coli* with plasmids. Gene 96:23–28. https://doi.org/10.1016/0378-1119(90)90336-P.

71. Starosta AL, Lassak J, Peil L, Atkinson GC, Woolstenhulme CJ, Virumäe K, Buskirk A, Tenson T, Remme J, Jung K, Wilson DN. 2014. A conserved proline triplet in Val-tRNA synthetase and the origin of elongation factor P. Cell Rep 9:476–483. https://doi.org/10.1016/j.celrep.2014.09.008.

72. Galili G. 1995. Regulation of lysine and threonine synthesis. Plant Cell 7:899–906. https://doi.org/10.1105/tpc.7.7.899.

73. Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254. https://doi.org/10.1016/0003-2697(76)90527-3.

74. Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685. https://doi.org/10.1038/227680a0.

75. Ladner CL, Yang J, Turner RJ, Edwards RA. 2004. Visible fluorescent detection of proteins in polyacrylamide gels without staining. Anal Biochem 326:13–20. https://doi.org/10.1016/j.ab.2003.10.047.

76. Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675. https://doi.org/10.1038/nmeth.2089.

77. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858. https://doi.org/10.1038/nprot.2015.053.

78. Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22:195–201. https://doi.org/10.1093/bioinformatics/bti770.

79. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. 2009. The SWISS-MODEL Repository and associated resources. Nucleic Acids Res 37:D387–D392. https://doi.org/10.1093/nar/gkn750.

80. Guex N, Peitsch MC, Schwede T. 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-Pdb Viewer: a historical perspective. Electrophoresis 30(Suppl 1):S162–S173. https://doi.org/10.1002/elps.200900140.

81. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, Schwede T. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42:W252–W258. https://doi.org/10.1093/nar/gku340.

82. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. https://doi.org/10.1002/jcc.20084.

83. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. Nucleic Acids Res 28:235–242. https://doi.org/10.1093/nar/28.1.235.

84. Volkmer B, Heinemann M. 2011. Condition-dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling. PLoS One 6:e23126. https://doi.org/10.1371/journal.pone.0023126.

85. Cohen D, Mechold U, Nevenzal H, Yarmiyhu Y, Randall TE, Bay DC, Rich JD, Parsek MR, Kaever V, Harrison JJ, Banin E. 2015. Oligoribonuclease is a central feature of cyclic diguanylate signaling in *Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A 112:11359–11364. https://doi.org/10.1073/pnas.1421450112.

86. Sattler M, Schleucher J, Griesinger C. 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog Nucl Magn Reson Spectrosc 34:93–158. https://doi.org/10.1016/S0079-6565(98)00025-9.

87. Schwarzinger S, Kroon GJ, Foss TR, Chung J, Wright PE, Dyson HJ. 2001. Sequence-dependent correction of random coil NMR chemical shifts. J Am Chem Soc 123:2970–2978. https://doi.org/10.1021/ja003760i.

88. Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD. 1995. 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. J Biomol NMR 5:67–81.

89. Pervushin K, Riek R, Wider G, Wüthrich K. 1997. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci U S A 94: 12366–12371. https://doi.org/10.1073/pnas.94.23.12366.

90. Salzmann M, Pervushin K, Wider G, Senn H, Wüthrich K. 1998. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. Proc Natl Acad Sci U S A 95:13585–13590. https://doi.org/10.1073/pnas.95.23.13585.

91. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: a multidimensional spectral processing system based on UniX pipes. J Biomol NMR 6:277–293. https://doi.org/10.1007/BF00197809.

92. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. 2003. PRIMUS—a Windows-PC based system for small-angle scattering data analysis. J Appl Crystallogr 36:1277–1282. https://doi.org/10.1107/S0021889803012779.

93. Svergun DI, Barberato C, Koch MHJ. 1995. CRYSOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Crystallogr 28:768–773. https://doi.org/10.1107/S0021889895007047.

94. Skubák P, Pannu NS. 2013. Automatic protein structure solution from weak X-ray data. Nat Commun 4:2777. https://doi.org/10.1038/ncomms3777.

95. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 67:235–242. https://doi.org/10.1107/S0907444910045749.

96. de Graaff RA, Hilge M, van der Plas JL, Abrahams JP. 2001. Matrix methods for solving protein substructures of chlorine and sulfur from anomalous data. Acta Crystallogr D Biol Crystallogr 57:1857–1862. https://doi.org/10.1107/S0907444901016535.

97. Abrahams JP, Leslie AG. 1996. Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr D Biol Crystallogr 52:30–42. https://doi.org/10.1107/S0907444995008754.

98. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66: 213–221. https://doi.org/10.1107/S0907444909052925.

99. Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60:2126–2132. https://doi.org/10.1107/S0907444904019158.

100. Skubák P, Murshudov GN, Pannu NS. 2004. Direct incorporation of experimental phase information in model refinement. Acta Crystallogr D Biol Crystallogr 60:2196–2201. https://doi.org/10.1107/S0907444904019079.

101. Brunger AT. 2007. Version 1.2 of the crystallography and NMR system. Nat Protoc 2:2728–2733. https://doi.org/10.1038/nprot.2007.406.

102. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. 1998. Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 54:905–921. https://doi.org/10.1107/S0907444998003254.

103. Shen Y, Delaglio F, Cornilescu G, Bax A. 2009. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44:213–223. https://doi.org/10.1007/s10858-009-9333-z.

104. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75.

105. Cao B, Porollo A, Adamczak R, Jarrell M, Meller J. 2006. Enhanced recognition of protein transmembrane domains with prediction-based structural profiles. Bioinformatics 22:303–309. https://doi.org/10.1093/bioinformatics/bti784.

106. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED. 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59: 687–696. https://doi.org/10.1002/prot.20449.

4 Switching the post-translational modification of elongation factor P

Manuscript

Wolfram Volkwein[1,7], Ralph Krafczyk[1,7], Pravin Kumar Ankush Jagtap[2], Marina Parr[3,5], Elena Mankina[3], Jakub Macošek[2,4], Zhenghuan Guo[1], Maximilian Josef Ludwig Johannes Fürst[1,6], Miriam Pfab[1], Dmitrij Frishman[3,5], Janosch Hennig[2], Kirsten Jung[1] & Jürgen Lassak[1,8]*

[1]Center for integrated Protein Science Munich (CiPSM), Department of Biology I, Microbiology, Ludwig-Maximilians-Universität München, Munich, Germany

[2]Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL) Heidelberg, Heidelberg, Germany

[3]Department of Bioinformatics, Wissenschaftszentrum Weihenstephan, Technische Universität München, Freising, Germany

[4]Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences

[5]St. Petersburg State Polytechnic University, St. Petersburg, Russia

[6]Current address: Molecular Enzymology Group, University of Groningen, Nijenborgh 4, 9747AG, Groningen, The Netherlands

[7]These authors contributed equally

[8]Lead Contact

*Correspondence: juergen.lassak@lmu.de

Summary

Tripeptides with two consecutive prolines are the shortest and most frequent sequences causing ribosome stalling. The bacterial translation elongation factor P (EF-P) relieves the arrest, allowing protein biosynthesis to continue. A seven amino acids long loop between beta-strands β3/β4 is crucial for EF-P function and is modified at its tip by lysylation of lysine or rhamnosylation of arginine. Phylogenetic analyses unveiled an invariant proline in the -2 position of the modification site in EF-Ps that utilize lysine modifications such as *Escherichia coli*. Bacteria with arginine modifications such as *Pseudomonas putida* have selected against it. Focusing on the EF-Ps from these two model organisms we demonstrate the importance of the β3/β4 loop composition for functionalization with chemically distinct modifications. Ultimately we show that amino acid changes in *E. coli* EF-P are needed for switching the activation strategy from lysylation to rhamnosylation.

Introduction

Protein biosynthesis is a universally conserved three-step process that occurs on ribosomes and provides a platform for tRNA mediated amino acid delivery. During translation elongation aminoacyl-tRNAs bind to the ribosomal A-site and peptide bond formation is mediated with a peptidyl-tRNA located in the P-site. Relocation of the P-site tRNA to the E-site enables ribosome exiting. The speed of incorporating amino acids into the growing polypeptide chain varies and strongly depends on their chemical nature (1). Due to its rigid structure, particularly proline delays the peptidyl transfer reaction, being both a poor A-site donor and P-site acceptor substrate (1-6). When translating stretches of two or more prolines, ribosomes become arrested (5, 7-15). Thus consecutive prolines are disfavored in evolution (16). However, the structural benefits of polyproline sequences in proteins (17, 18) seems to outweigh the translational drawback and favored the evolution of a specialized universal elongation factor termed e/aIF5A in eukaryotes/archaea and EF-P in bacteria (5, 8, 11). Upon polyproline mediated stalling e/aIF5A and EF-P are recruited to the ribosome and bind between the P- and E-tRNA binding sites (19-24).

With its three domains, EF-P (Figure 1A) spans both ribosomal subunits and forms an L-shaped, tRNA mimicking structure (25, 26). Whereas the two OB-fold domains (Oligonucleotide Binding) II and III are likely to be involved in P-site tRNA^Pro (27) and E-site codon (24) recognition, the EF-P KOW-like N-domain I is crucial for the catalytic activity. Specifically, a seven amino acid long apical loop region between beta strands three and four (β3Ωβ4) protrudes towards the peptidyl transferase center (19, 24). A conserved positively charged residue at the loop tip mediates stabilization and positioning of the CCA-end of the P-site tRNA^Pro in a way favorable for peptide bond formation (5, 6, 28). EF-P activity is further

enhanced by post-translational extensions of this specific tip residue (5, 28). Interestingly the underlying bacterial modifications appear to be chemically diverse (29) (Figure 1B). In a subset of bacteria including the Gram-negative model organism *Escherichia coli*, a lysine residue K34 is β-lysylated (30-32) with (*R*)-β-lysine (33) at the ε-amino group, employing the catalytic activity of the EF-P specific ligase EpmA (YjeA, PoxA, GenX) (34). Subsequent hydroxylation by EpmC (YfcM) (35, 36) presumably at the (*R*)-β-lysyl-lysine C5 atom (24) completes the modification, but is negligible for function (37). A chemically related amino acid - 5-amino-pentanolyl-lysine - was found in *Bacillus subtilis* EF-P (38). By contrast, activity of a distinct EF-P subset encoded in the β-proteobacterial subdivision and certain γ-proteobacteria such as *Pseudomonas putida* and *Shewanella oneidensis* depends on α-rhamnosylation of arginine at the equivalent position (28, 39, 40). This glycosylation is mediated by the GT-B fold glycosyltransferase EarP (41, 42) belonging to the enzyme family GT104 according to the CAZy database (43).

Despite their distinct chemical nature both lysine as well as arginine modifications of EF-P promote proline-proline peptide bond formation. This raises the question which protruding residue and modification puts which selective pressure onto the EF-P β3Ωβ4 sequence to ultimately fulfill the same function. Using bioinformatics and site directed mutagenesis, we were able to show that only very few sequential changes in β3Ωβ4 of EF-P are needed to switch modification specificity and even allow functionalization. Beside the crucial lysine/arginine our analyses pointed towards amino acids located at the 2nd position N-terminal of the modification site. While bacteria that encode EF-P with protruding lysine contain an invariant proline, those with protruding arginine instead strictly select against it. We argue that the presence or absence of this specific proline orients β3Ωβ4 in a way that allows EF-P activation with modifications similar to either (*R*)-β-lysylation or α-rhamnosylation.

Results and discussion

Phylogenetic analysis of EF-P β3Ωβ4

Our study began with a bioinformatics analysis of EF-P β3Ωβ4. In a first step we constructed a phylogenetic tree based on 4421 sequences including both EF-P and IF5A sequences, with the latter used as an outgroup. To define modification specific protein subsets, EpmABC and EarP orthologs were collected as described previously (28) (Figure S1A, Table S1). The EF-P modification system present in *B. subtilis* was excluded in this study, as the full pathway is still poorly understood (38, 44). A first weblogo of β3Ωβ4 numbered according to the *E. coli* protein (amino acids 31 to 37) was generated based on the complete EF-P dataset (Figure 1A, Table S1). In line with earlier reports (30) the vast majority (78.91%) of EF-Ps have a lysine at the β3Ωβ4 tip (K34), whereas arginine is the second most frequent

amino acid occurring in 15.33% of the proteins (Table S1). The remaining 5.76% contain A (0.05%), G (0.07%), H (0.02), M (0.68%), N (1.97%), Q (2.78%) and S (0.18%) in this position. We next extracted two subsets of proteins with either a protruding lysine (lysine type) or arginine (arginine type) (Figure 1C). This analysis revealed a highly conserved proline in the 2nd position N-terminal of the modification site (P32) in the lysine type subset being almost absent in the arginine type EF-Ps. Consistently, bacteria with EpmA pathway have the conserved proline whereas bacteria with EarP do not (Figure 1D). The two modification systems thus appear to be mutually exclusive (28) (Figure S1B). Based on these observations, EF-P sequences were grouped according to the presence or absence of P32 (Figure 1E). Beside lysine (97.25%) we also found that asparagine in the protruding position strongly co-occurs with proline (97.7%), whereas other types of amino acids co-occur with P32 extremely rarely: arginine (2.95%), alanine (0%), glycine (0%), histidine (0%), methionine (0%), glutamine (0%) and serine (0%) (Figure 1D, Table S1).

These findings led us to speculate that P32 might play a role in orienting the protruding residue for proper functionality and thus be a determinant of the chemical nature of the post-translational modification. Phylogenetic analysis of β3Ωβ4 in EF-P and the variability in this crucial structural element also prompted us to investigate the evolutionary order of events resulting in the observed co-occurrence patterns between the residues occupying positions 34 and 32 as well as in the emergence of the EpmA or EarP modifying enzymes. To this end we performed a phylogenetic tree reconstruction using the maximum likelihood method from the phytools R package (45). As the lysine at the β3Ωβ4 tip was found in more than three quarters of all EF-P sequences, and is also conserved in the eukaryotic and archaeal orthologs e/aIF5A, we hypothesized that this amino acid is evolutionary ancient. Indeed, we found EF-P with a protruding lysine to be most likely (62%) at the root of our tree with subsequent emergence of the first arginine, followed by asparagine, glutamine and methionine (Figure S1D).

When reconstructing evolutionary scenarios at position 32 (Figure 2A), threonine is the most likely amino acid in an EF-P progenitor despite the fact that proline is found in about 80% of extant sequences. Interestingly, threonine in the equivalent position is highly conserved in the archaeal/eukaryotic EF-P homolog a/eIF5A as well (46). As a common ancestor of EF-P/IF5A most likely emerged before diversification into the three domains of life one can speculate that threonine was disfavored in most bacterial lineages and outcompeted by proline to match the specific requirements of prokaryotic ribosomes. Based on the reconstructed phylogenetic order of events the emergence of proline might also has favored the evolution of the EpmABC modification system (Figures S1C and S1D). By contrast, an evolutionary distinct EF-P branch, which lacks proline at position 32 and with a protruding arginine, allowed for the recruitment of EarP and rhamnosylation as an activation strategy.

Although our data suggest that EpmABC is phylogenetically older than EarP, the latter seems to have evolved independently, but interestingly from a P32-containing progenitor (Figures S1D and S1C). It is also noteworthy that our data imply an advantage of arginine over lysine at the β3Ωβ4 tip during early evolution of EF-P, which was lost once proline P32 had emerged (Figures S1E and S1D).

Mutagenesis of β3Ωβ4 of *E. coli* and *P. putida* EF-P

To investigate the importance of the amino acid composition of β3Ωβ4 we initially chose the EF-P from *E. coli* (EF-P$_{Eco}$) being dependent on EpmABC (R)-β-lysylation of lysine K34. Accordingly, when analyzing β3Ωβ4 we first focused on K34 and its impact on EF-P$_{Eco}$ activity. EF-P activity was measured *in vivo* using a previously established β-galactosidase dependent reporter system (11) (Figure 3A). The assay is based on the effective translation of the polyproline motif containing acid stress responsive transcriptional regulator CadC (47, 48) and activation of its cognate promoter P$_{cadBA}$ fused to *lacZ* (11). Consequently, β-galactosidase activity is low in *E. coli* cells lacking *efp* but becomes elevated when complementing with a wild-type EF-P$_{Eco}$ copy in trans (Figure 3B). Interestingly the absence of the modification system allows for residual P$_{cadBA}$ activation which could be further enhanced upon *efp*$_{Eco}$ overexpression (Figure S2A). Presumably the increased copy number of unmodified EF-P$_{Eco}$ can partially compensate for a lack of modification as the lysine K34 side chain forms important stabilizing contacts with the CCA-end of the P-site tRNA^{Pro} (24). We next substituted K34 by any other amino acid to be found in the protruding position of β3Ωβ4 of other bacteria (Figure 1A, Table S1). β-galactosidase activity was completely lost when exchanging lysine by alanine (K34A$_{Eco}$), asparagine (K34N$_{Eco}$), glutamine (K34Q$_{Eco}$) or methionine (K34M$_{Eco}$) (Figure 3B). Presumably species with EF-Ps containing these tip residues underwent further evolutionary adjustments or might have developed yet unknown post-translational modifications that promote polyproline biosynthesis. By contrast we observed residual EF-P activity when complementing with K34R$_{Eco}$. On the one hand side chain similarities of arginine with lysine presumably preserve certain of the above-mentioned interactions that might be otherwise absent. On the other hand, the significant decrease in activity with K34R$_{Eco}$ compared to unmodified EF-P$_{Eco}$ points towards a non-stimulating or even negative effect, possibly caused by the guanidino group. Having demonstrated that substitution of K34 is hardly tolerated, we went on to analyze the impact of its sequence context residues. Consistent with our phylogenetic analysis (Figures 1C and 1E) an exchange of P32 (P32S$_{Eco}$, P32G$_{Eco}$) abolishes EF-P activity, as can be concluded from low β-galactosidase activities (Figure 3B). Similarly, substitution of G33 (G33A$_{Eco}$, G33S$_{Eco}$) is not tolerated and leads to a loss of function. In comparison, when mutating G35 (G35N$_{Eco}$) and Q36 (Q36S$_{Eco}$) a residual rescue activity is retained. Altogether our analysis of EF-P$_{Eco}$

β3Ωβ4 unveils important determinants for protein function and thus explains their high degree of conservation.

As second EF-P, we investigated the one from *P. putida* KT2440 (EF-P$_{Ppu}$) being activated by EarP mediated arginine R32 α-rhamnosylation. We generated the substitution variants K29R$_{Ppu}$, S30P$_{Ppu}$, R32K$_{Ppu}$, N33G$_{Ppu}$ and S34Q$_{Ppu}$ according to amino acids predominantly found in the lysine type weblogo (Figure 1C). We also constructed K29A$_{Ppu}$, S30A$_{Ppu}$, S30G$_{Ppu}$, G31A$_{Ppu}$, G31S$_{Ppu}$, N33D$_{Ppu}$, S34A$_{Ppu}$ and A35S$_{Ppu}$ to further study the impact of the corresponding positions on EF-P activity and rhamnosylation efficiency. An in vitro time course analysis was performed (Figure 3C) with wild-type EF-P$_{Ppu}$ as well as its variants S30P$_{Ppu}$, G31A$_{Ppu}$, R32K$_{Ppu}$, N33G$_{Ppu}$ and S34Q$_{Ppu}$ using 50 µM TDP-β-L-rhamnose (=10fold K_M; Figure S3A; see also supplemental materials and methods). This revealed relative rhamnosylation rates with S30P$_{Ppu}$ (1% of wild-type activity) and G31A$_{Ppu}$ (1% of wild-type activity) being slowest, while N33G$_{Ppu}$ and S34Q$_{Ppu}$ reach 62% and 12% compared to wild-type EF-P$_{Ppu}$, respectively (Figure 3C). We next assessed *in vivo* EF-P$_{Ppu}$ activity employing the *E. coli* P$_{cadBA}$::*lacZ* reporter strain. This is possible as cross-complementation with a combination of EF-P$_{Ppu}$ and EarP$_{Ppu}$ (Figure 3D) restores wild-type levels of β-galactosidase activity in Δ*efp* cells. We note that all EF-P$_{Ppu}$ variants – except changes of R32 – were fully modified by EarP$_{Ppu}$ when co-expressing them in *E. coli in vivo* (Figure 3D). Accordingly, one can assume that when analyzing EF-P$_{Ppu}$ variants a corresponding P$_{cadBA}$::*lacZ* mutant phenotype rather reflects impaired protein function than reduced rhamnosylation rate. In this regard we found S30P$_{Ppu}$ to be almost inactive (4% of wild-type β-galactosidase activity) and with this matching the result of the corresponding *E. coli* EF-P substitution P32S. On the contrary, the alanine and glycine substitutions S30A$_{Ppu}$ and S30G$_{Ppu}$ reached 39% and 96% of wild-type β-galactosidase activity, respectively. These data support our observation of a strong selection against proline in the arginine type EF-Ps, but at the same time allowing for a certain degree of freedom in the -2 position of the modification site. Substitutions in positions N33, S34 and A35 as well as K29 are also tolerated without significant EF-P activity loss *in vivo* (Figure 3D). Similar to G33 in EF-P$_{Eco}$ (32) (Figure 3B) the position equivalent G31 in EF-P$_{Ppu}$ is crucial for both modification efficiency (Figure 3C) and protein function (Figure 3D), which might be explained by sterically hindering interactions with either the ribosome or EarP caused by longer side chains. Interestingly and in contrast to K34R$_{Eco}$ R32K$_{Ppu}$ is not only inactive but the β-galactosidase activity measured with this variant is even below the level of Δ*efp* harboring the empty vector control. This drastic phenotype indicates an inhibitory effect on polyproline translation. Notably, we saw the same when testing unmodified EF-P of both *P. putida* or *S. oneidensis* (28) (Figure 3D). A similar phenomenon was also observed by others when analyzing the growth of the *P. aeruginosa*

EF-P R32K variant (39). The most plausible explanation is a distinct orientation of the protruding residue in dependence of β3Ωβ4 composition.

Activation of *E. coli* EF-P hybrids by rhamnosylation

Having established the impact of single substitution variants on function for both *E. coli* and *P. putida* EF-P, we were curious whether we can adjust the β3Ωβ4 loop composition in a way that allows for cross-interaction, cross-rhamnosylation and ultimately cross-activation of *E. coli* EF-P by EarP. Based on the knowledge we gained from our experiments here and earlier studies (41, 42) we chose EF-P$_{Eco}$ hybrids with a swapped β3Ωβ4 as well as also included the K34R$_{Eco}$ and the double substitution P32S/K34R$_{Eco}$ to be initially analyzed.

Despite their sequential diversity EF-P$_{Eco}$ and EF-P$_{Ppu}$ are structurally similar (32, 49), which led us to hypothesize about a possible cross-interaction of EarP$_{Ppu}$ even with wild-type EF-P$_{Eco}$. To test this, we constructed an *E. coli* Δ*cyaA* deletion, in which we integrated the *lux* operon (*luxCDABE*) at the *lac* locus (Figure 4A). The resultant strain KV1 was used in combination with the Euromedex bacterial two-hybrid system (50). This system is based on functional reconstitution of split *Bordetella pertussis* adenylate cyclase CyaA, which in turn activates the *lac* promoter P$_{lac}$ being dependent on cAMP receptor protein CAP (Figure 4A). As proof of principle, self-interaction of the GCN4 leucine zipper (50) was assessed in *E. coli* KV1 and the commercially available bacterial two-hybrid strains *E. coli* BTH101 (Euromedex) and *E. coli* DHM1 (Euromedex) on X-Gal containing screening plates. For this purpose, the reporter strains were co-transformed with the plasmids pKT25-zip (Euromedex) and pUT18C-zip (Euromedex) that encode for protein hybrids of the leucine zipper and the corresponding CyaA fragment. While all reporter strains responded with comparable β-galactosidase mediated color formation on culture plates, KV1 exhibits an additional light output. In liquid culture, we could quantify the light output over a course of approximately 40 hours. In this regard, *E. coli* KV1 transformed with pKT25-zip and pUT18C-zip, emitted a maximum of 12×10^6 relative light units (RLU) after ~10 hours. These results demonstrate that *E. coli* KV1 enables continuous monitoring of protein-protein interaction by detection of light emission while maintaining the possibility for blue/white screening assays on X-Gal containing culture plates (Figure 4B).

Having demonstrated the functionality of our reporter KV1 we asked whether we can detect even the transient interaction that occurs between the enzyme EarP$_{Ppu}$ and its (non)cognate partner EF-P$_{(Eco)/Ppu}$. To this end C-terminal fusions of T25 and T18 (T25-EarP$_{Ppu}$, T18-EF-P$_{Ppu}$, T18-EF-P$_{Eco}$, T18-EF-P K34R$_{Eco}$, T18-EF-P P32S K34R$_{Eco}$, T18-EF-P$_{Eco}$ loop$_{Ppu}$) were generated using commercially available expression vectors (Euromedex). Interactions were assessed by continuous monitoring of luminescence emission from *E. coli* KV1 (Figure S2D). Co-expression of the cognate interaction partners EarP$_{Ppu}$ and EF-P$_{Ppu}$ resulted in a maximum of 7,000 RLU (Figure 4C). When we co-produced EarP$_{Ppu}$ and the non-cognate

EF-P$_{Eco}$, a maximal light emission of 255 RLU was observed. This value was significantly higher than the luminescence generated in any combination that employs solely T18 or T25 (maximal RLU of <100) and thus clearly demonstrates cross-interaction of EarP$_{Ppu}$ and EF-P$_{Eco}$. We also observed interaction between EarP and EF-P$_{Eco}$ loop$_{Ppu}$, K34R$_{Eco}$ and P32S/K34R$_{Eco}$ however their strengths were indistinguishable from the one with wild-type EF-P$_{Eco}$. We additionally performed NMR titration experiments employing EF-P$_{Eco}$ and its variants with EarP$_{Ppu}$ to determine their interaction at a molecular level in vitro (Figure 4D, S4A, S4B). EF-P$_{Eco}$ interacts with EarP$_{Ppu}$ as shown by the decrease in total amount of peak intensities in the EF-P$_{Eco}$ ^{15}N-HSQC spectrum upon EarP$_{Ppu}$ titration. Physical interaction leads to an increased molecular tumbling time, which in turn decreases transverse relaxation times and peak intensities. The interaction was substantially enhanced in the K34R$_{Eco}$ variant, resulting in even lower peak intensities. This was expected, as R34 makes important contacts with EarP and its cognate EF-P in *Neisseria meningitidis* (42). In contrast to K34R$_{Eco}$ we observed reduced interaction strength in the P32S variant as peak intensities were stronger than for EF-P$_{Eco}$ wild type. This result might be counterintuitive, however, only if one ignores that EF-P must not only be efficiently rhamnosylated by EarP, but at the same time has to interact optimally with the P-site tRNA on the ribosome. In this light, the substitution of proline can be regarded as an evolutionary consequence to maintain functionality at the expense of rhamnosylation efficiency. In line with the findings for K34R$_{Eco}$ and P32S$_{Eco}$ the K34R/P32S$_{Eco}$ double substitution variant showed intermediate interaction with EarP$_{Ppu}$ compared to K34R$_{Eco}$ and increased further with the EF-P$_{Eco}$ loop$_{Ppu}$ variant. In addition to K34R$_{Eco}$ and P32S$_{Eco}$ the EF-P$_{Eco}$ loop$_{Ppu}$ construct bears two additional substitutions at positions 35 and 36, which seem to be important for EF-P/EarP interaction. Thus, we can interpret our finding as an adjustment to compensate for the negative interaction effect, that we saw with P32S$_{Eco}$. It is possible that substitution of proline P32 cause substantial changes in the loop dynamics due to its rigid nature. To test this, we performed ^{15}N R_1, R_2, and steady-state heteronuclear {^{1}H}–^{15}N-NOE relaxation experiments on EF-P$_{Eco}$ and its variants and compared it with EF-P$_{Ppu}$. Our analysis suggests that substitution of single EF-P$_{Eco}$ β3Ωβ4-loop residues with residues from EF-P$_{Ppu}$ or even with the complete β3Ωβ4 does not significantly alter the NMR relaxation properties of β3Ωβ4 and hence its dynamics (Figure S4C-F). Thus, differences observed in the interaction of EF-P$_{Eco}$ and its variants with EF-P$_{Ppu}$ can be attributed to the molecular nature of resulting interactions rather than changes in the loop dynamics.

Knowing that EF-P$_{Eco}$ and EarP$_{Ppu}$ do cross-interact, we next assessed cross-rhamnosylation and activation. First, we tested the EF-P$_{Eco}$ variant in which solely the modification site K34 was substituted to arginine. Strikingly, this single substitution was sufficient to allow *in vivo* modification in an EarP dependent manner, however did not result in EF-P activation (Figure

4E). Moreover, we did not observe significant rhamnosylation in vitro presumably as a result of suboptimal contacts between EF-P$_{Eco}$ and EarP$_{Ppu}$ (42) (Figure 4F, Figure S3B). As important interaction sites between EF-P and its corresponding modification system are predominantly located within the first 65 amino acids (31, 32, 41, 42) we next swapped the N-domain of EF-P$_{Eco}$ with the one from EF-P$_{Ppu}$ (EF-P$_{Eco}$ domainI$_{Ppu}$). In line with our expectations EF-P$_{Eco}$ domainI$_{Ppu}$ was readily modified both *in vivo* and in vitro (Figures 4E-4F and S3B) and allowed activation comparable to wild-type EF-P$_{Ppu}$. It also shows that binding to the ribosome occurs presumably at the same position, despite an identity below 30% between the OB-domains of the two EF-Ps. In a subsequent step we tested the variant with swapped β3Ωβ4 - EF-P$_{Eco}$ loop$_{Ppu}$ with in total four amino acid substitutions P32S, K34R, G35N and Q36S. Subsequent β-galactosidase assays showed that these alterations are sufficient to allow for partial EF-P activation by rhamnosylation. The remaining discrepancy in β-galactosidase activities compared with wild-type EF-P$_{Ppu}$ is presumably a result of a less effective interaction with EarP as can be concluded from the in vitro interaction studies (Figure 4D) as well as the time course experiments with EF-P$_{Eco}$ loop$_{Ppu}$ (Figure 4F and S3B). We also note that the activity of EF-P$_{Eco}$ loop$_{Ppu}$ is not restricted to three consecutive prolines, but also alleviates ribosome stalling at other diprolyl arrest motifs as demonstrated for APP, DPP, PPD, PPG and PPN (Figure S2B).

Our data shows that the presence of P32 is crucial for activation of EF-P$_{Eco}$ on the one hand and prevents activation of EF-P$_{Ppu}$ on the other hand. Although our *in vivo* analysis with P32S/K34R$_{Eco}$ did not result in a significant upshift of β-galactosidase activity in combination with EarP$_{Ppu}$, we measured slightly increased Miller units when using EarP of *S. oneidensis* MR-1 instead (Figure S2C). Sengoku and co-workers previously showed that amino acids immediately downstream of the modification site interact with EarP. Consequently, we additionally substituted G35 and Q36 for asparagine and serine. Congruent with this assumption both resultant EF-P$_{Eco}$ variants P32S/K34R/G35N$_{Eco}$ and P32S/K34R/Q36S$_{Eco}$ alleviated CadC translation when coproducing EarP$_{Ppu}$, exhibited by a 2fold and 3fold increase in β-galactosidase activities, respectively. However, neither the double substitution K34R/G35N$_{Eco}$, K34R/Q36S$_{Eco}$ nor the triple exchange K34R/G35N/Q36$_{Eco}$ had an alleviating effect on translational arrest. Thus, our analysis clearly shows that cross-activation of EF-P$_{Eco}$ by EarP$_{Ppu}$ is strictly prohibited in the presence of proline P32. On the contrary cross-modification solely depends on the protruding residue to be arginine. Combined with our *in vitro* interaction analysis and the fact that a substitution of P32 even weakens the contact of EF-P$_{Eco}$ with EarP$_{Ppu}$, we conclude that specifically the selection against proline is an adaptation to rescue polyproline stalled ribosomes with α-rhamnosylarginine rather than for efficient modification. On the other hand, our data also implies that additional adjustments in

the β3Ωβ4 sequence composition have been made to compensate for the negative effect on rhamnosylation by EarP in EF-Ps lacking P32.

Concluding remarks

In this study we provide a comprehensive analysis of EF-P β3Ωβ4 and how its sequence composition allows functionalization with chemically and structurally distinct modifications. It might also help to predict the type of novel, yet undiscovered EF-P post-translational activation strategies in the >50% of bacteria which do not encode any known modification enzyme. Our assumption is supported by the recent identification of lysine 5-amino-pentanolylation which takes place in *B. subtilis* and presumably a few other firmicutes (51). This EF-P activation strategy chemically resembles β-lysylation and also occurs on a β3/β4 loop with an invariant proline in position 32. Further, in certain prokaryotes that have a β3Ωβ4 similar to EF-P$_{Eco}$ with lysine at the loop tip or alternatively an asparagine (Figure 1E), one can identify a deoxyhypusine synthase (DHS) like protein (52). In eukaryotes and archaea DHS elongates a lysine in the EF-P ortholog IF5A by an amino-butyryl moiety (53, 54). Accordingly, it is plausible that the bacterial ortholog might attach an analogous modification onto the respective EF-Ps although the experimental connection remains elusive.

The evolutionary flexibility in modification systems and β3Ωβ4 sequence composition is not fully understood yet. However, one could speculate that beside the universally conserved role in alleviating ribosome stalling at polyproline stretches (8, 11) diverse EF-Ps might have extended functionality. In this regard the Green and the Alepuz lab (55, 56) reported that IF5A also acts on non-polyproline arrest motifs and even facilitates termination. Although EF-P activity seems to be restricted to the alleviation of translational arrest situations at consecutive prolines (11, 15) it should be noted that all global analyses thus far were performed solely in *E. coli* (10, 15) and *Salmonella enterica* (9) both of which depend on (*R*)-β-lysylation of lysine. One might therefore speculate whether other EF-Ps with distinct modifications and β3Ωβ4 sequence composition might have expanded functions similar to eIF5A. EarP dependent EF-Ps might therefore be of particular interest. First evolved in β-proteobacteria, this EF-P type seems to have spread into certain γ-proteobacterial orders and other phyla (28). Conversely however, horizontal gene transfer events of EpmABC dependent EF-Ps into the β-proteobacterial subdivision hardly occur. This in turn could indicate a selection in favor of EF-P arginine rhamnosylation caused either by an expanded target spectrum or improved functionality.

The results of this study also demonstrate the possibility of switching the EarP acceptor substrate specificity. The interaction of EarP with its cognate EF-P has been shown to be both sequence- and structure dependent (42). Our data show that a substitution of lysine to

arginine in the EF-P of *E. coli* K34*Eco* is already sufficient to allow for rhamnosylation in an EarP dependent manner. As the EF-Ps of *E. coli* and *P. putida* share only 30% identity in the EarP-interacting EF-P_N domain, sequence specific contacts between EF-P and EarP (42) might only enhance interaction strength between the two proteins. This is further supported by our corresponding bacterial two hybrid and *in vitro* NMR analyses. The recognition motif for the AIDA-associated heptosyltransferase Aah has been described as a "short b-strand– short acceptor loop–short b-strand" (57). Analogously the two beta-strands bracketing β3Ωβ4 might constitute a structural recognition motif for EarP dependent rhamnosylation. Determining the minimal recognition motif is of particular interest as this information allows for targeted rhamnosylation even for proteins other than EF-P. Thus, our study also lays the foundation to evolve EarP into a glycosynthase that can ultimately be used in heterologous production of eukaryotic glycoproteins.

Material and methods

Bioinformatics analysis

Amino acid sequence of EF-P/IF5A, EpmA, EpmB, EpmC and EarP were collected by extracting sequences that have EF-P_N, tRNA-synthetase II, Radical_SAM, EpmC, and DUF2331 domains, respectively, from the Pfam and InterPro databases with subsequent filtration, as described in the Supplementary Section. The final dataset contains 4421, 858, 4894, 317 and 306 sequences of EF-P/IF5A, EpmA, EpmB, EpmC and EarP, respectively (Table S1). Multiple sequence alignments (MSA) and phylogenetic trees were built using MAFFT and FastTree (58, 59). Taxonomic assignments for the EF-P/IF5A and modification system proteins were determined based on the NCBI Taxonomy IDs. Using a rooted phylogenetic tree of the EF-P KOW-like N-domain I, maximum likelihood ancestral state reconstruction was performed with the phytools R package (45).

Bacterial stains, oligonucleotides and plasmids

A description of all strains, plasmids and oligonucleotides used in this study can be found in the Supplemental Information (Table S2, Table S3 and Table S4). *E. coli* strain KV1 for bacterial two-hybrid analysis was constructed as follows: The *luxCDABE* operon from *Photorhabdus luminescens* was amplified from pBAD/HisA-Lux (60) and integrated into *E. coli* LF1 as essentially described previously by (61). To keep the ability of blue/white screening, a synthetic ribosomal binding site (62, 63) which precedes the *lacZ* ORF was introduced. Afterwards *cyaA* was deleted using Red®/ET® recombination technology and the kanamycine cassette was removed using the 709-FLPe/amp expression vector in accordance to the Quick & Easy *E. coli* Gene Deletion Kit (Gene Bridges, Germany). In the same way, *epmA* was deleted in the *E. coli* Δ*efp* reporter strain MG-CR-efp-KanS, resulting in the Δ*efp/*Δ*epmA* reporter strain MG-CR-*efp*-*epmA*-KanR. The Δ*efp* reporter strain MG-CR-

efp-KanS itself was generated by removing the kanamycin resistance cassette from MG-CR-*efp* (28) using also the Quick & Easy *E. coli* Gene Deletion Kit of Gene Bridges according to the manufacturer's instructions.

β-galactosidase activity assay

E. coli Δ*efp* (MG-CR-efp-KanS) or Δ*efp*/Δ*epmA* (MG-CR-efp-epmA-KanR) reporter strain cells, in which *lacZ* expression is controlled by the *cadBA* promoter, were grown over night in 100 mM sodium-phosphate buffered Miller modified LB (pH 5.8) under microaerobic conditions, with minimal agitation at 37 °C. On the next day, cells were harvested by centrifugation, and the β-galactosidase activities were determined as described (64) and are given in relative Miller units (MU) (65).

NMR experiments

NMR ^{1}H, ^{15}N HSQC titration and relaxation experiments were performed at 298 K on an 800 MHz Bruker Avance III spectrometer equipped with a TXI cryogenic probehead on ^{15}N-labeled EF-P$_{Eco}$ and its variants and unlabeled EarP$_{Ppu}$. Backbone assignment of EF-P$_{Eco}$ and EF-P$_{Eco}$loop$_{Ppu}$ was performed on ^{15}N, ^{13}C isotope labeled protein using HNCACB, CBCA(CO)NH and HNCA experiments. Assignments of all other EF-P variants were done by tracing the peaks from the EF-P$_{Eco}$ and EF-P$_{Eco}$ loop$_{Ppu}$ spectra. For NMR titrations, 2fold excess of unlabeled EarP was titrated to ^{15}N labelled EF-P$_{Eco}$ variants and NMR relaxation experiments were performed at concentrations ranging between 0.15-0.18 mM EF-P$_{Eco}$ variants except for P32S$_{Eco}$ variant for which 0.09 mM protein was used for due to lower expression yield. Detailed experimental parameters are descried in the supplementary text.

Author contribution

D.F., M.P., E.M. and J.L. performed bioinformatic analyses. NMR studies were performed by J.H., P.K.A.J. and J.M.. The corresponding proteins were produced and purified by W.V.. M.Pf. performed isoelectric focusing experiments. R.K. and Z.G. performed *in vitro* rhamnosylation assays. All other biochemical and genetic analyses of β3Ωβ4 substitution variants of *E. coli* and *P. putida* were conducted by W.V. and R.K. W.V. and M.J.L.J.F. performed the biochemical analysis with EarP from *S. oneidensis* with contributions from J.L.. J.L., J.H., K.J. and D.F. designed the study. The manuscript was written by W.V., R.K., K.J., E.M., P.K.A.J., J.H. and J.L..

Declaration

The authors declare no competing interest.

References

1. **Pavlov MY, Watts RE, Tan Z, Cornish VW, Ehrenberg M, Forster AC.** 2009. Slow peptide bond formation by proline and other N-alkylamino acids in translation. Proc Natl Acad Sci USA **106:**50-54.
2. **Muto H, Ito K.** 2008. Peptidyl-prolyl-tRNA at the ribosomal P-site reacts poorly with puromycin. Biochem Biophys Res Commun **366:**1043-1047.
3. **Wohlgemuth I, Brenner S, Beringer M, Rodnina MV.** 2008. Modulation of the rate of peptidyl transfer on the ribosome by the nature of substrates. J Biol Chem **283:**32229-32235.
4. **Johansson M, Ieong KW, Trobro S, Strazewski P, Aqvist J, Pavlov MY, Ehrenberg M.** 2011. pH-sensitivity of the ribosomal peptidyl transfer reaction dependent on the identity of the A-site aminoacyl-tRNA. Proc Natl Acad Sci USA **108:**79-84.
5. **Doerfel LK, Wohlgemuth I, Kothe C, Peske F, Urlaub H, Rodnina MV.** 2013. EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. Science **339:**85-88.
6. **Doerfel LK, Wohlgemuth I, Kubyshkin V, Starosta AL, Wilson DN, Budisa N, Rodnina MV.** 2015. Entropic contribution of elongation factor P to proline positioning at the catalytic center of the ribosome. J Am Chem Soc **137:**12997-13006.
7. **Tanner DR, Cariello DA, Woolstenhulme CJ, Broadbent MA, Buskirk AR.** 2009. Genetic identification of nascent peptides that induce ribosome stalling. J Biol Chem **284:**34809-34818.
8. **Gutierrez E, Shin BS, Woolstenhulme CJ, Kim JR, Saini P, Buskirk AR, Dever TE.** 2013. eIF5A promotes translation of polyproline motifs. Mol Cell **51:**35-45.
9. **Hersch SJ, Wang M, Zou SB, Moon KM, Foster LJ, Ibba M, Navarre WW.** 2013. Divergent protein motifs direct elongation factor P-mediated translational regulation in Salmonella enterica and Escherichia *coli*. MBio **4:**e00180-00113.
10. **Peil L, Starosta AL, Lassak J, Atkinson GC, Virumae K, Spitzer M, Tenson T, Jung K, Remme J, Wilson DN.** 2013. Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. Proc Natl Acad Sci USA **110:**15265-15270.
11. **Ude S, Lassak J, Starosta AL, Kraxenberger T, Wilson DN, Jung K.** 2013. Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. Science **339:**82-85.

12. **Woolstenhulme CJ, Parajuli S, Healey DW, Valverde DP, Petersen EN, Starosta AL, Guydosh NR, Johnson WE, Wilson DN, Buskirk AR.** 2013. Nascent peptides that block protein synthesis in bacteria. Proc Natl Acad Sci USA **110:**E878-887.

13. **Elgamal S, Katz A, Hersch SJ, Newsom D, White P, Navarre WW, Ibba M.** 2014. EF-P dependent pauses integrate proximal and distal signals during translation. PLoS Genet **10:**e1004553.

14. **Starosta AL, Lassak J, Peil L, Atkinson GC, Virumae K, Tenson T, Remme J, Jung K, Wilson DN.** 2014. Translational stalling at polyproline stretches is modulated by the sequence context upstream of the stall site. Nucleic Acids Res **42:**10711-10719.

15. **Woolstenhulme CJ, Guydosh NR, Green R, Buskirk AR.** 2015. High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. Cell Rep **11:**13-21.

16. **Qi F, Motz M, Jung K, Lassak J, Frishman D.** 2018. Evolutionary analysis of polyproline motifs in *Escherichia coli* reveals their regulatory role in translation. PLoS Comput Biol **14:**e1005987.

17. **Adzhubei AA, Sternberg MJ, Makarov AA.** 2013. Polyproline-II helix in proteins: structure and function. J Mol Biol **425:**2100-2132.

18. **Starosta AL, Lassak J, Peil L, Atkinson GC, Woolstenhulme CJ, Virumae K, Buskirk A, Tenson T, Remme J, Jung K, Wilson DN.** 2014. A conserved proline triplet in Val-tRNA Synthetase and the origin of elongation factor P. Cell Rep **9:**476-483.

19. **Blaha G, Stanley RE, Steitz TA.** 2009. Formation of the first peptide bond: the structure of EF-P bound to the 70S ribosome. Science **325:**966-970.

20. **Saini P, Eyler DE, Green R, Dever TE.** 2009. Hypusine-containing protein eIF5A promotes translation elongation. Nature **459:**118-121.

21. **Melnikov S, Mailliot J, Rigger L, Neuner S, Shin BS, Yusupova G, Dever TE, Micura R, Yusupov M.** 2016. Molecular insights into protein synthesis with proline residues. EMBO Rep **17:**1776-1784.

22. **Melnikov S, Mailliot J, Shin BS, Rigger L, Yusupova G, Micura R, Dever TE, Yusupov M.** 2016. Crystal Structure of Hypusine-Containing Translation Factor eIF5A Bound to a Rotated Eukaryotic Ribosome. J Mol Biol **428:**3570-3576.

23. **Schmidt C, Becker T, Heuer A, Braunger K, Shanmuganathan V, Pech M, Berninghausen O, Wilson DN, Beckmann R.** 2016. Structure of the hypusinylated eukaryotic translation factor eIF-5A bound to the ribosome. Nucleic Acids Res **44:**1944-1951.

24. **Huter P, Arenz S, Bock LV, Graf M, Frister JO, Heuer A, Peil L, Starosta AL, Wohlgemuth I, Peske F, Novacek J, Berninghausen O, Grubmuller H, Tenson T, Beckmann R, Rodnina MV, Vaiana AC, Wilson DN.** 2017. Structural basis for polyproline-mediated ribosome stalling and rescue by the translation elongation factor EF-P. Mol Cell **68:**515-527 e516.

25. **Hanawa-Suetsugu K, Sekine S, Sakai H, Hori-Takemoto C, Terada T, Unzai S, Tame JR, Kuramitsu S, Shirouzu M, Yokoyama S.** 2004. Crystal structure of elongation factor P from *Thermus thermophilus* HB8. Proc Natl Acad Sci USA **101:**9595-9600.

26. **Katz A, Solden L, Zou SB, Navarre WW, Ibba M.** 2014. Molecular evolution of protein-RNA mimicry as a mechanism for translational control. Nucleic Acids Res **42:**3261-3271.

27. **Katoh T, Wohlgemuth I, Nagano M, Rodnina MV, Suga H.** 2016. Essential structural elements in tRNA^{Pro} for EF-P-mediated alleviation of translation stalling. Nat Commun **7:**11657.

28. **Lassak J, Keilhauer EC, Fürst M, Wuichet K, Gödeke J, Starosta AL, Chen JM, Søgaard-Andersen L, Rohr J, Wilson DN, Häussler S, Mann M, Jung K.** 2015. Arginine-rhamnosylation as new strategy to activate translation elongation factor P. Nat Chem Biol **11:**266-270.

29. **Lassak J, Wilson DN, Jung K.** 2016. Stall no more at polyproline stretches with the translation elongation factors EF-P and IF-5A. Mol Microbiol **99:**219-235.

30. **Bailly M, de Crecy-Lagard V.** 2010. Predicting the pathway involved in post-translational modification of elongation factor P in a subset of bacterial species. Biol Direct **5:**3.

31. **Navarre WW, Zou SB, Roy H, Xie JL, Savchenko A, Singer A, Edvokimova E, Prost LR, Kumar R, Ibba M, Fang FC.** 2010. PoxA, YjeK, and elongation factor P coordinately modulate virulence and drug resistance in *Salmonella enterica*. Mol Cell **39:**209-221.

32. **Yanagisawa T, Sumida T, Ishii R, Takemoto C, Yokoyama S.** 2010. A paralog of lysyl-tRNA synthetase aminoacylates a conserved lysine residue in translation elongation factor P. Nat Struct Mol Biol **17:**1136-1143.

33. **Behshad E, Ruzicka FJ, Mansoorabadi SO, Chen D, Reed GH, Frey PA.** 2006. Enantiomeric free radicals and enzymatic control of stereochemistry in a radical mechanism: the case of lysine 2,3-aminomutases. Biochemistry **45:**12639-12646.

34. **Roy H, Zou SB, Bullwinkle TJ, Wolfe BS, Gilreath MS, Forsyth CJ, Navarre WW, Ibba M.** 2011. The tRNA synthetase paralog PoxA modifies elongation factor-P with (R)-beta-lysine. Nat Chem Biol **7:**667-669.

35. **Peil L, Starosta AL, Virumae K, Atkinson GC, Tenson T, Remme J, Wilson DN.** 2012. Lys34 of translation elongation factor EF-P is hydroxylated by YfcM. Nat Chem Biol **8:**695-697.

36. **Kobayashi K, Katz A, Rajkovic A, Ishii R, Branson OE, Freitas MA, Ishitani R, Ibba M, Nureki O.** 2014. The non-canonical hydroxylase structure of YfcM reveals a metal ion-coordination motif required for EF-P hydroxylation. Nucleic Acids Res **42:**12295-12305.

37. **Bullwinkle TJ, Zou SB, Rajkovic A, Hersch SJ, Elgamal S, Robinson N, Smil D, Bolshan Y, Navarre WW, Ibba M.** 2013. (*R*)-beta-lysine-modified elongation factor P functions in translation elongation. J Biol Chem **288:**4416-4423.

38. **Rajkovic A, Hummels KR, Witzky A, Erickson S, Gafken PR, Whitelegge JP, Faull KF, Kearns DB, Ibba M.** 2016. Translation control of swarming proficiency in *Bacillus subtilis* by 5-amino-pentanolylated elongation factor P. J Biol Chem **291:**10976-10985.

39. **Rajkovic A, Erickson S, Witzky A, Branson OE, Seo J, Gafken PR, Frietas MA, Whitelegge JP, Faull KF, Navarre W, Darwin AJ, Ibba M.** 2015. Cyclic Rhamnosylated Elongation Factor P Establishes Antibiotic Resistance in Pseudomonas aeruginosa. MBio **6:**e00823.

40. **Yanagisawa T, Takahashi H, Suzuki T, Masuda A, Dohmae N, Yokoyama S.** 2016. *Neisseria meningitidis* translation elongation factor P and its active-site arginine residue are essential for cell viability. PLoS ONE **11:**e0147907.

41. **Krafczyk R, Macosek J, Jagtap PKA, Gast D, Wunder S, Mitra P, Jha AK, Rohr J, Hoffmann-Roder A, Jung K, Hennig J, Lassak J.** 2017. Structural Basis for EarP-Mediated Arginine Glycosylation of Translation Elongation Factor EF-P. mBio **8:**e01412-01417.

42. **Sengoku T, Suzuki T, Dohmae N, Watanabe C, Honma T, Hikida Y, Yamaguchi Y, Takahashi H, Yokoyama S, Yanagisawa T.** 2018. Structural basis of protein arginine rhamnosylation by glycosyltransferase EarP. Nat Chem Biol **14:**368-374.

43. **Coutinho PM, Deleury E, Davies GJ, Henrissat B.** 2003. An evolving hierarchical family classification for glycosyltransferases. J Mol Biol **328:**307-317.

44. **Witzky A, Hummels KR, Tollerson R, 2nd, Rajkovic A, Jones LA, Kearns DB, Ibba M.** 2018. EF-P posttranslational modification has variable impact on polyproline translation in *Bacillus subtilis*. mBio **9:**e00306-00318.

45. **Revell LJ.** 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol **3:**217-223.

46. **Dever TE, Gutierrez E, Shin BS.** 2014. The hypusine-containing translation factor eIF5A. Crit Rev Biochem Mol Biol **49:**413-425.

47. **Buchner S, Schlundt A, Lassak J, Sattler M, Jung K.** 2015. Structural and functional analysis of the signal-transducing linker in the pH-responsive one-component system CadC of *Escherichia coli*. J Mol Biol **427:**2548-2561.

48. **Schlundt A, Buchner S, Janowski R, Heydenreich T, Heermann R, Lassak J, Geerlof A, Stehle R, Niessing D, Jung K, Sattler M.** 2017. Structure-function analysis of the DNA-binding domain of a transmembrane transcriptional activator. Sci Rep **7:**1051.

49. **Choi S, Choe J.** 2011. Crystal structure of elongation factor P from Pseudomonas aeruginosa at 1.75 A resolution. Proteins **79:**1688-1693.

50. **Karimova G, Pidoux J, Ullmann A, Ladant D.** 1998. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proc Natl Acad Sci USA **95:**5752-5756.

51. **Hummels KR, Witzky A, Rajkovic A, Tollerson R, 2nd, Jones LA, Ibba M, Kearns DB.** 2017. Carbonyl reduction by YmfI in *Bacillus subtilis* prevents accumulation of an inhibitory EF-P modification state. Mol Microbiol **106:**236-251.

52. **Brochier C, Lopez-Garcia P, Moreira D.** 2004. Horizontal gene transfer and archaeal origin of deoxyhypusine synthase homologous genes in bacteria. Gene **330:**169-176.

53. **Wolff EC, Park MH, Folk JE.** 1990. Cleavage of spermidine as the first step in deoxyhypusine synthesis. The role of NAD. J Biol Chem **265:**4793-4799.

54. **Prunetti L, Graf M, Blaby IK, Peil L, Makkay AM, Starosta AL, Papke RT, Oshima T, Wilson DN, de Crecy-Lagard V.** 2016. Deciphering the Translation Initiation Factor 5A Modification Pathway in Halophilic Archaea. Archaea **2016:**7316725.

55. **Schuller AP, Wu CC, Dever TE, Buskirk AR, Green R.** 2017. eIF5A functions globally in translation elongation and termination. Mol Cell **66:**194-205 e195.

56. **Pelechano V, Alepuz P.** 2017. eIF5A facilitates translation termination globally and promotes the elongation of many non polyproline-specific tripeptide sequences. Nucleic Acids Res **45:**7326-7338.

57. **Charbonneau ME, Cote JP, Haurat MF, Reiz B, Crepin S, Berthiaume F, Dozois CM, Feldman MF, Mourez M.** 2012. A structural motif is the recognition site for a new family of bacterial protein *O*-glycosyltransferases. Mol Microbiol **83:**894-907.

58. **Katoh K, Standley DM.** 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol **30:**772-780.

59. **Price MN, Dehal PS, Arkin AP.** 2010. FastTree 2-approximately maximum-likelihood trees for large alignments. PLoS ONE **5:**e9490.

60. **Volkwein W, Maier C, Krafczyk R, Jung K, Lassak J.** 2017. A versatile toolbox for the control of protein levels using N^ϵ-acetyl-L-lysine dependent amber suppression. ACS Synth Biol **6:**1892-1902.

61. **Fried L, Lassak J, Jung K.** 2012. A comprehensive toolbox for the rapid construction of *lacZ* fusion reporters. J Microbiol Methods **91:**537-543.

62. **Espah Borujeni A, Channarasappa AS, Salis HM.** 2014. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. Nucleic Acids Res **42:**2646-2659.

63. **Salis HM, Mirsky EA, Voigt CA.** 2009. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol **27:**946.

64. **Tetsch L, Koller C, Haneburger I, Jung K.** 2008. 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.

65. **Miller JH.** 1992. A short course in bacterial genetics: a laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory, Cold Spring Harbor N.Y.
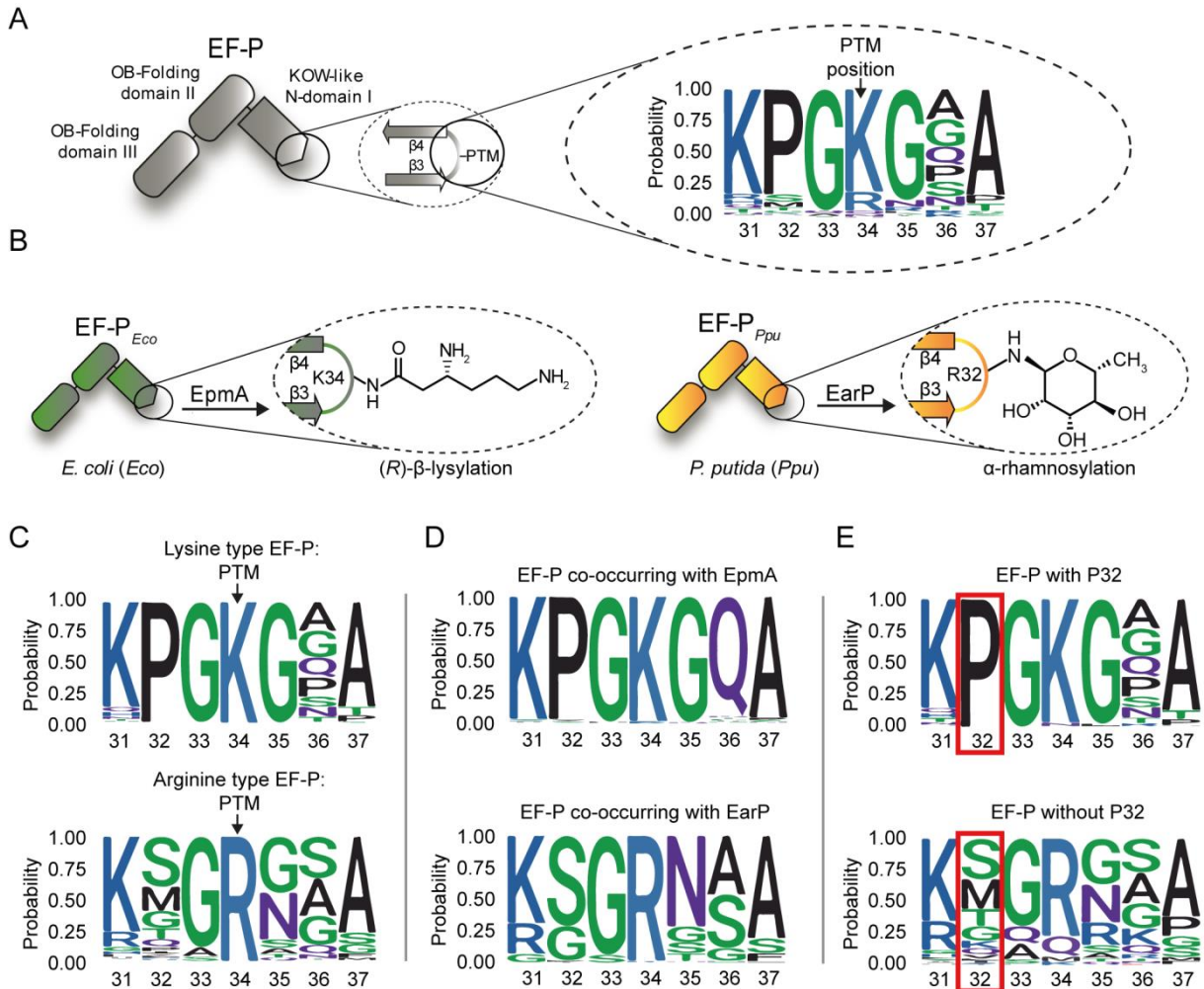
Figures and legends

Figure 1. Composition of β3Ωβ4 of the EFP_N superfamily (A) Domain architecture of EF-P: The first magnification illustrates β3Ωβ4 of the KOW-like EF-P_N-domain I including the post-translational modification (PTM) site. The second magnification depicts the general amino acid composition of the seven amino acid long loop between the strands β3 and β4 based on 4421 sequences. (B) PTMs present either in *E. coli* EF-P (left, (R)-β-lysylation) or *P. putida* EF-P (right, α-rhamnosylation). The modifying enzymes EpmA and EarP are indicated. (C) Weblogo of β3Ωβ4 for the EF-Ps with lysine (lysine type EF-P) or an arginine (arginine type EF-P) at position 34 according to the numbering of the *E. coli* ortholog. (D) Weblogo of β3Ωβ4 for the EF-Ps of bacteria co-occurring with EpmA (upper logo) or EarP (lower logo). (E) Weblogo of β3Ωβ4 for the EF-Ps containing (upper logo) or not containing (lower logo) a proline at position 32 according to the numbering of the *E. coli* ortholog.
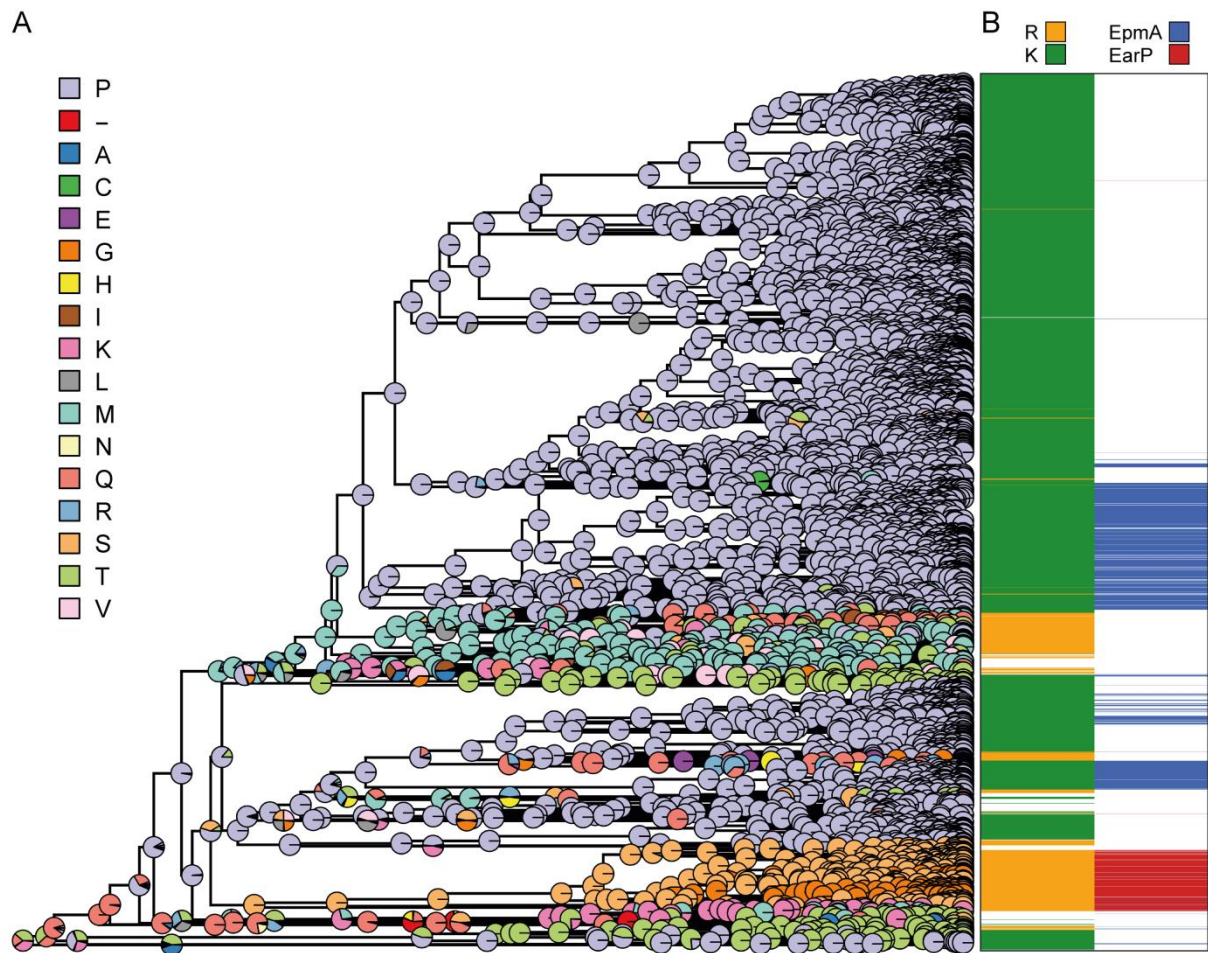
Figure 2. EF-P phylogeny and β3Ωβ4 sequence evolution (A) Phylogenetic tree reconstruction of the 32nd position (-2 relative to the β3Ωβ4 tip) of the EF-P KOW-like N-domain I: The colored circle areas correspond to the likelihood of a certain amino acid (see color code legend) to occupy the position at the phylogenetic knot. (B) Colors in the left bar indicate the amino acid located at the 34th position (β3Ωβ4 tip). Colors in the right bar indicate the corresponding type of modification system. Each colored line corresponds to one bacterial species in the EF-P data set.
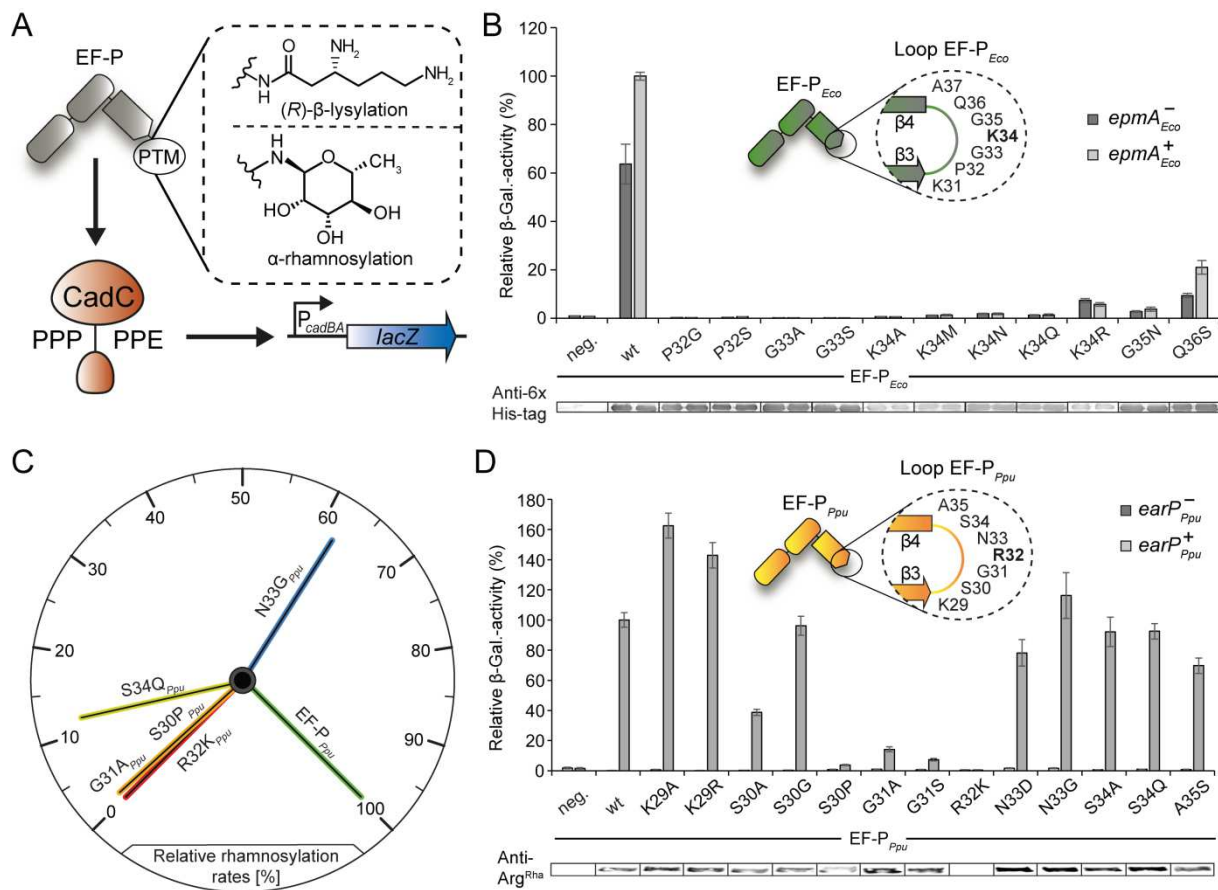
Figure 3. Characterization of single amino acid substitutions in EF-P β3Ωβ4 of *E. coli* and *P. putida* (A) LacZ based *E. coli* reporter for the detection of EF-P functionality: Post-translationally modified EF-P ((R)-β-lysylated or α-rhamnosylated) promotes production of the polyproline containing pH-sensor CadC. At mild acidic pH and the concomitant presence of lysine CadC induces *lacZ* expression by activation of the cadBA promoter (B) *In vivo* activity measurements of *E. coli* EF-P β3Ωβ4 variants using the *E. coli* reporter strain described in (A). Here, either efp (*E. coli* MG-CR-*efp*-KanS, *epmA+Eco*, dark grey bars) or both, *efp* as well as *epmA* were deleted (*E. coli* MG-CR-*efp*-*epmA*-KanR, *epmA-Eco*, light grey bars) and complemented with a plasmid encoded *E. coli* His6-tagged EF-P variant. β-galactosidase activities are given in relative Miller units with the wild-type EF-P*Eco* in the *epmA+* background set to 100%. EF-P production was confirmed by Western blot analysis using Anti-His6. (C) *In vivo* activity analysis of *P. putida* EF-P β3Ωβ4 using the *E. coli* reporter strain described in (A) in which *efp* is deleted (*E. coli* MG-CR-*efp*-KanS). *E. coli* Δ*efp* cells were complemented either with plasmid encoded *P. putida* His6-tagged EF-P variant solely (*earP-Ppu*, dark grey bars) or in combination with *EarPPpu* (*earP+Ppu*, light grey bars). β-galactosidase activities are given in relative Miller units with the wild-type EF-P*Ppu* in *earP+* cells set to 100%. Rhamnosylation of EF-P was confirmed by Western blot analysis using Anti-rhamnosylarginine antibody (Anti-Arg[Rha]). (D) Relative *in vitro* rhamnosylation rates of EF-P*Ppu* wild type (set to 100%) and amino acid substitution variants.
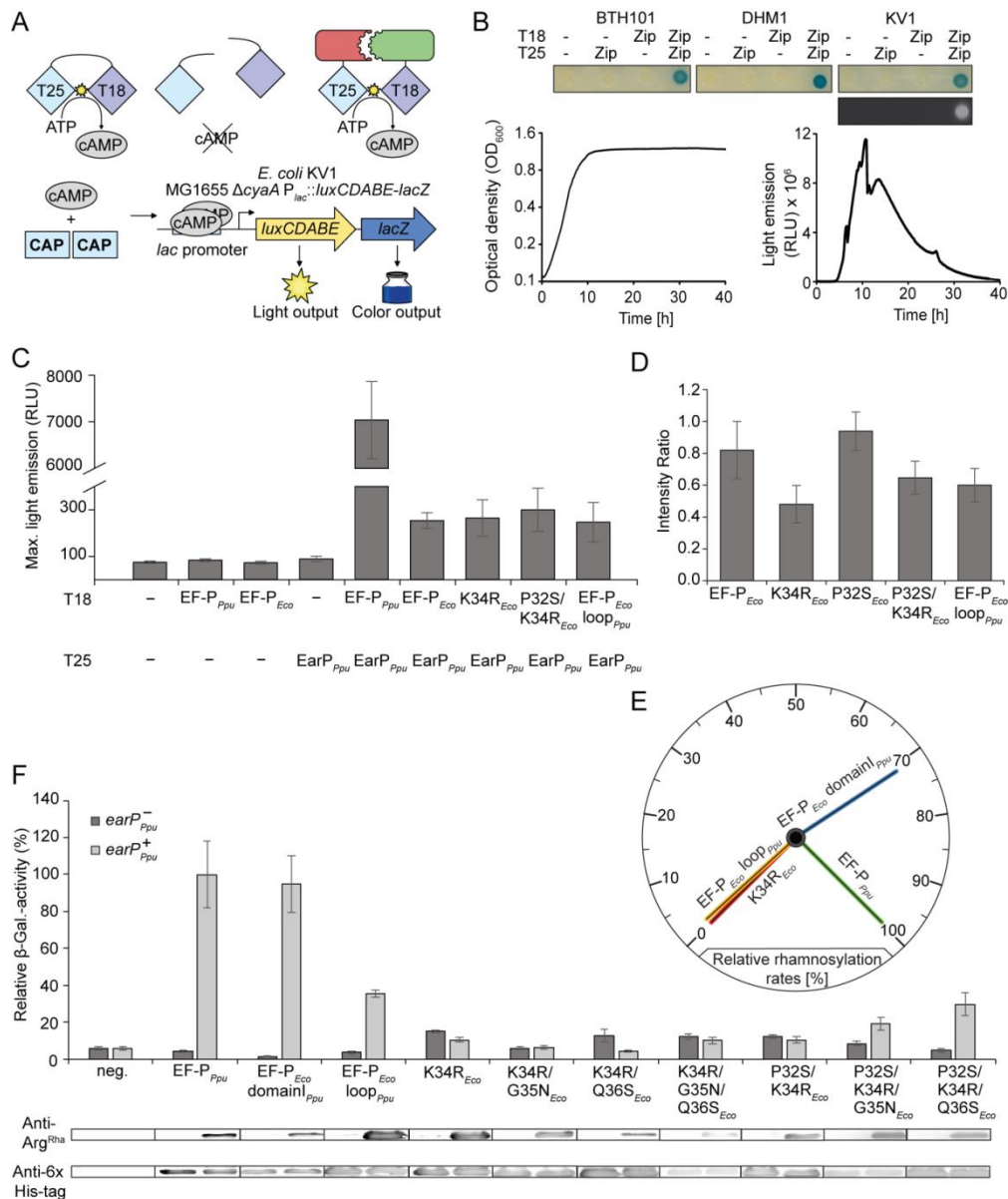
Figure 4. Cross-modification and activation of *E. coli* EF-P by EarP of *P. putida* (A) Molecular principle of the Euromedex bacterial two-hybrid system with *E. coli* KV1 (adjusted from the Euromedex KIT Manual): The *Bordetella pertussis* adenylate cyclase CyaA apoprotein catalyzes the formation of cyclic AMP from ATP. Splitting the enzyme into two parts – T18 and T25 – renders CyaA inactive even upon co-expression. Fusing T18 and T25 to interacting proteins brings the fragments into close proximity and thus allows reconstitution of the Apo-CyaA. In *E. coli* KV1, the cyclic AMP dependent lac promoter precedes a translational fusion of the *lux*-operon and *lacZ*, allowing the indirect measurement of protein-protein interactions by light emission and colorimetric detection. (B) Proof of principle of *E. coli* KV1 as an *in vivo* reporter for protein-protein interaction using the self-interacting leucine zipper of GCN4. The colorimetric detection in *E. coli* KV1 cells was assessed semi quantitatively based on the formation of blue colonies on LB (Miller)-plates containing 40 µg/mL X-Gal. *E. coli* KV1 light emission was measured in a time course experiment recording relative luminescence (RLU) and optical density at 600 nm (OD600) in time intervals of 10 minutes. (C) *In vivo* interaction analysis of a T25-EarP$_{Ppu}$ fusion or T25 solely with T18 fusions to cognate (EF-P$_{Ppu}$), and non-cognate EF-Ps (EF-P$_{Eco}$) as well as its corresponding substitution variants. The maximal light emission from a 40 h time course experiment is given in RLU. 95 % confidence intervals of at least 6 replicates are shown. (D) NMR-single titration experiment: Average change in the intensity ratio of EF-P$_{Eco}$ and its variants upon titration with 2fold EarP$_{Ppu}$. Error bars represent standard deviation of the intensity ratios within each titration. (E) Relative *in vitro* rhamnosylation rates of EF-P$_{Ppu}$ (set to 100%) and EF-P$_{Eco}$ variants. (F) *In vivo* rhamnosylation and functionality analysis of EF-P hybrids co-expressed with EarP from *P. putida*. Measurements were performed in the *E. coli* reporter strain (MG-CR-efp-KanS). *E. coli* Δ*efp* cells were complemented either with plasmid encoded His6-tagged EF-P$_{Eco/Ppu}$ variants solely (*earP-Ppu*, dark grey bars) or in combination with EarP$_{Ppu}$ (*earP+Ppu*, light grey bars). β-galactosidase activities are given in relative Miller units with the wild-type EF-P$_{Ppu}$ in *earP+* cells set to 100%. Rhamnosylation of EF-P was confirmed by Western blot analysis using Anti-Arg^{Rha}. In the hybrid, EF-P$_{Eco}$ domainI$_{Ppu}$, the EF-P KOW-like N-domain I of *E. coli* was swapped with the one from *P. putida*. The hybrid, EF-P$_{Eco}$ loop$_{Ppu}$, was generated by replacing β3Ωβ4 from *E. coli* by the corresponding β3Ωβ4 from *P. putida*.

60

5 A versatile toolbox for the control of protein levels using N$^{\varepsilon}$-acetyl-l-lysine dependent amber suppression

Volkwein W, Maier C, <u>Krafczyk R</u>, Jung K, Lassak J. 2017. A versatile toolbox for the control of protein levels using N$^{\varepsilon}$-acetyl-l-lysine dependent amber suppression. ACS Synth Biol 6:1892-1902.

A Versatile Toolbox for the Control of Protein Levels Using N^ε-Acetyl-L-lysine Dependent Amber Suppression

Wolfram Volkwein, Christopher Maier, Ralph Krafczyk, Kirsten Jung,*[ID] and Jürgen Lassak*[ID]

Center for integrated Protein Science Munich (CiPSM) at the Department of Biology I, Microbiology, Ludwig-Maximilians-Universität München, Großhaderner Strasse 2-4, 82152 Martinsried, Germany

Ⓢ *Supporting Information*

ABSTRACT: The analysis of the function of essential genes *in vivo* depends on the ability to experimentally modulate levels of their protein products. Current methods to address this are based on transcriptional or post-transcriptional regulation of mRNAs, but approaches based on the exploitation of translation regulation have so far been neglected. Here we describe a toolbox, based on amber suppression in the presence of N^ε-acetyl-L-lysine (AcK), for translational tuning of protein output. We chose the highly sensitive luminescence system LuxCDABE as a reporter and incorporated a UAG stop codon into the gene for the reductase subunit LuxC. The system was used to measure and compare the effects of AcK- and N^ε-(*tert*-butoxycarbonyl)-L-lysine (BocK) dependent amber suppression in *Escherichia coli*. We also demonstrate here that, in combination with transcriptional regulation, the system allows protein production to be either totally repressed or gradually adjusted. To identify sequence motifs that provide improved translational regulation, we varied the sequence context of the amber codon and found that insertion of two preceding prolines drastically decreases luminescence. In addition, using LacZ as a reporter, we demonstrated that a strain encoding a variant with a Pro-Pro amber motif can only grow on lactose when AcK is supplied, thus confirming the tight translational regulation of protein output. In parallel, we constructed an *E. coli* strain that carries an isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible version of the AcK-tRNA synthetase (AcKRS) gene on the chromosome, thus preventing mischarging of noncognate substrates. Subsequently, a diaminopimelic acid auxotrophic mutant ($\Delta dapA$) was generated demonstrating the potential of this strain in regulating essential gene products. Furthermore, we assembled a set of vectors based on the broad-host-range pBBR *ori* that enable the AcK-dependent amber suppression system to control protein output not only in *E. coli*, but also in *Salmonella enterica* and *Vibrio cholerae*.

KEYWORDS: *translational regulation, gene silencing, PylRS, Salmonella enterica, Vibrio cholerae, genetic screen*
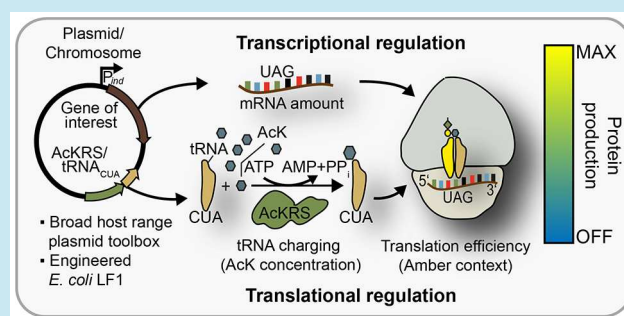
Most gene silencing strategies target gene expression at the level of transcription by exploiting the activity of transcriptional regulators, *e.g.*, TetR,[1] LacI[2] or the CRISPRi system.[3] Alternatively, gene expression can be tightly regulated post-transcriptionally by the use of, *e.g.*, microRNAs or small interfering RNAs.[4] Here, we show that translational regulation can also be employed to sensitively adjust protein levels.

Canonically, 20 different amino acids can be utilized for polypeptide synthesis by the translational apparatus. However, the genetic code can be translationally expanded to direct the incorporation of two natural but noncanonical amino acids (ncAAs), namely selenocysteine (Sec)[8] and pyrrolysine (Pyl).[9] Both Sec-tRNA and Pyl-tRNA recognize codons—UGA and UAG, respectively—that are usually reserved for translation termination. Whereas Sec is found in a variety of proteins in all domains of life,[10,11] Pyl is restricted to a small number of proteins in a few archaeal and bacterial species.[12] In methanogenic archaea such as *Methanosarcina barkeri*, the pyrrolysyl tRNA-synthetase (PylRS) charges its cognate tRNA_CUA (encoded by *pylT*) (Figure 1A) to suppress the amber stop codon UAG.[13] Due to its substrate promiscuity and its narrow distribution, PylRS has become a powerful means of expanding the genetic code to include synthetic ncAAs.[14] Today, over 150 substrates can be incorporated into proteins by means of amber suppression.[15,16] This in turn has enabled researchers not only to investigate post-translational modifications,[17] but also to introduce fluorescent labels,[18] and ultimately allows one to engineer proteins with modified functions. In 2015 Rovner *et al.* introduced the orthogonal translation system consisting of an aminoacyl-tRNA synthetase and its cognate UAG-reading tRNA from *Methanococcus janaschii*, which is optimized for the incorporation of *p*-acetyl-L-phenylalanine, into an *E. coli* MG1655 derivative.[19] With this strain, they found that three amber mutations, one in each of the open reading frames of the essential genes *murG, dnaA* and *serS*, can be rescued by an external supply of 4-acetyl-L-phenylalanine. Along the same line Yusuke Kato and co-workers demonstrated
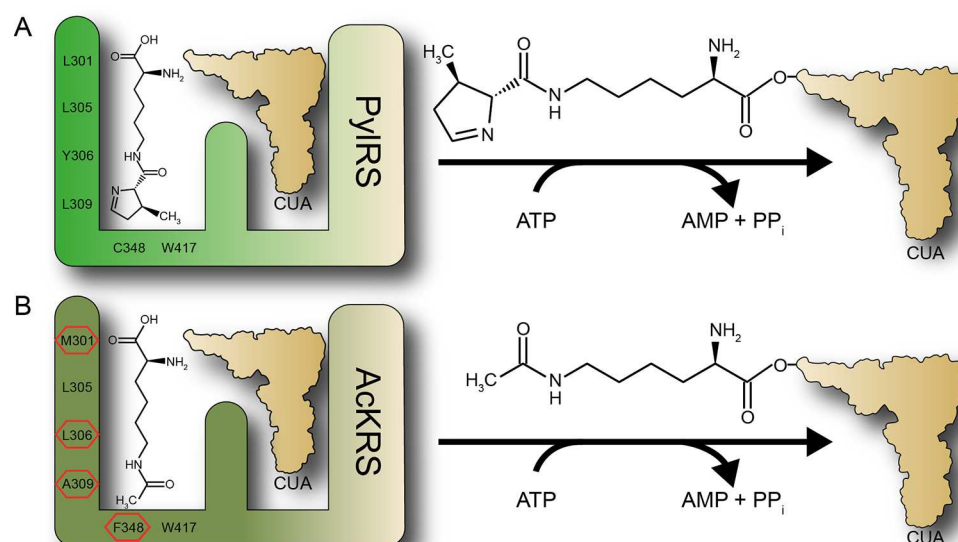
Figure 1. PylRS and AcKRS based amber suppression. (A) The pyrrolysyl-tRNA synthetase (PylRS) of *Methanosarcina mazei* loads its natural substrate pyrrolysine onto its cognate suppressor tRNA$_{CUA}$ (encoded by *pylT*). The six amino acids that form the pyrrole ring binding pocket are indicated.[5,6] (B) Acetyl lysyl-tRNA synthetase (AcKRS), a PylRS variant bearing the highlighted amino acids at positions 301, 306, 309 and 348, which enable it to interact specifically with N^ε-acetyl-L-lysine (AcK).[7]

the tunable translational control using site-specific 3-iodo-L-tyrosine incorporation in *Escherichia coli*.[20,21] Thus, the amber suppression system can be used to efficiently deplete essential gene products at the translational level.

In the present study we describe the assembly of a versatile toolbox for control of protein levels in diverse bacterial species using N^ε-acetyl-L-lysine (AcK)-dependent amber suppression.

RESULTS AND DISCUSSION

Selection of N^ε-Acetyl-L-lysine for Amber Suppression. Our study began with the selection of a suitable substrate to be incorporated by PylRS-mediated amber suppression. We chose N^ε-acetyl-L-lysine (AcK) for this purpose because an evolved variant of PylRS has been reported to charge tRNA$_{CUA}$ with AcK,[7,22] and AcK occurs frequently in bacterial proteins as a product of the post-translational modification of the ε-amino group of lysine residues.[23−26] Lysine acetylation can be catalyzed by protein acetylases such as YfiQ (PatZ) in *E. coli*,[27] but can also result from a nonenzymatic reaction with acetyl phosphate.[24,26] Note that protein acetylation is not a permanent modification, but can be reversed by deacetylases, *e.g.*, CobB, in bacteria.[27] This would allow the production of an unaltered protein when a native lysine codon is replaced by UAG and read by AcK-dependent amber suppression.[22,28] Apart from the fact that AcK is found naturally in proteins, it has no effect on bacterial growth (Supplementary Figure S1), is commercially available and is relatively cheap. For example, supplementing 1 L of culture medium with 1 mM AcK costs around 5€, and is thus about as expensive as IPTG. Taken together, these considerations make AcK an ideal substrate. We used the previously described PylRST expression system from *Methanosarcina mazei*[29] and introduced the mutations reported by Umehara *et al.*[7] to render the enzyme specific for AcK, resulting in the N^ε-acetyl lysyl-tRNA synthetase/tRNA$_{CUA}$ pair (AcKRST) (Figure 1B).

Development of a Luminescence-Based Amber Suppression Reporter. Amber suppression is conventionally

quantified using either green fluorescent protein (GFP)[30] or β-galactosidase (LacZ)[7] as the reporter. We decided to use luminescence as the readout, and chose the system based on expression of the *luxCDABE* operon from *Photorhabdus luminescens* (Lux) (Figure 2A, B). In this system *luxA* and *luxB* encode the α- and β-subunits of the luciferase heterodimer (Figure 2A), while the reductase, transferase and synthase specified by LuxC, D and E, respectively, together form the complex that synthesizes and regenerates the myristyl aldehyde substrate required for the bioluminescent reaction (Figure 2A).[31,32] The Lux system is highly sensitive and allows one to quantify amber suppression in living cells in real time.[33]

To make luminescence development not only dependent on amber suppression but also tunable at the transcriptional level we cloned the *luxCDABE* cassette into pBAD/HisA, placing it under the control of the L-arabinose (Ara)-inducible promoter P$_{BAD}$. The stop sequence TAG was inserted at position 3 in the *luxC* gene (Figure 2B)

We note that alterations at the 5′ end of coding sequences might affect translation efficiency due to changes in mRNA folding[34−36] and accordingly one should be careful when inserting TAG here. On the other hand, the region close to the protein start site often lacks structural features important for protein function and therefore tolerates manipulation without perturbation of functionality. In the case of LuxC, secondary structure prediction did not identify any α-helix or β-sheet in the first six amino acids and thus giving first hint for a tolerance toward insertions. Furthermore, there is relative low sequence conservation in this region and some LuxC homologues even have N-terminal extensions, *e.g.*, LuxC of *Photobacterium phosphoreum* altogether indicating this region to be suitable for TAG insertion.

We analyzed Lux-amber reporter (Lux^{Am}) activity in *E. coli* BW25113 which, in addition to pBAD/HisA-Lux(Am), harbors the vector pACYCDuet-AcKRST that codes for AcKRS and its cognate tRNA$_{CUA}$ (Figure 1B, 2C, Table 1). We grew the cells in LB and supplemented the medium with either 0.2% Ara, 1
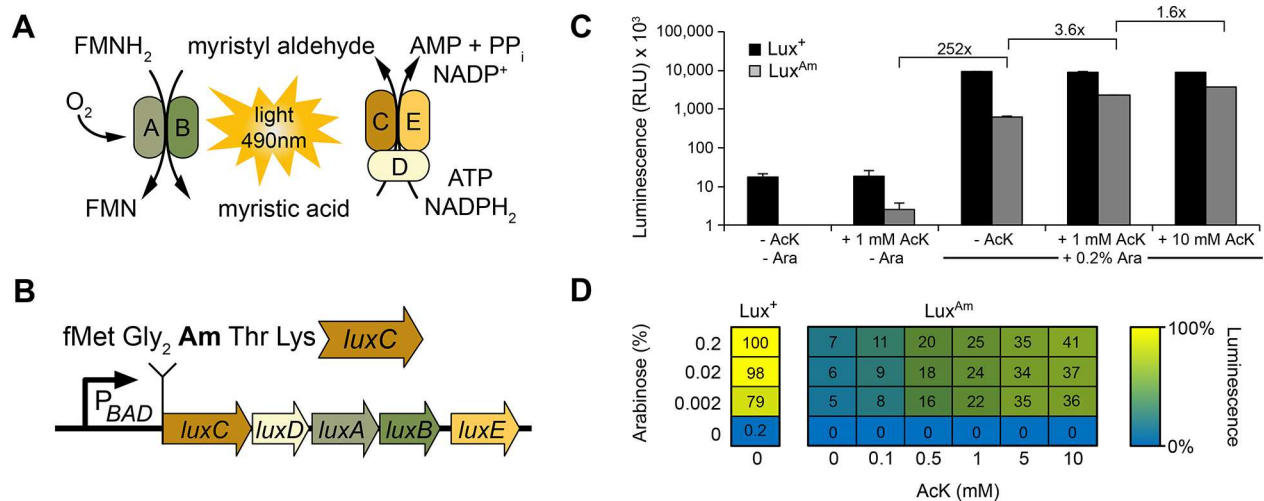
Figure 2. A luminescence-based amber suppression bioreporter (A) Basic principle of the luminescence system (Lux) from *Photorhabdus luminescens*. The *lux* genes are organized in the *luxCDABE* operon, which encodes the multienzyme complex LuxCDE and the luciferase heterodimer LuxAB. These two orchestrate substrate synthesis/regeneration of myristyl aldehyde and generate luminescence, respectively. (B) The Lux^{Am} luminescence reporter: The *luxCDABE* operon of *P. luminescens* was placed under the control of the Ara-inducible promoter P$_{BAD}$, and the third codon of the open reading frame of *luxC* was replaced by an amber stop codon, resulting in the Lux-amber reporter (Lux^{Am}). (C) Luminescence production of Lux^{Am} in *E. coli* BW25113. A construct in which the wildtype *luxC* ORF was left unaltered was used as a positive control (Lux$^+$): Cells harboring the corresponding Lux reporter plasmid in combination with pACYCDuet-AcKRST, which codes for the AcKRS and its cognate tRNA$_{CUA}$, were grown in LB overnight and luminescence production was monitored in response to Ara and/or AcK, and is given in relative light units (RLU). Error bars represent the standard deviation of data from three different experiments. (D) Parameter optimization for Lux$^+$- and Lux^{Am}-encoding *E. coli* strains. Ara and AcK concentrations were varied and maximal luminescence was determined. The resulting luminescence values ranged from "below detection" (0) RLU (blue) to 9.0×10^6 RLU (yellow).

mM AcK or both compounds together (Figure 2C). The Lux reporter plasmid pBAD/HisA-Lux, in which the wildtype *luxC* ORF was unchanged, served as a positive control (Lux$^+$). The maximal luminescence was determined and is given in relative light units (RLU). As expected, cells harboring pBAD/HisA-Lux produced light after induction with 0.2% Ara, irrespective of the presence or absence of AcK (Figure 2C). In contrast, in Lux^{Am}-expressing *E. coli* cells, luminescence was best stimulated by a combination of the two substances. However, under Ara-inducing conditions, we detected significant light emission (1×10^6 RLU) even in the absence of AcK. This result can be explained by translational read-through mediated by binding of near-cognate tRNAs[37] or by misaminoacylation of tRNA$_{CUA}$.[38]

Strikingly, while Lux$^+$-harboring cells showed a certain degree of light production in the absence of Ara, the Lux^{Am}-containing cells remained completely dark (Figure 2C). Thus, the combination of transcriptional and translational control completely abolishes protein output, in line with an earlier report.[19]

In the next step we systematically varied both the concentration of Ara and AcK and compared the luminescence of Lux^{Am} with Lux$^+$ cells (Figure 2D, Supplementary Figure S2). As before, light production by the Lux$^+$ strain was essentially unaffected by increasing concentrations of AcK (Supplementary Figure S2), and reached a maximum when the growth medium contained ≥0.02% Ara (Figure 2C, Supplementary Figure S2). In contrast, luminescence in the Lux^{Am} strain responds to the increasing presence of AcK, which allows for adjustment of the light output over a range that extends from "below detection" (blue) to more than 3.8×10^6 RLUs (green), corresponding to 41% of the maximal light intensity that we could measure with Lux$^+$ (Figure 2C, D, Supplementary Figure S2).

Additionally, we used the Lux^{Am} luminescence reporter to investigate the potential of N^ε-(*tert*-butoxycarbonyl)-L-lysine (BocK) as alternative substrate for translational protein level control. BocK can be incorporated into the genetic code employing the wildtype PylRST.[39] Like AcK, BocK is commercially available and comparably cheap. In contrast to AcKRS, PylRS is not promiscuous to charge tRNA$_{CUA}$ with natural amino acids,[7,40] and thus we were curious to compare the two aminoacyl-tRNA-synthetases in our test system with respect to misacylation and amber suppression efficiency. To this end we transformed *E. coli* BW25113 cells harboring pBAD/HisA-Lux(Am) in combination with pACYCDuet-PylRST,[29] pACYCDuet-AcKRST or the empty vector pACYCDuet-1, respectively, in order to discriminate between misacylation and translational read through. Subsequently, light production was analyzed in the resulting strains in the presence of 0.2% Ara with and without the supplement of 10 mM BocK and AcK, respectively (Supplementary Figure S3). Independent of PylRS and AcKRS we measured a base level luminescence of about 2.0×10^5 RLU. This value further increased in cells harboring pACYCDuet-AcKRST to about 5.0×10^5 RLU without the addition of AcK, indicating mischarging of tRNA$_{CUA}$ by endogenously available noncognate substrates. By contrast, in cells bearing pACYCDuet-PylRST we did not observe such an increase in luminescence clearly showing, that natural amino acids are not recognized by PylRS. Therefore, BocK might be a good alternative to AcKRS in order to control protein output translationally. Note that the BocK supplement resulted in an approximately 5-fold increase in luminescence in cells expressing *pylRS/pylT*. This was slightly (1.0×10^6 RLU) lower compared to the maximal value measured with AcKRST. Thus, and similar to AcK dependent amber suppression, the translational regulation with BocK/PylRST allows a tight

protein output control when combined with inducible gene expression.

Adjustment of the Amber Codon Context Permits Modulation of Protein Output. In the preceding section we showed that the Lux^{Am} strain produces significant levels of luminescence even in the absence of the cognate AcKRS substrate AcK (Figure 2C), which prompted us to ask whether an additional amber codon in the reporter gene might eliminate this background signal. Indeed, the introduction of a second TAG into *luxC* completely abolished light production, despite supplementation with AcK (Table 1). This implies that the second amber codon brings translation of *luxC* to a halt. We therefore undertook further engineering to enable suppression to compete with termination.[41,42]

Table 1. Influence of the 5′ Amber Context on Luminescence Production[a]

motif	AcK (1 mM)	mean × 10^3/SD × 10^3 (RLU)
Lux$^+$	−	7165 ± 392
Lux$^+$	+	7252 ± 486
Lux^{Am}	−	689 ± 26
Lux^{Am}	+	2406 ± 77
Lux^{2Am}	−	0
Lux^{2Am}	+	0
Lux$^{RA(Am)}$	−	919 ± 68
Lux$^{RA(Am)}$	+	3400 ± 22
Lux$^{DA(Am)}$	−	194 ± 33
Lux$^{DA(Am)}$	+	496 ± 51
Lux$^{HH(Am)}$	−	1434 ± 21
Lux$^{HH(Am)}$	+	3717 ± 633
Lux$^{KDP(Am)}$	−	20 ± 5
Lux$^{KDP(Am)}$	+	35 ± 26
Lux$^{PP(Am)}$	−	112 ± 35
Lux$^{PP(Am)}$	+	691 ± 72
Lux$^{KDPP(Am)}$	−	3 ± 3
Lux$^{KDPP(Am)}$	+	78 ± 7

[a]Cells harboring the corresponding Lux reporter plasmid in combination with pACYCDuet-AcKRST were grown in LB supplemented with 0.2% L-arabinose overnight and luminescence production was monitored in response to AcK and is given in relative light units (RLU).

It is known that the context of an internal stop codon can have a huge impact on translational readthrough.[43,44] This led us to analyze the influence of sequence motifs known to have significant effects on amber suppression. AcK incorporation into the nascent chain competes with translational termination mediated by the release factor RF1. This competition is reflected by the fact that even at saturating AcK concentrations the luminescence of Lux^{Am} cells does not reach the level observed in the Lux$^+$ *E. coli* strain (Figure 2C). However, the efficiency of recognition of the amber stop codon by RF1 is modulated by the base that follows it. While in *E. coli* a 3′ uracil best ensures termination, amber suppression is favored by a neighboring adenine.[45,46] The threonine sense codon ACU that follows the UAG in the Lux^{Am} reporter should therefore promote suppression and was left intact. Accordingly, we focused on the sequence context 5′ to the amber codon. Although the codon directly upstream of the stop site is of importance,[47] in 1994 Mottagui-Tabar and co-workers found that the penultimate codon has a major influence on termination in *E. coli*.[43] In that study, the authors described a

hierarchy of suppression, with arginine codons in this position showing the lowest and aspartate codons stimulating the highest suppression efficiency when UGA was the stop codon. On the basis of these findings, we decided to investigate the influence of R$_{-2}$ A$_{-1}$ and D$_{-2}$ A$_{-1}$ on amber suppression. We also included H$_{-2}$ H$_{-1}$, a context motif that was identified in a screen for efficient incorporation of ncAAs.[44] Moreover, we chose K$_{-3}$ D$_{-2}$ P$_{-1}$ and P$_{-2}$ P$_{-1}$, as well as a combination of the two (K$_{-4}$ D$_{-3}$ P$_{-2}$ P$_1$), to precede the stop site. These sequence contexts are known to cause ribosome stalling and should thus interfere with translation termination.[48−50]

All Lux^{Am} variants were constructed essentially as depicted in Figure 2B, by insertion of the motif after Gly$_2$ in the *luxC* gene. Luminescence was measured under P$_{BAD}$-inducing conditions as described above, and in the presence and absence of 1 mM AcK, respectively (Table 1).

As before (Figure 2C), the strain containing the Lux^{Am} reporter construct produced a maximal level of luminescence of about 2.4 × 10^6 RLU upon supplementation with AcK. This is a 3.5-fold increase relative to cells that were grown without AcK. In comparison, the maximal light intensities emitted by Lux-RA(Am) and Lux-DA(Am) cells grown in the presence of AcK were 3.4 × 10^6 and 5.0 × 10^5 RLU, respectively, and differed significantly from each other. This might be explained by altered *luxC* expression, as regulatory elements are located within the first 50 codons.[51] Regardless of these differences, the relative increases in luminescence in the presence of AcK were comparable, at 3.7-fold for Lux$^{RA(Am)}$ and 2.6-fold for Lux$^{DA(Am)}$. Mottagui-Tabar *et al.* showed that, in principle, arginine and aspartate in the second last position to UAG follow the same rules as when UGA is the stop codon, but that the effect with UAG was less pronounced.[43] This latter factor may account for the small differences in light production seen between amber-suppressed Lux$^{RA(Am)}$ and Lux$^{DA(Am)}$.

In agreement with the findings of Pott *et al.*, we observed increased luminescence for Lux$^{HH(Am)}$ compared to the original Lux^{Am} bioreporter.[44] However, this is largely attributable to increased translational read-through, as indicated by the increase in light production seen in the absence of AcK (Table 1). The analysis of amber context motifs that promote ribosome stalling—KDP, PP and KDPP—revealed a strong decrease in luminescence, which was unaffected by the presence of AcK. This decrease is presumably the result of the recruitment of ribosome rescue systems, such as tmRNA, which ultimately lead to abortion of translation.[49] Of the three arrest motifs tested, only the PP motif present in the Lux$^{PP(Am)}$ reporter had a significant effect on amber suppression. In this case, the addition of AcK to the growth medium was correlated with an increase of >6-fold in luminescence—the highest relative change observed with any of the motifs tested. Thus, a diprolyl amber context would appear to be especially beneficial in experiments where coupling with transcriptional regulation must be avoided.

A Diproline Amber Context Allows Tight Regulation of β-Galactosidase Protein Levels. We next tested whether the integration of an amber stop codon into the bacterial chromosome might allow for tighter regulation of protein production in *E. coli*. For these experiments, we chose the β-galactosidase LacZ, which cleaves the disaccharide lactose into D-galactose and D-glucose. This enzyme is essential for growth of *E. coli* cells on lactose as sole carbon source. We constructed two *E. coli* strains, bearing either an Am (TAG) or a PP(Am) (CCG CCG TAG) insertion after the ninth codon in the *lacZ*

coding sequence (Figure 3A). Similar to LuxC, our decision for the insertion position was based on a lack of structural elements as well as comparatively low sequence conservation in the first ten N-terminal amino acids of LacZ.[54] In addition to the
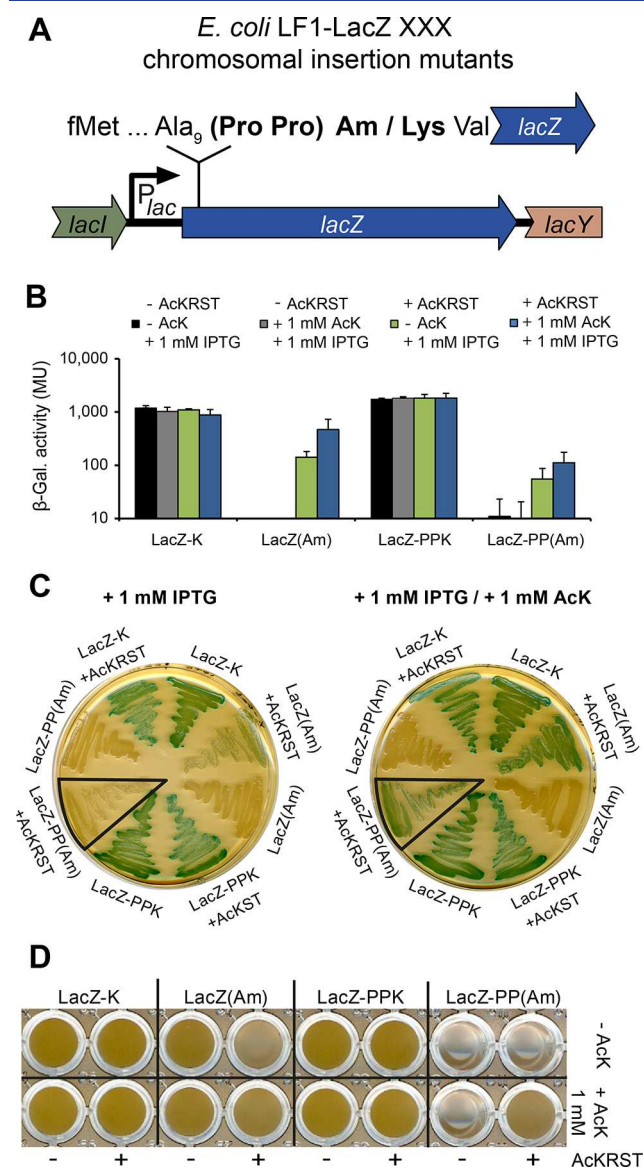

Figure 3. Regulation of LacZ levels in *E. coli* by amber suppression. (A) Genetic organization of the *lacZ* gene in *E. coli* LF1-LacZ strains. *E. coli* LF1 strains were constructed by inserting either an amber (LF1-LacZ(Am)) or PP amber sequence (LF1-LacZ-PP(Am)) after the ninth codon of the *lacZ* gene. Corresponding lysine codon insertions served as positive controls—LF1-LacZ-K and LF1-LacZ-PPK, respectively. (B) Quantitative analysis of amber suppression of β-galactosidase LacZ. To measure the β-galactosidase activity, cells were grown in LB overnight and reporter activity was determined according to Miller,[52] in the presence/absence of the amber suppression machinery (AcKRST). Cells were incubated with IPTG, AcK or a combination of both substances. The activity is given in Miller Units (MU). (C) Qualitative analysis of amber suppression of LacZ. *E. coli* LF1 strains described in (A) were grown on LB agar supplemented with 80 μM X-Gal, 1 mM IPTG and 1 mM AcK as indicated. (D) Growth analysis of *E. coli* LF1 strains with lactose as sole carbon source. LF1 strains with and without AcKRST were grown in M9 minimal medium[53] supplemented with 20 mM lactose in the presence/absence of 1 mM AcK in microtiter plates.

resulting strains LF1-LacZ(Am) and LF1-LacZ-PP(Am) and to further exclude any effects on protein activity we engineered two strains containing a lysine codon instead of TAG for use as positive controls (LF1-LacZ-K and LF1-LacZ-PPK). These two strains were compared with wildtype MG1655 cells and found to be phenotypically indistinguishable with respect to both growth on lactose and β-galactosidase activity (data not shown). By contrast, both LF1-LacZ(Am) and LF1-LacZ-PP(Am) exhibited strongly diminished LacZ production (Figure 3B). Moreover, whereas β-galactosidase activity was still detectable in LF1-LacZ(Am) in the absence of AcK (141 MU; MU = Miller Units), only background levels (55 MU) were measured in strain LF1-LacZ-PP(Am). As expected, addition of 1 mM AcK to the medium increased LacZ production in both cases, to 468 MU (LF1-LacZ(Am)) and 111 MU (LF1-LacZ-PP(Am)), respectively. In agreement with our previous finding for the luminescence reporter Lux[PP(Am)], the PP(Am) context resulted in a more pronounced decrease in translational output, confirming that it further reduces leakiness in the system. The virtually complete abrogation of LacZ production is further illustrated by the fact that the color change from white to blue of *E. coli* colonies encoding LacZ-PP(Am) is strictly dependent on AcK (Figure 3C). In contrast, the LF1-LacZ(Am) strain turns blue when grown on LB/X-Gal agar plates containing the *lacZ* inducer IPTG (Figure 3C). Furthermore, LF1-LacZ-PP(Am) containing AcKRST was only able to grow on lactose when AcK was concomitantly supplied, while the residual β-galactosidase activity in LF1-LacZ(Am) cells (Figure 3B) was sufficient to promote growth even in the absence of AcK (Figure 3D). Thus, these data further corroborate our findings with the Lux reporters (Figure 2C, D; Table 1).

The ability to tightly control LacZ synthesis in this way demonstrates the potential of our system for AcK-dependent translational regulation of essential genes with a specific amber context.

Chromosomal Integration of AcKRS Enhances Regulated Amber Suppression. Having shown that amber suppression is a potent molecular tool for the regulation of translational output, we set out to engineer *E. coli* in such a way that it can serve as a host for investigation of the function of essential genes in an AcK-dependent manner. In a previous publication we reported the construction of a strain named LF1, into which we had inserted a kanamycin resistance cassette and a wildtype copy of *rpsL* (conferring streptomycin sensitivity) into the *lac* locus.[55]

The two genes are flanked by FRT sites that mediate Red/ET (Gene Bridges) recombination, which permits scarless insertion of any DNA sequence of interest within 1 day. We used this method to integrate *pylS*/*pylT* into the LF1 strain. While *pylS* was set under the control of P$_{lac}$, *pylT* (tRNA$_{CUA}$) expression was kept constitutively expressed employing P$_{proK}$. The genomic context of the resultant insert in the new strain LF1-AcKRST is depicted in Figure 4A. To test amber suppression in *E. coli* LF1-AcKRST, the cells were transformed with pBAD/HisA-Lux(Am), and luminescence was monitored in response to the external supply of IPTG and AcK. As expected, no luminescence was detectable in the absence of both substances. Similarly, supplementation with 1 mM AcK alone did not lead to any measurable light production, while addition of IPTG to the medium on its own resulted in only very weak luminescence of approximately 1.0×10^4 RLU. In stark contrast, the simultaneous presence of AcK and IPTG
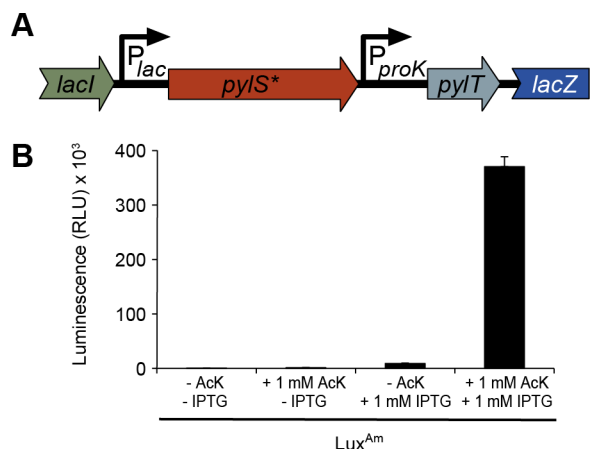
Figure 4. *E. coli* LF1-AcKRST. (A) Genetic organization of the *E. coli* strain LF1-AcKRST. N^ε-acetyl lysyl-tRNA-synthetase (*pylS**) was cloned under the control of the IPTG-inducible *lac* promoter (P_{lac}), while the cognate tRNA$_{CUA}$ (*pylT*) was placed under the control of the constitutive promoter P_{proK}. (B) AcK-dependent amber suppression in *E. coli* LF1-AcKRST. Cells were transformed with the Lux-amber reporter (Lux^{Am}). The resulting strains were then grown overnight in LB, and luminescence development was recorded in response to IPTG and AcK. The maximal luminescence normalized to the OD$_{600}$ from a 16 h time course experiment is depicted. Error bars represent the standard deviation of data from three different experiments.

increased the light intensity to about 3.7×10^5 RLU. Thus, the *E. coli* strain LF1-AcKRST provides an on/off switch for the production of specific proteins. It should be noted that 0.2% Ara was always added to the medium, so that the *luxCDABE* operon was fully induced under all conditions tested. Recall that when *pylS**/*pylT* was provided *in trans* (pACYCDuet-AcKRST) significant luminescence was observed in the absence of AcK (Figure 2C). Thus, the reduction in *pylS**/*pylT* copy

number from about 20−30 in the plasmid-based system[56] to one in the chromosome almost completely abolishes mischarging of tRNA$_{CUA}$. Furthermore, the coupling of chromosomal integration of the amber suppression machinery (AcKRST) with IPTG-inducible expression of *pylS** eliminates the physiological burden that goes along with the constitutive expression of orthogonal tRNAs and aminoacyl tRNA synthetases[57] (Supplementary Figure S4).

To proof functionality of the parental strain *E. coli* LF1-AcKRST in regulating the expression of essential gene products, we aimed to generate a mutant with a defect in cell wall biosynthesis. For this purpose, we chose the gene encoding dihydrodipicolinate synthase DapA, being crucial for the generation of meso diaminopimelic acid (DAP).[58] Accordingly, a *dapA* deletion causes DAP auxotrophy. To this end the *E. coli* strain LF1-AcKRST/*dapA:cam*^{r} was generated using Red/ET (Gene Bridges) recombination. As expected LF1-AcKRST/*dapA:cam*^{r} (DAP$^-$) can only grow in LB supplemented with as little as 75 μM DAP (Supplementary Figure S5).

To trace the observed auxotrophy back to the deletion of *dapA* we constructed a plasmid pBAD/HisA-dapA coding for an N-terminal His$_6$-tagged copy of DapA. Ectopic expression of His$_6$-DapA in LF1-AcKRST/*dapA:cam*^{r} permitted growth in the absence of DAP after induction with 0.2% Ara (Figure 5A). Beside the fact, that this result clearly links the observed phenotype to the gene deletion, it also demonstrates that the N-terminal extension of the protein sequence does not interfere with the enzymatic activity of DapA. Consequently, a plasmid pBAD/HisA-dapA(Am) was derived from pBAD/HisA-dapA. This daughter plasmid deviates from the parental only by an amber motif as a third codon in the original *dapA* ORF which in turn follows a 39 amino acids (aa) long N-terminal extension encoded in the pBAD/HisA backbone.

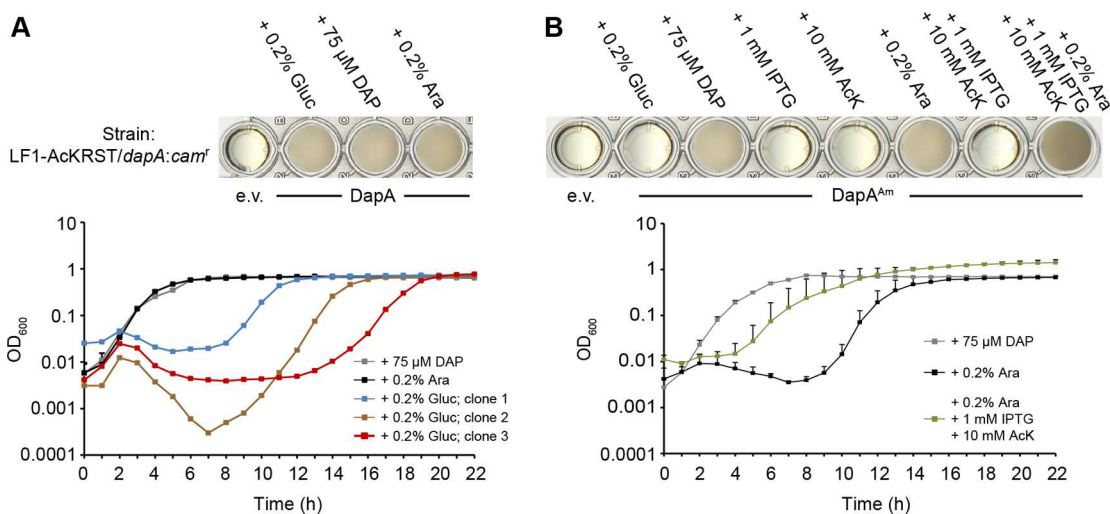We investigated the power of coupled transcriptional and translational regulation by controlling the growth of the DAP$^-$

Figure 5. Transcriptional and translational control of DAP auxotrophy. Growth analysis of the LF1-AcKRST/*dapA:cam*^{r} strain harboring the pBAD/HisA-dapA or pBAD/HisA-dapA(Am) plasmid. Cells were grown in a microtiter plate over a time course of 22h in LB with the indicated supplements. Afterwards pictures of the bacterial cultures were taken. (A) Growth curve and picture of the LF1-AcKRST/*dapA:cam*^{r} strain harboring a plasmid borne copy of *dapA* (pBAD/HisA-dapA), or the empty vector e.v. (pBAD/HisA) as negative control, in response to 0.2% (w/v) L-arabinose (Ara), 0.2% (w/v) D-glucose (Gluc) or 75 μM diaminopimelic acid (DAP). Ara and DAP growth curves are depicted as triplicates, Gluc growth curves are depicted as single measurements. (B) Growth curve and picture of the LF1-AcKRST/*dapA:cam*^{r} strain harboring the empty vector (e.v.) as negative control or pBAD/HisA-dapA(Am) (DapA^{Am}) containing a plasmid borne copy of DapA with an inserted amber codon, in response to Ara, DAP, Gluc, N^ε-acetyl-L-lysine (AcK) and isopropyl β-D-1-thiogalactopyranoside (IPTG) as indicated. Error bars in the growth curve represent the standard deviation of data from three different experiments.
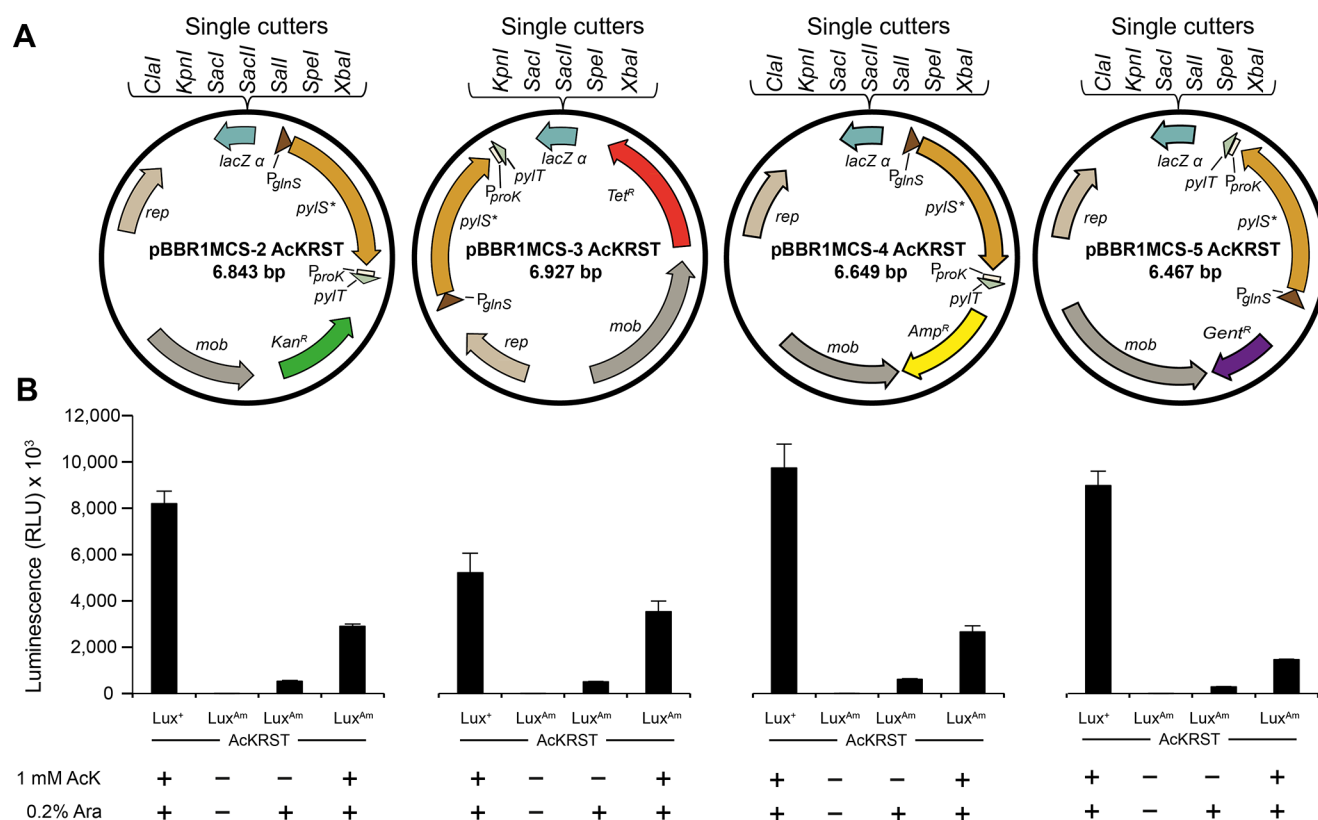
Figure 6. A set of broad-host-range cloning vectors for amber suppression in diverse bacteria. (A) Maps of pBBR1MCS-X AcKRST derivatives. The *pylS**/*pylT* cassette was integrated adjacent to the multiple cloning site in the available broad-host-range plasmids pBBR1MCS-2 to 5. (B) Analysis of amber suppression in cells bearing the pBBR1MCS-X AcKRST derivatives. pBBR1MCS-X AcKRST vectors were transferred into *E. coli* BW25113 together with the Lux-amber reporter (Lux^{Am}) or Lux reporter (Lux^{+}), as positive control. The resulting strains were grown overnight in LB and luminescence development was recorded in response to Ara and AcK. The maximal luminescence normalized to the OD$_{600}$ from a 16 h time course experiment is depicted. Error bars represent the standard deviation of data from three different experiments.

strain in the presence of a plasmid borne gene copy. Therefore, the *E. coli* DAP^{-} strain LF1-AcKRST/*dapA:cam*^{r} was transformed with pBAD/HisA-dapA and pBAD/HisA-dapA(Am), respectively and cells were initially cultivated overnight in LB supplemented with 75 μM DAP. The next day cells were inoculated into LB + 0.2% D-glucose (Gluc) (to catabolically repress P$_{BAD}$ induction), LB + 75 μM DAP and LB + 0.2% Ara (to induce P$_{BAD}$ induction). Growth of the strain harboring the pBAD/HisA-dapA(Am) was additionally monitored in LB + 1 mM IPTG (to induce expression of P$_{lac}$ regulated AckRS), in LB + 1 mM IPTG + 1 mM AcK (to allow Amber suppression) and ultimately in LB containing a combination of all three components (0.2% Ara, 1 mM IPTG and 1 mM AcK) (Figure 5B). As expected, this coupling of AcK dependent amber suppression with Ara induced transcriptional activation permitted growth in the absence of DAP. Notably, growth behavior of cells cultivated with 0.2% L-Ara, 1 mM IPTG and 1 mM AcK was beside a short lag-phase almost identical to what we observed for cells that were supplemented with 75 μM of DAP (Figure 5B). On the other hand, we could not detect an increase in OD$_{600}$ in the cultures lacking either DAP or 0.2% Ara. However, the sole addition of Ara was sufficient to promote overnight growth of LF1-AcKRST/*dapA:cam*^{r} with pBAD/HisA-dapA(Am). Compared to cultures containing all three supplements (Ara, IPTG and AcK) we observed an extended lag phase and even an initial drop in the starting OD$_{600}$ (Figure 5B). The latter might be explained by cell lysis due to instability of the bacterial cell wall, whereas the first 10 h

without any significant change in OD$_{600}$ indicates the development of a suppressor mutation that promotes growth even under translationally repressing conditions. Similar results were observed when monitoring the growth of cells containing pBAD/HisA-dapA in which P$_{BAD}$ dependent gene expression was catabolically repressed[59] (+0.2% Gluc) (Figure 5A).

Thus, the coupling of transcriptional and translational regulation does not only allow a tunable protein output over several orders of magnitude but also prevents a fast development of suppressor mutation in cases where prolonged incubation is necessary.

Four Broad-Host-Range Cloning Vectors Allow Amber Suppression in Diverse Bacterial Species. The demonstration that amber suppression permits tight regulation of protein output in *E. coli* raised the question of the system's transferability to other bacteria. To answer this question, we made use of the various pBBR1MCS-X derivatives available,[60] bearing resistance cassettes for kanamycin, tetracycline, ampicillin and gentamycin, respectively.[61] The pBBR1 origin of replication present in these plasmids is recognized in a wide range of bacterial hosts, and they are also mobilizable for conjugational transfer.[60]

Moreover, the vectors do not belong to the Inc.P, Inc.Q or Inc.W incompatibility group, and can therefore be coinherited with p15A- or ColE1-dependent replicons.[61] Consequently, we cloned the genes of the amber suppression machinery into pBBR1MCS-2 to 5, keeping the original multiple cloning sites intact, and retaining the potential for blue/white screening

(Figure 6A). The function of the resulting derivatives was validated by transformation of *E. coli* with the Lux[Am] reporter and subsequent analysis of AcK-dependent amber suppression (Figure 6B).

Next we cloned a copy of the NanoLuc luciferase (Nluc) into pBBR1MCS-2 AcKRST (Figure 7A) under the control of the P[BAD] promoter. The plasmids generated in this way (pBBR1MCS-2 AcKRST Nluc, pBBR1MCS-2 AcKRST Nluc-(Am), pBBR1MCS-2 AcKRST Nluc PP(Am)) were then transferred into *Salmonella enterica* LT2 and *Vibrio cholerae* El Tor C6706 cells, and the wildtype NanoLuc was used as the positive control (Nluc). AcK-dependent amber suppression was analyzed with a NanoLuc variant, containing a UAG as third codon (Nluc[Am]). Again, a position close to the protein start was selected because the enzyme reportedly tolerates N-terminal fusions.[62] In addition, we constructed a variant with PP amber context (Nluc[PP(Am)]) (Figure 7A). The dependence of luminescence production in *S. enterica* and *V. cholerae* cells on supplementation with Ara and/or AcK was then analyzed. As expected, cells encoding a NanoLuc without any amber insertion emit luminescence when grown in the presence of Ara (Figure 7B, C). In contrast, significant luminescence production by cells harboring the Nluc[Am] or Nluc[PP(Am)] reporter required the simultaneous addition of Ara and AcK. As in *E. coli*, the PP(Am) context further diminished the low level of light production seen in the absence of AcK. Thus, our results are 1:1 transferable from *E. coli* to *S. enterica* and *V. cholerae* and with this, amber suppression is applicable to diverse bacteria.

■ CONCLUDING REMARKS

In the present study we investigated the application of AcK-dependent amber suppression for the regulation of translational outputs. We showed that coupling of transcriptional and translational regulation is an effective way to fully switch off the synthesis of specific proteins. Manipulation of the concentrations of both the transcriptional inducer and AcK permits gradual adjustment of the protein output. Furthermore, and in line with previous reports, we observed that the incorporation of an amber stop codon into a plasmid-borne gene copy resulted in a measurable level of protein synthesis by translational read-through (Figure 2C), which can be modulated by changing the amber context[43,46,49] (Table 1). Such read-through might be beneficial when reduced levels of an essential gene product are needed to maintain viability, but nevertheless cause a mutant phenotype. If necessary, the translational output can be further increased by ectopic expression of the *pylS*/*pylT* amber suppression system. It should be noted that in our system mischarging becomes virtually irrelevant when *pylS*/*pylT* was integrated in the chromosome and combined with inducible expression. This is exemplified by the fact that *E. coli* LF1-AcKRST cells encoding a Lux[Am] reporter remain completely dark in the absence of IPTG (Figure 4). Only the concomitant presence of IPTG and AcK allowed efficient amber suppression. Thus, chromosomal incorporation of the amber suppression machinery might generally increase specificity for ncAAs.

Altogether, our analyses of AcK-dependent amber suppression as a means of regulating protein production provide the basis for use of the system as a versatile translational control mechanism for investigation of the function of essential genes. Thus, the components available in this molecular toolbox enable fine-tuning of protein synthesis, not only in *E. coli* but
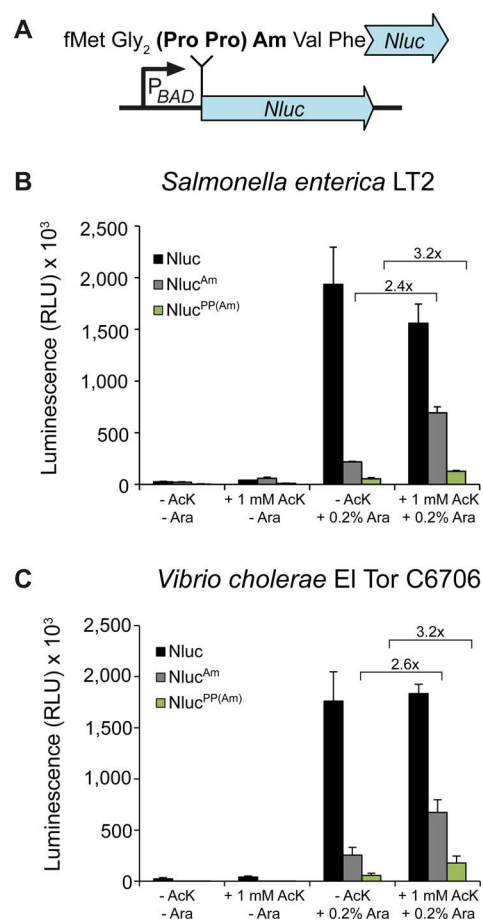
Figure 7. Amber suppression in *Salmonella enterica* and *Vibrio cholerae*. (A) An amber (Nluc[Am]) or PP amber motif (Nluc[PP(Am)]) was inserted after codon 2 of the NanoLuc gene and cloned into the pBBR1MCS-2 AcKRST backbone. A vector bearing the wildtype NanoLuc gene served as the positive control (Nluc). pBBR1MCS-2 AcKRST Nluc, pBBR1MCS-2 AcKRST Nluc(Am) and pBBR1MCS-2 AcKRST Nluc PP(Am) were transferred into *Vibrio cholerae* (B) and *Salmonella enterica* (C). Luminescence was measured from overnight cultures using the Nano-Glo luciferase assay system. Error bars represent the standard deviation of data from three different experiments.

also in other Gram-negative bacteria such as *S. enterica* or *V. cholerae*.

■ MATERIALS AND METHODS

Bacterial Stains, Oligonucleotides and Plasmids. Strains, primers and plasmids used in this study are listed in Supplementary Tables S1−S3. All oligonucleotides used for PCR amplification and sequencing were synthesized by Sigma-Aldrich. The sequences of all constructed plasmids were verified by Sanger sequencing.

Molecular Biological Methods. *E. coli*, *S. enterica* and *V. cholerae* cells were routinely grown in LB and where indicated in M9 minimal medium.[63] Solidification of the medium was achieved by adding 1.5% (w/v) Agar−Agar (Carl Roth). If necessary, media were supplemented with antibiotics at the following concentrations: ampicillin sodium salt (100 μg/mL), chloramphenicol (30 μg/mL), kanamycin sulfate (50 μg/mL), tetracycline hydrochloride (15 μg/mL), gentamycin sulfate (15 μg/mL), and streptomycin sulfate (50 μg/mL). For blue-white selection, LB-agar plates were additionally supplemented with 80 μM 5-bromo-4-chloro-3-indolyl β-D-galacto-

pyranoside (X-Gal) (Sigma-Aldrich) and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (Sigma-Aldrich) to induce the *lac* promoter P_{lac}. Unless indicated otherwise, 0.2% (w/v) L-arabinose was added to induce the P_{BAD} promoter.

All enzymes and kits were used according to the manufacturers' directions. DNA fragments were purified from agarose gels using a high-yield PCR cleanup and gel extraction kit (Sued-Laborbedarf). Restriction endonucleases were purchased from New England Biolabs. Sequence amplifications were performed by PCR using Phusion high-fidelity DNA polymerase (Finnzymes) or Taq DNA polymerase (New England Biolabs). All mutants were constructed by one- or two-step PCR using mismatched primer pairs.[64]

β-Galactosidase activity assays were performed as essentially described previously.[55]

Plasmid and Strain Construction. pACYCDuet-AcKRST was constructed as follows: A PCR fragment containing the respective mutations in the *pylS* gene was generated as described previously by Umehara *et al.* 2012[7] and cloned into pACYCDuet-PylRST using the internal *Xho*I and *Not*I restriction sites (Supplementary Table S2, Supplementary Table S3).

LuxCDABE-containing pBAD/HisA constructs were generated as follows: The *luxCDABE* operon from *Photorhabdus luminescens* was amplified from pBBR1MCS-5 TT-RBS-lux,[65] and then cloned into pBAD/HisA (Invitrogen) using *Nco*I and *Kpn*I as restriction sites. The resulting vector pBAD/HisA-Lux served as a template for all other *luxCDABE*-bearing plasmids used in this study. To construct *luxC* amber versions we used mismatched primer pairs (Supplementary Table S3) for overlap-extension PCR and cloned the resulting fragments into pBAD/HisA-Lux using *Nco*I and *Bsr*GI, thereby replacing the parental *luxC* fragment. The resulting plasmids, pBAD/HisA-Lux(Am), pBAD/HisA-Lux(2Am), pBAD/HisA-Lux-RA-(Am), pBAD/HisA-Lux-DA(Am), pBAD/HisA-Lux-HH(Am), pBAD/HisA-Lux-KDP(Am), pBAD/HisA-Lux-PP(Am) and pBAD/HisA-Lux-KDPP(Am) are listed and described in Supplementary Table S2.

The constructs pBAD/HisA-dapA and pBAD/HisA-dapA-(Am) were generated as follows: From genomic *E. coli* DNA, *dapA* was amplified and the amber motif was introduced using a mismatched primer (Supplementary Table S3). The resulting fragments were than cloned into the pBAD/HisA backbone using *Xho*I and *Eco*RI as restriction sites. The resulting plasmids are listed in Supplementary Table S2.

To generate pBAD/HisA-Kan and plasmid derivatives (Supplementary Table S2) the ampicillin resistance cassette was replaced by *nptI*, which was amplified from pBBR1MCS-2 (Supplementary Table S3). The PCR fragment was introduced restriction-free into pBAD/HisA as described.[66]

The broad-host-range cloning vectors containing *pylS*/pylT* were constructed as follows: *pylS*/pylT* was amplified from pACYCDuet-AcKRST and the resulting PCR fragment was cloned into pBBR1MCS-2 to 5 plasmids using either the *Nsi*I or the *Age*I restriction site (Supplementary Table S2, Supplementary Table S3).

The Nluc reporter plasmids were generated by amplifying the sequence encoding NanoLuc luciferase from pNL1.1 (Promega). The respective amber motifs were introduced using mismatched primer pairs in a standard PCR (Supplementary Table S3). The fragments obtained were each combined with the arabinose-inducible P_{BAD} promoter fragment (amplified from pBAD/HisA, *via* overlap-extension

PCR), and cloned into pBBR1MCS-2 AcKRST using *Xba*I and *Kpn*I restriction sites. The resulting plasmids, pBBR1MCS-2 AcKRST Nluc, pBBR1MCS-2 AcKRST Nluc(Am), and pBBR1MCS-2 AcKRST Nluc PP(Am) are listed in Supplementary Table S2.

E. coli LF1-LacZ-K, LF1-LacZ(Am), LF1-LacZ-PPK, and LF1-LacZ-PP(Am) strains, as well as LF1-AcKRST, were constructed essentially as described previously by Fried *et al.*[55] and are listed in Supplementary Table S1.

E. coli LF1-AcKRST/*dapA:cam*^{r} was generated by using the pRed/ET recombination technology in accordance to the technical protocol of the Quick & Easy *E. coli* Gene Deletion Kit (Gene Bridges) and is listed in Supplementary Table S1.

Lux-Based Luminescence Activity Assay. Single colonies were transferred into LB supplemented with the appropriate antibiotics, and grown aerobically overnight at 37 °C. On the next day, cells were inoculated at an OD_{600} of 0.01 into fresh LB supplemented with the corresponding antibiotics. Cells were then grown aerobically at 37 °C in a microtiter plate in a Tecan Infinite F500 system (TECAN) and luminescence development in response to AcK and Ara was monitored at 10 min intervals over the course of 16 h. Light units were normalized to OD_{600} and are thus expressed in relative light units (RLU). Subsequently RLUs were background corrected by subtracting the luminescence level detected in cells containing a vector without a *luxCDABE* operon. Each measurement was performed in triplicate.

The luminescence assay for the *E. coli* LF1-AcKRST strain was performed as described above but with the difference that the cells on the second day were inoculated in LB, containing besides the antibiotic additionally 0.2% Ara (w/v). Subsequently, luminescence development in response to 1 mM AcK and 1 mM IPTG was detected as described before.

NanoLuc Luciferase Assay in *S. enterica* and *V. cholerae*. Single colonies of *S. enterica* and *V. cholerae* were transferred into LB supplemented with kanamycin sulfate (50 μg/mL) and grown overnight aerobically at 37 °C. On the next day, these cultures were used to inoculate the main cultures to an OD_{600} of 0.01. To determine AcK- and Ara-dependent luminescence development, all main cultures were grown aerobically overnight at 37 °C in LB supplemented with kanamycin sulfate (50 μg/mL) and containing either 1 mM AcK, 0.2% Ara (w/v), or a combination of both compounds. A culture grown in the absence of both served as negative control. OD_{600} was determined from overnight cultures and the Nano-Glo luciferase assay reagent (Promega) was prepared according to manufacturer's instructions. Then, the experimental cultures were mixed 1:1 with Nano-Glo luciferase assay reagent, transferred into a 96-well microtiter plate, and luminescence was determined 5 min later with a Tecan Infinite F500 system (TECAN). Light units were normalized to OD_{600} and expressed in relative light units (RLU). Each measurement was performed in triplicate.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.7b00048.

Table S1−S4; Figures S1−S5 (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: jung@lmu.de.
*E-mail: juergen.lassak@lmu.de.

ORCID

Kirsten Jung: 0000-0003-0779-6841
Jürgen Lassak: 0000-0003-3936-3389

Author Contributions

The study was designed by W.V., K.J. and J.L. and directed as well as coordinated by K.J. and J.L.; W.V. and C.M. constructed all strains and performed the enzyme assays; R.K. performed the parameter optimization for the Lux reporters. All authors contributed to the writing of the manuscript

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

ABBREVIATIONS

AcK, N^ε-acetyl-L-lysine; AcKRS, N^ε-acetyl lysyl-tRNA-synthetase; ncAA, noncanonical amino acid; Sec, selenocysteine; Pyl, pyrrolysine; PylRS, pyrrolysyl-tRNA-synthetase; PylRST, pyrrolysyl-tRNA-synthase/tRNA$_{CUA}$ pair; AcKRST, N^ε-acetyl lysyl-tRNA-synthetase/tRNA$_{CUA}$ pair; Nluc, NanoLuc luciferase

REFERENCES

(1) Guo, X. V., Monteleone, M., Klotzsche, M., Kamionka, A., Hillen, W., Braunstein, M., Ehrt, S., and Schnappinger, D. (2007) Silencing *Mycobacterium smegmatis* by using tetracycline repressors. *J. Bacteriol.* 189, 4614−4623.

(2) Higuchi, M., Tsutsumi, R., Higashi, H., and Hatakeyama, M. (2004) Conditional gene silencing utilizing the *lac* repressor reveals a role of SHP-2 in *cagA*-positive *Helicobacter pylori* pathogenicity. *Cancer Sci.* 95, 442−447.

(3) Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., and Lim, W. A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152, 1173−1183.

(4) Filipowicz, W., Jaskiewicz, L., Kolb, F. A., and Pillai, R. S. (2005) Post-transcriptional gene silencing by siRNAs and miRNAs. *Curr. Opin. Struct. Biol.* 15, 331−341.

(5) Kavran, J. M., Gundllapalli, S., O'Donoghue, P., Englert, M., Soll, D., and Steitz, T. A. (2007) Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci. U. S. A.* 104, 11268−11273.

(6) Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., and Yokoyama, S. (2008) Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase. *J. Mol. Biol.* 378, 634−652.

(7) Umehara, T., Kim, J., Lee, S., Guo, L. T., Soll, D., and Park, H. S. (2012) *N*-acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess *N*-acetyl lysine specificity *in vitro* and *in vivo*. *FEBS Lett.* 586, 729−733.

(8) Yoshizawa, S., and Bock, A. (2009) The many levels of control on bacterial selenoprotein synthesis. *Biochim. Biophys. Acta, Gen. Subj.* 1790, 1404−1414.

(9) Krzycki, J. A. (2013) The path of lysine to pyrrolysine. *Curr. Opin. Chem. Biol.* 17, 619−625.

(10) Bulteau, A. L., and Chavatte, L. (2015) Update on selenoprotein biosynthesis. *Antioxid. Redox Signaling* 23, 775−794.

(11) Stock, T., and Rother, M. (2009) Selenoproteins in Archaea and Gram-positive bacteria. *Biochim. Biophys. Acta, Gen. Subj.* 1790, 1520−1532.

(12) Gaston, M. A., Jiang, R., and Krzycki, J. A. (2011) Functional context, biosynthesis, and genetic encoding of pyrrolysine. *Curr. Opin. Microbiol.* 14, 342−349.

(13) James, C. M., Ferguson, T. K., Leykam, J. F., and Krzycki, J. A. (2001) The amber codon in the gene encoding the monomethylamine methyltransferase isolated from *Methanosarcina barkeri* is translated as a sense codon. *J. Biol. Chem.* 276, 34252−34258.

(14) Wang, L., Xie, J., and Schultz, P. G. (2006) Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* 35, 225−249.

(15) Wan, W., Tharp, J. M., and Liu, W. R. (2014) Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta, Proteins Proteomics* 1844, 1059−1070.

(16) Dumas, A., Lercher, L., Spicer, C. D., and Davis, B. G. (2015) Designing logical codon reassignment - Expanding the chemistry in biology. *Chem. Sci.* 6, 50−69.

(17) Chatterjee, C., and Muir, T. W. (2010) Chemical approaches for studying histone modifications. *J. Biol. Chem.* 285, 11045−11050.

(18) Lang, K., Davis, L., and Chin, J. W. (2015) Genetic encoding of unnatural amino acids for labeling proteins. *Methods Mol. Biol.* 1266, 217−228.

(19) Rovner, A. J., Haimovich, A. D., Katz, S. R., Li, Z., Grome, M. W., Gassaway, B. M., Amiram, M., Patel, J. R., Gallagher, R. R., Rinehart, J., and Isaacs, F. J. (2015) Recoded organisms engineered to depend on synthetic amino acids. *Nature* 518, 89−93.

(20) Minaba, M., and Kato, Y. (2014) High-yield, zero-leakage expression system with a translational switch using site-specific unnatural amino acid incorporation. *Appl. Environ. Microbiol.* 80, 1718−1725.

(21) Kato, Y. (2015) Tunable translational control using site-specific unnatural amino acid incorporation in *Escherichia coli*. *PeerJ* 3, e904.

(22) Neumann, H., Peak-Chew, S. Y., and Chin, J. W. (2008) Genetically encoding N^ε-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* 4, 232−234.

(23) Kim, D., Yu, B. J., Kim, J. A., Lee, Y. J., Choi, S. G., Kang, S., and Pan, J. G. (2013) The acetylproteome of Gram-positive model bacterium *Bacillus subtilis*. *Proteomics* 13, 1726−1736.

(24) Weinert, B. T., Iesmantavicius, V., Wagner, S. A., Scholz, C., Gummesson, B., Beli, P., Nystrom, T., and Choudhary, C. (2013) Acetyl-phosphate is a critical determinant of lysine acetylation in *E. coli*. *Mol. Cell* 51, 265−272.

(25) Zhang, K., Zheng, S., Yang, J. S., Chen, Y., and Cheng, Z. (2013) Comprehensive profiling of protein lysine acetylation in *Escherichia coli*. *J. Proteome Res.* 12, 844−851.

(26) Kuhn, M. L., Zemaitaitis, B., Hu, L. I., Sahu, A., Sorensen, D., Minasov, G., Lima, B. P., Scholle, M., Mrksich, M., Anderson, W. F., Gibson, B. W., Schilling, B., and Wolfe, A. J. (2014) Structural, kinetic and proteomic characterization of acetyl phosphate-dependent bacterial protein acetylation. *PLoS One* 9, e94816.

(27) Lima, B. P., Antelmann, H., Gronau, K., Chi, B. K., Becher, D., Brinsmade, S. R., and Wolfe, A. J. (2011) Involvement of protein acetylation in glucose-induced transcription of a stress-responsive promoter. *Mol. Microbiol.* 81, 1190−1204.

(28) AbouElfetouh, A., Kuhn, M. L., Hu, L. I., Scholle, M. D., Sorensen, D. J., Sahu, A. K., Becher, D., Antelmann, H., Mrksich, M., Anderson, W. F., Gibson, B. W., Schilling, B., and Wolfe, A. J. (2015) The *E. coli* sirtuin CobB shows no preference for enzymatic and nonenzymatic lysine acetylation substrate sites. *MicrobiologyOpen* 4, 66−83.

(29) Gattner, M. J., Vrabel, M., and Carell, T. (2013) Synthesis of ε-N-propionyl-, ε-N-butyryl-, and ε-N-crotonyl-lysine containing histone H3 using the pyrrolysine system. *Chem. Commun. 49*, 379−381.

(30) Kuhn, S. M., Rubini, M., Fuhrmann, M., Theobald, I., and Skerra, A. (2010) Engineering of an Orthogonal Aminoacyl-tRNA Synthetase for Efficient Incorporation of the Non-natural Amino Acid O-Methyl-L-tyrosine using Fluorescence-based Bacterial Cell Sorting. *J. Mol. Biol. 404*, 70−87.

(31) Belas, R., Mileham, A., Cohn, D., Hilman, M., Simon, M., and Silverman, M. (1982) Bacterial bioluminescence: isolation and expression of the luciferase genes from *Vibrio harveyi*. *Science 218*, 791−793.

(32) Close, D., Xu, T., Smartt, A., Rogers, A., Crossley, R., Price, S., Ripp, S., and Sayler, G. (2012) The evolution of the bacterial luciferase gene cassette (*lux*) as a real-time bioreporter. *Sensors 12*, 732−752.

(33) Schultz, D. W., and Yarus, M. (1990) A simple and sensitive *in vivo* luciferase assay for tRNA-mediated nonsense suppression. *J. Bacteriol. 172*, 595−602.

(34) Kudla, G., Murray, A. W., Tollervey, D., and Plotkin, J. B. (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science 324*, 255−258.

(35) Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., and Bluthgen, N. (2013) Efficient translation initiation dictates codon usage at gene start. *Mol. Syst. Biol. 9*, 675.

(36) Goodman, D. B., Church, G. M., and Kosuri, S. (2013) Causes and effects of N-terminal codon bias in bacterial genes. *Science 342*, 475−479.

(37) O'Donoghue, P., Prat, L., Heinemann, I. U., Ling, J., Odoi, K., Liu, W. R., and Soll, D. (2012) Near-cognate suppression of amber, opal and quadruplet codons competes with aminoacyl-tRNA^{Pyl} for genetic code expansion. *FEBS Lett. 586*, 3931−3937.

(38) Gundllapalli, S., Ambrogelly, A., Umehara, T., Li, D., Polycarpo, C., and Soll, D. (2008) Misacylation of pyrrolysine tRNA *in vitro* and *in vivo*. *FEBS Lett. 582*, 3353−3358.

(39) Ambrogelly, A., Palioura, S., and Soll, D. (2007) Natural expansion of the genetic code. *Nat. Chem. Biol. 3*, 29−35.

(40) Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., and Yokoyama, S. (2008) Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N^ε-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem. Biol. 15*, 1187−1197.

(41) Johnson, D. B., Xu, J., Shen, Z., Takimoto, J. K., Schultz, M. D., Schmitz, R. J., Xiang, Z., Ecker, J. R., Briggs, S. P., and Wang, L. (2011) RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol. 7*, 779−786.

(42) Johnson, D. B. F., Wang, C., Xu, J. F., Schultz, M. D., Schmitz, R. J., Ecker, J. R., and Wang, L. (2012) Release Factor One Is Nonessential in *Escherichia coli*. *ACS Chem. Biol. 7*, 1337−1344.

(43) Mottagui-Tabar, S., Bjornsson, A., and Isaksson, L. A. (1994) The second to last amino acid in the nascent peptide as a codon context determinant. *EMBO J. 13*, 249−257.

(44) Pott, M., Schmidt, M. J., and Summerer, D. (2014) Evolved sequence contexts for highly efficient amber suppression with noncanonical amino acids. *ACS Chem. Biol. 9*, 2815−2822.

(45) Phillips-Jones, M. K., Watson, F. J., and Martin, R. (1993) The 3′ codon context effect on UAG suppressor tRNA is different in *Escherichia coli* and human cells. *J. Mol. Biol. 233*, 1−6.

(46) Xu, H., Wang, Y., Lu, J., Zhang, B., Zhang, Z., Si, L., Wu, L., Yao, T., Zhang, C., Xiao, S., Zhang, L., Xia, Q., and Zhou, D. (2016) Re-exploration of the Codon Context Effect on Amber Codon-Guided Incorporation of Noncanonical Amino Acids in *Escherichia coli* by the Blue-White Screening Assay. *ChemBioChem 17*, 1250−1256.

(47) Bjornsson, A., Mottagui-Tabar, S., and Isaksson, L. A. (1996) Structure of the C-terminal end of the nascent peptide influences translation termination. *EMBO J. 15*, 1696−1704.

(48) Hayes, C. S., Bose, B., and Sauer, R. T. (2002) Proline residues at the C terminus of nascent chains induce SsrA tagging during translation termination. *J. Biol. Chem. 277*, 33825−33832.

(49) Tanner, D. R., Cariello, D. A., Woolstenhulme, C. J., Broadbent, M. A., and Buskirk, A. R. (2009) Genetic identification of nascent peptides that induce ribosome stalling. *J. Biol. Chem. 284*, 34809−34818.

(50) Ude, S., Lassak, J., Starosta, A. L., Kraxenberger, T., Wilson, D. N., and Jung, K. (2013) Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science 339*, 82−85.

(51) Karlin, S., Mrazek, J., and Campbell, A. M. (1998) Codon usages in different gene classes of the *Escherichia coli* genome. *Mol. Microbiol. 29*, 1341−1355.

(52) Miller, J. H. (1992) *A Short Course in Bacterial Genetics: A Laboratory Manual for Escherichia coli and Related Bacteria*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 456 pp.

(53) Epstein, W., and Kim, B. S. (1971) Potassium transport loci in *Escherichia coli* K-12. *J. Bacteriol. 108*, 639−644.

(54) Juers, D. H., Heightman, T. D., Vasella, A., McCarter, J. D., Mackenzie, L., Withers, S. G., and Matthews, B. W. (2001) A structural view of the action of *Escherichia coli* (*lacZ*) β-Galactosidase. *Biochemistry 40*, 14781−14794.

(55) Fried, L., Lassak, J., and Jung, K. (2012) A comprehensive toolbox for the rapid construction of *lacZ* fusion reporters. *J. Microbiol. Methods 91*, 537−543.

(56) Lutz, R., and Bujard, H. (1997) Independent and tight regulation of transcriptional units in *Escherichia coli via* the LacR/O, the TetR/O and AraC/I$_1$-I$_2$ regulatory elements. *Nucleic Acids Res. 25*, 1203−1210.

(57) Hong, S. H., Kwon, Y. C., and Jewett, M. C. (2014) Non-standard amino acid incorporation into proteins using *Escherichia coli* cell-free protein synthesis. *Front. Chem.*, DOI: 10.3389/fchem.2014.00034.

(58) Laber, B., Gomis-Ruth, F. X., Romao, M. J., and Huber, R. (1992) *Escherichia coli* dihydrodipicolinate synthase. Identification of the active site and crystallization. *Biochem. J. 288* (Pt 2), 691−695.

(59) Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P$_{BAD}$ promoter. *J. Bacteriol. 177*, 4121−4130.

(60) Antoine, R., and Locht, C. (1992) Isolation and molecular characterization of a novel broad-host-range plasmid from *Bordetella bronchiseptica* with sequence similarities to plasmids from gram-positive organisms. *Mol. Microbiol. 6*, 1785−1799.

(61) Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M., 2nd, and Peterson, K. M. (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene 166*, 175−176.

(62) Boute, N., Lowe, P., Berger, S., Malissard, M., Robert, A., and Tesar, M. (2016) NanoLuc Luciferase - A Multifunctional Tool for High Throughput Antibody Screening. *Front. Pharmacol. 7*, 27.

(63) Miller, J. H. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbor, NY.

(64) Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene 77*, 51−59.

(65) Gödeke, J., Heun, M., Bubendorfer, S., Paul, K., and Thormann, K. M. (2011) Roles of two *Shewanella oneidensis* MR-1 extracellular endonucleases. *Appl. Environ. Microbiol. 77*, 5342−5351.

(66) van den Ent, F., and Lowe, J. (2006) RF cloning: a restriction-free method for inserting target genes into plasmids. *J. Biochem. Biophys. Methods 67*, 67−74.

6   Concluding discussion

Protein-glycosylation is a universal strategy for post-translational modification and has been linked to numerous processes that are important for bacterial physiology and pathogenicity (71, 100). These capacities are mediated by the considerable structural and functional changes that can result from carbohydrate linkage to polypeptides (16). Almost all reported cases of nitrogen-linked glycosylation (N-glycosylation) occur on asparagine (65). The EF-P specific rhamnosyltransferase EarP is one of only two known glycosyltransferases linked to arginine N-glycosylation. Both of these enzymes are involved in mediating bacterial pathogenicity (113, 114, 116). Moreover, arginine rhamnosylation has so far not been detected in humans (126) Therefore, this process offers opportunities for application as an orthologous protein glycosylation platform in synthetic biology and as an efficient antibacterial target in medicine. However, understanding of the molecular mechanism of the glycoconjugate biosynthesis must precede application.

The structural and biochemical data presented here substantially contributes to a better understanding of arginine glycosylation and allows us to derive a mechanistic model of EarP mediated EF-P modification (Figure 6). In a first step, the nucleotide sugar donor TDP-β-L-rhamnose is bound and oriented in the C-terminal domain of the GT-B glycosyltransferase EarP by the concerted action of several highly conserved amino acids (Chapter 3, Figure 6B). Subsequently, structural elements of the KOW-like EF-P N-domain are recognized and bound by the N-terminal EarP domain, thereby occluding the active site from the environment (Chapters 3, 4 and (127)). By a mechanism that is so far not fully understood, the negatively charged side chains of two aspartates activate the acceptor arginine, which in turn engages in a nucleophilic attack onto the anomeric carbon of the nucleotide sugar (Chapter 3, Figure 6C). This reaction results in the formation of α-rhamnosyl-arginine on the protein acceptor, thus representing an inverting mode of glycosyl transfer (Chapter 2). Finally, the products are released from the active site and EarP engages in another cycle of rhamnosylation (Figure 6A).
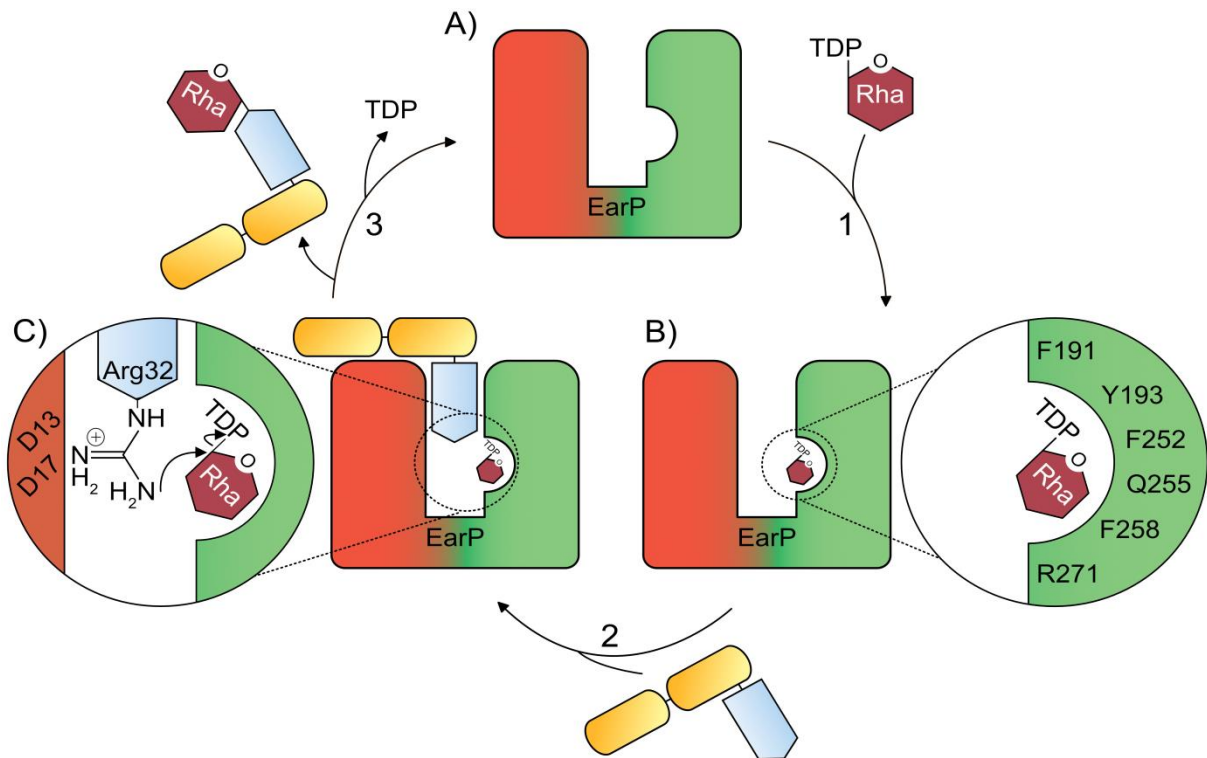
Figure 6. Proposed glycosylation mechanism of the rhamnosyltransferase EarP from *P. putida*. A) Ground state; both the donor and acceptor binding site of EarP (orange and green) are unoccupied. **B)** Donor-bound state; TDP-β-L-rhamnose (TDP-Rha, red) is bound and oriented within the binding pocket in the C-terminal domain. **C)** Catalytic state; The putative catalytic amino acids D13 and D17 facilitate the nucleophilic attack onto the anomeric centre of TDP-β-L-rhamnose by activating the arginine 32 guanidinium of EF-P (yellow and blue). Binding and dissociation events are indicated by arrows: 1. TDP-Rha binding; 2. EF-P binding; 3. EF-P, rhamnose and TDP dissociation.

6.1 EarP is a GT-B fold glycosyltransferase

Crystal structure analysis identified EarP as a GT-B fold glycosyltransferase (Chapter 3). Enzymes of this fold consist of two physically separate domains that are connected by a flexible linker (42). Canonically, the N-terminal domain is involved in acceptor binding, whereas the nucleotide sugar binding pocket is localized in the C-terminal domain (26). Accordingly, EarP binds its donor substrate TDP-β-L-rhamnose in a highly conserved pocket in the protein C-domain (Chapter 3, Figure 6B). Here, the aromatic side chains of F191, F252 and F258 surround and coordinate the thymine moiety of the substrate. The deoxyribose is specifically recognized by hydrogen bonds with Q255. A hydrogen bonding network established by the side chains of R271 and Y193, as well as the backbone amides of E273 and D274 orient the phosphate. The side chain of R271 is localized in a position suitable to neutralize the developing negative charges on the phosphate upon transfer and thus to stabilize the deglycosylated dinucleotide product (31, 42) (Chapter 3, Figure 6B). This donor binding site architecture is broadly observed for the GT-B fold, which generally exhibits a high degree of structural conservation in the nucleotide sugar binding domain (26).

In contrast, the N-terminal acceptor binding domain is structurally more diverse, most likely reflecting the capability of GT-B fold glycosyltransferases to adapt to a wide variety of different acceptor substrates such as carbohydrates, lipids or proteins (26). Structural analysis of the EF-P-EarP complex from *Neisseria meningitidis* (PDB: 5WXK) confirmed that the N-terminal domain of EarP is responsible for binding of the acceptor substrate (127). Thus, the general principles that are observed for GT-B protein-glycosyltransferases with respect structure and substrate binding also apply for EarP (27). Conversely, the molecular targets of the rhamnosyltransferase are quite unusual. EarP is so far the only known glycosyltransferase that activates a cytosolic protein and the sugar that is transferred in order to confer this activity is rarely used in protein glycosylation (65, 116). In addition, EarP also represents the first example of GT-B mediated arginine glycosylation. As the chemical nature of the arginine guanidinium is different from the asparagine amido group, the mechanism for activation of this side chain remains to be conclusively determined.

6.2 EarP is an inverting glycosyltransferase

NMR analysis of *in vivo* modified EF-P of *P. putida* demonstrated that the product of the rhamnosylation reaction is α-rhamnosyl-arginine (Chapter 2). As the nucleotide sugar donor substrate for the glycosylation reaction was previously identified to be TDP-β-L-rhamnose (116), these results unambiguously identified EarP as an inverting glycosyltransferase. Wang and colleagues later confirmed our findings using an alternative approach. They used chemical synthesis to generate peptides that correspond to the *P. aeruginosa* EF-P acceptor loop and carried either α- or β-mono-rhamnosyl-arginine (128). The elution profiles of these glycopeptides during nano-UHPLC analyses were compared to the one of the corresponding native peptide released from purified EF-P after Lys-C digest. As the native peptide co-eluted with the synthesized α-rhamnosyl-arginine containing glycopeptide, the inversion of the anomeric configuration during transfer was confirmed. The specific stereochemical outcome of the glycosylation reaction is a key characteristic of a given glycosyltransferase (27). Inversion and retention result from fundamentally different reaction mechanisms which are in turn based on the microenvironment within the active site (129). Structural comparison of glycosyltransferases that result in the same stereochemical outcome can therefore be utilized in order to deduce the underlying molecular mechanism (Chapter3).

A well-studied and structurally supported mechanism for bacterial N-glycosylation is that of the inverting asparagine olygosaccharyltransferase PglB of *C. lari* (52, 54). This enzyme catalyzes the transfer of a preassembled glycan from a lipid donor to a multitude of periplasmic acceptor proteins (130). The catalytic activity of GTs mainly includes activation of the acceptor substrate, which in turn attacks the anomeric carbon (29, 42). Therefore, this mechanism of activation provides valuable information despite the utilization of a chemically

distinct donor substrate (131). In PglB, two negatively charged amino acids are responsible for activation of the acceptor asparagine. These side chains form hydrogen bonds with the amido group. Thereby the partial double-bond character is abolished and the acceptor is activating by a twisted amide mechanism. The necessary interactions are further stabilized by a coordinated divalent cation and additional side chain hydrogen bond interactions (52, 54) (Figure 7A).

Bioinformatic and structural analyses of EarP led to the identification of two invariant aspartate residues (D13 and D17 in *P. putida* EarP) that are located within the active site of the rhamnosyltransferase (Chapter 3). The side chains of these amino acids might exert a similar function as the catalytically important amino acids in PglB (Figure 7A and 7B). Substitution of either of these amino acids to alanine leads to inactivation of EarP. NMR-analysis confirmed that the corresponding protein variants are correctly folded and capable of binding the sugar substrate. Additional bacterial two-hybrid based protein-protein interaction studies showed that interaction with the acceptor substrate is not inhibited by these mutations either (Chapter 3). Based on these results, we suggest a PglB-like mechanism for the activation of arginine (Figure 7). In this model, D13 and D17 form hydrogen bond interactions with the acceptor. These interactions in turn abolish electron delocalization on the guanidinium, yielding a lone electron pair on the nitrogen that is capable of a nucleophilic attack onto the anomeric carbon (Figure 7B).

As GT-B glycosyltransferases are in general functionally independent of divalent cations, involvement of charged metal ions seems unlikely. Instead, an extended hydrogen bond network based on side chain interactions could be compensating for the lack of the positive charge. Due to low electron density of structural elements of the EarP N-terminal domain, the position of the catalytically relevant amino acid side chains of D13 and D17, but not of the surrounding and possibly supporting residues, could be derived from the crystal structure (Chapter 3). Thus it remained unclear, whether the EarP active site is suitably organized to coordinate a twisted amide-like mechanism of activation.
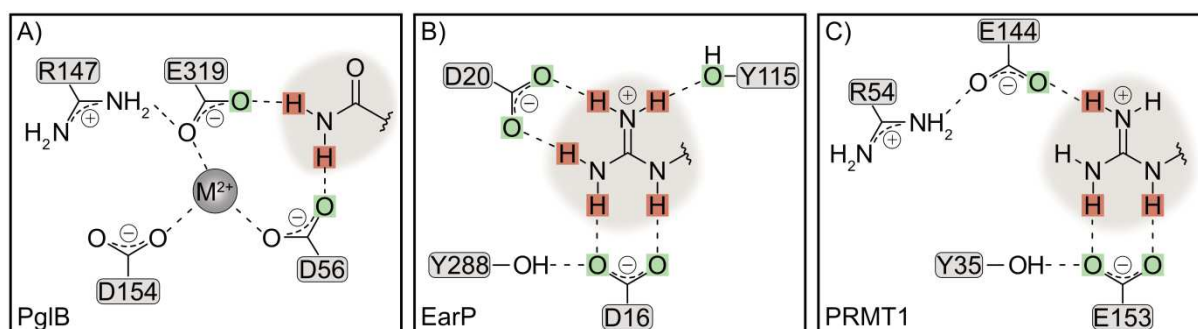
Figure 7. Chemical structures of the PglB, EarP and PRMT1 catalytic sites. Amino acids involved in building the active site are indicated by single letter code specifiers. Hydrogen bonds are indicated by dashed lines. Hydrogen atoms of the acceptor substrates that are engaging in hydrogen bond interaction with the corresponding modification enzyme are highlighted by a red background. Oxygen atoms of the modification enzyme that are engaging in hydrogen bond interaction with the acceptor substrate are highlighted by a green background. **A)** PglB of *C. jejuni* as described in (52). The position of the divalent metal ion (M^{2+}) is indicated by a grey sphere. The amido group of the acceptor asparagine is highlighted by a grey backdrop. **B)** EarP of *N. meningitidis* as described in (127). The guanidinium of the acceptor arginine is highlighted by a grey backdrop. **C)** PRMT1 of *Rattus norvegicus* (right) as described in (132). The guanidinium of the acceptor arginine is highlighted by a grey backdrop.

The two aspartates D16 and D20 of *N. meningitidis* EarP (corresponding to D13 and D17 of *P. putida* EarP) are located in close proximity of the acceptor arginine and form salt bridge interactions with the amino groups of the guanidinium (127). Together with additional side chain interactions, these negatively charged amino acids comprise the main site of contact. Based on this structure, D20 was suggested to act as catalytic base that deprotonates one of the amino groups in order to activate the acceptor arginine. This mechanism is orthologous to the one that is generally accepted to be utilized by most inverting O-glycosyltransferases and has also been proposed for N-linked glycosylation (29, 62, 133, 134) (Figure 4D).

In the case of asparagine modification by PglB, the specific architecture of the active site, the involvement of a divalent cation, as well as the orientation of the acceptor substrate seem to favor the twisted amide mechanism over acid/base catalyzed activation (52, 54). The structural information on *N. meningitidis* EarP confirms that the rhamnosyltransferase does not coordinate a divalent cation and reveals that there are no obvious interactions with positively charged amino acid side chains that could compensate for the lack of this cofactor (127). Therefore, EarP lacks a structural feature that is considered to be important for catalysis by the twisted amide mechanism. The acid/base activation on the other hand has been shown to involve a proton donor that is part of the Asn-X-Ser/Thr glycosylation motif on the acceptor substrate. Neither the protein acceptor EF-P, nor the glycosyltransferase EarP exhibit an appropriate polar residue that reaches into vicinity of the active site and could fulfil this function (127). Additionally, the high pK_a value of the guanidinium ($pK_a = 12.10$) raises the question, how deprotonation by the weak base aspartate would occur under physiological conditions.

Catalytic deprotonation of arginine by negatively charged amino acids was originally proposed for the *Rattus norvegicus* protein-arginine-methyltransferase 1 (PRMT1) but later

shown not to play a role in promoting this transfer reaction (132, 135). Interactions similar to those that are associated with the twisted amide mechanism were instead proposed to properly align the substrate (Figure 7C). These interactions abolish the planarity of the guanidinium in order to prime it for the nucleophilic attack (136). The orientation of the catalytically important amino acids relative to the acceptor arginine in PRMT1 is highly similar to that in EarP, suggesting that a shared mechanism for substrate activation might be possible (Figure 7B and 7C). Specifically, localization and hydrogen bond formation of D16, D20 and Y288 in EarP of *N. meningitidis* are essentially mimicked by E144, E153 and Y35 in PRMT1 (127, 132, 136) (Chapter 3). While this similarity again indicates that a twisted amide-like mechanism for activation might be viable for the rhamnosyltransferase EarP, experimental evidence for this hypothesis is lacking.

Taken together, the functional importance of several amino acid side chains has been demonstrated (Chapter 3). However, their exact molecular role in activation of the acceptor substrate has not yet been conclusively determined. Comparison with reported mechanisms of N-linked post-translational modification provides theoretical insights (Figure 7). However, further experimental analyses will ultimately have to prove, whether a twisted amide-like mechanism or acid/base activation is used to prime arginine for glycosylation.

6.3 Rhamnosyl-arginine is a novel bacterial glycoconjugate

EarP is an inverting glycosyltransferase that mediates the formation of α-rhamnosyl-arginine on the protein acceptor EF-P. Based on this information, polyclonal antibodies targeting this modification were raised in rabbits (Chapter 2, 3). Single mono-rhamnosyl-arginine containing peptide haptens fused to bovine serum albumin were used as immunogens. The purified anti-rhamnosyl-arginine antibodies (anti-Arg^{Rha}) sensitively and specifically bind rhamnosylated EF-P (EF-P^{Rha}), while showing no cross-reactivity with the unmodified elongation factor (Chapter 2, 3). As EF-P rhamnosylation is found in clinically relevant bacteria such as *P. aeruginosa* and *N. meningitidis* (116, 127), antibodies with this specificity could be used in diagnostic applications in order to detect these pathogens.

As a biochemical tool, anti-Arg^{Rha} allows for detection of product formation during *in vitro* rhamnosylation reactions. This enables both the determination of kinetic parameters and the quantification of intracellular TDP-β-ʟ-rhamnose concentrations (Chapter 3). *E. coli*, *P. putida* and *P. aeruginosa* cells contain high micromolar concentrations of this nucleotide sugar. This is in line with previous studies reporting a TDP-β-ʟ-rhamnose concentration of 1 mM in the Gram-positive lactic acid bacterium *Lactococcus lactis* (137). Gram-positive bacteria have been shown to exhibit ʟ-rhamnose containing cell wall polysaccharides that are critical for both pathogenicity and cell viability (138-140). These polymers are synthesized using TDP-β-ʟ-rhamnose as a sugar donor (126), which explains the high demand of this

metabolite within the cells. The O-antigen and the outer core oligosaccharide region of lipopolysaccharide (LPS) of several pathogenic Gram-negative bacteria also contain L-rhamnose derived from this precursor (Figure 8). Besides this, the γ-proteobacterium *P. aeruginosa* is considered to be the most prominent producer of rhamnolipids (126). Due to this frequent use of L-rhamnose in various molecular contexts, it is conceivable that *P. aeruginosa* has a high demand for the activated sugar precursor. Interestingly, *E. coli* and *P. putida* do not produce rhamnolipids and have not been shown to contain L-rhamnose in their LPS. Yet, these bacteria exhibit levels of TDP-β-L-rhamnose that are comparable or even higher than those of *P. aeruginosa* (Chapter 3). The biosynthesis of this metabolite is energy-intensive and involves the subsequent action of RmlA, RmlB, RmlC and RmlD (126, 138) (Figure 8). Due to this elaborate biosynthesis pathway, an important cellular function can be assumed. The high levels of TDP-β-L-rhamnose might not only be necessary to supply for the synthesis of cell wall components or secondary metabolites, but also for the glycosylation of so far unknown protein targets.
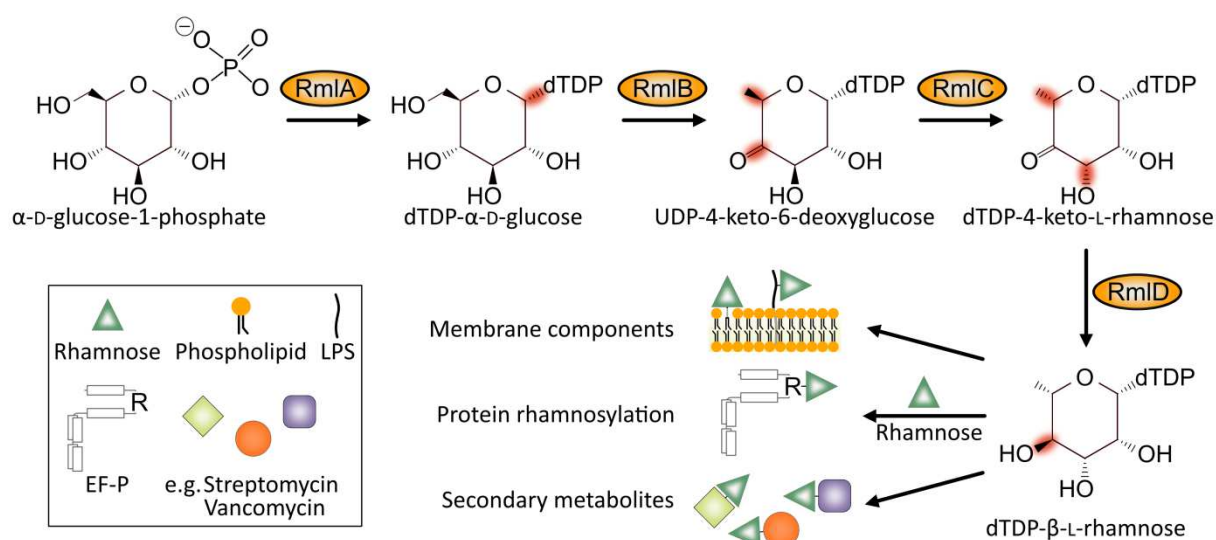
Figure 8. TDP-β-L-rhamnose biosynthetic pathway. Chemical structures of the metabolites used and generated during TDP-β-L-rhamnose biosynthesis are shown. The biosynthesis enzymes RmlA (glucose-1-phosphate thymidyltransferase), RmlB (dTDP-D-glucose-4,6-dehydratase), RmlC (dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase) and RmlD, (dTDP-4-keto-6-deoxy-L-mannose reductase) are indicated by yellow background. The chemical changes occurring during each reaction are indicated by red backgrounds at the corresponding positions. Rhamnose is transferred onto membrane components and proteins and used in the synthesis of secondary metabolites such as streptomycin and vancomycin.

Two distinct pathways for arginine linked N-glycosylation have been identified (113, 114, 116, 117). The effector glycosyltransferase NleB1 of enteropathogenic *E. coli* is injected into the host cell via a type III secretion system and inhibits the innate immune response (113, 114). Specifically, the protein target TRADD is modified with a GlcNAc moiety at a highly conserved arginine and thereby inactivated. Conversely, the rhamnosyltransferase EarP activates its acceptor substrate EF-P by transferring a rhamnose moiety onto the arginine guanidinium (116, 117). These two cases exemplify that the molecular role of arginine

glycosylation is functionally diverse. Consequently, glycosylated arginine residues might be involved in various cellular processes such as stabilization of tertiary structures or even cell-cell recognition that are to date believed to be modulated only by asparagine linked N-glycosylation (100).

In eukaryotes, all N-linked glycoproteins are initially modified with a core polysaccharide (67, 141). While this glycan is certainly modified to yield vastly different mature polysaccharide modifications, the glycan-peptide linkage will almost inevitably remain asparagine-GlcNAc. In bacteria, asparagine has been observed to be directly linked to diNAcBac, GlcNAc, galactose or glucose (142), therefore exhibiting a higher degree of structural diversity. A similar observation can be made for arginine-linked glycans that have been shown to encompass both GlcNAc and rhamnose modifications (113, 114, 116). GlcNAc is a ubiquitously used sugar that is - among involvement in other processes – not only essential for synthesis of the bacterial cell-wall but also as a component and precursor of the most widespread carbohydrate-asparagine linkages (141, 143). Conversely, rhamnose is so far considered as a rare sugar in protein glycosylation (65). This deoxyhexose has been shown to constitute the main component of the asparagine glycosylated platelet aggregation associated protein (PAAP) of *Streptococcus sanguinis*. However, the direct linkage to the amino acid occurs via GlcNAc (144). Thus, arginine glycosylation by EarP results not only in the formation of a novel glycoconjugate (Chapter 3), but also represents the first experimentally verified example of nitrogen-linked rhamnosylation. Considering the functional diversity of arginine glycosylation and the high concentration of TDP-β-L-rhamnose within bacterial cells, it seems likely, that both rhamnosyl-arginine and other arginine-linked glycoconjugates are also involved in other cellular processes.

6.4 EF-P-EarP interaction is sequence and structure dependent

The KOW-like EF-P N-domain is sufficient for rhamnosylation by the glycosyltransferase EarP. NMR titration and biochemical analyses suggest that most of the binding events occur around the EF-P acceptor loop that protrudes from the β-strands β3 and β4 (Chapter 3) (Figure 9 A and B). Besides the main contacts established by the catalytically relevant aspartates D16 and D20 further contacts were derived from the structure of the *N. meningitidis* EarP-EF-P complex. Specifically, K36 and K55 of *N. meningitidis* EF-P form salt bridges with E114 and E89 of EarP (127). Interestingly, position equivalent residues are missing in EF-P of *P. putida.* Moreover, the EarP homologue of *P. putida* interacts both with its cognate EF-P as well as with the non-cognate orthologue of *E. coli* (Chapter 4). The orthologous elongation factors exhibit a sequence identity of only ~30 % and *E. coli* EF-P does not contain lysine residues corresponding to those that were proposed to play a role in protein-protein interaction (Chapter 4). As these amino acids are neither universally

conserved in their corresponding homologues nor necessary for interaction, it seems that the observed contacts are specific for the *N. meningitidis* EarP-EF-P pair. Their role might therefore be to support and enhance a general and sequence-independent mode of EF-P recognition and binding by EarP.

The target site of the bacterial AIDA-associated heptosyltransferase (Aah) was identified as a "short β-strand–short acceptor loop–short β-strand" (145). Similarly, the acceptor loop of elongation factor P is bracketed by two β-strands and thus exhibits a structurally reminiscent architecture (123, 127, 146) (Figure 9 A and B). As EarP can bind and potentially even modify sequentially different EF-Ps (Chapter 4), substrate recognition based mainly on overall structure seems plausible. This mode of acceptor binding is in line with the co-evolution of specific interactions that support the general, structure-based protein-protein interaction. The structure-dependent substrate specificity of Aah was assessed using short peptide stretches with varying sequence and conserved structure (145). Similar experiments conducted on EarP might lead to the identification of a minimal binding motif. Additionally, the positional constraints on the acceptor arginine within the loop could be assessed by locating the acceptor amino acid to different positions within the unstructured region. These experiments could provide further insights into the mechanism of substrate recognition and the architecture of the EarP active site.
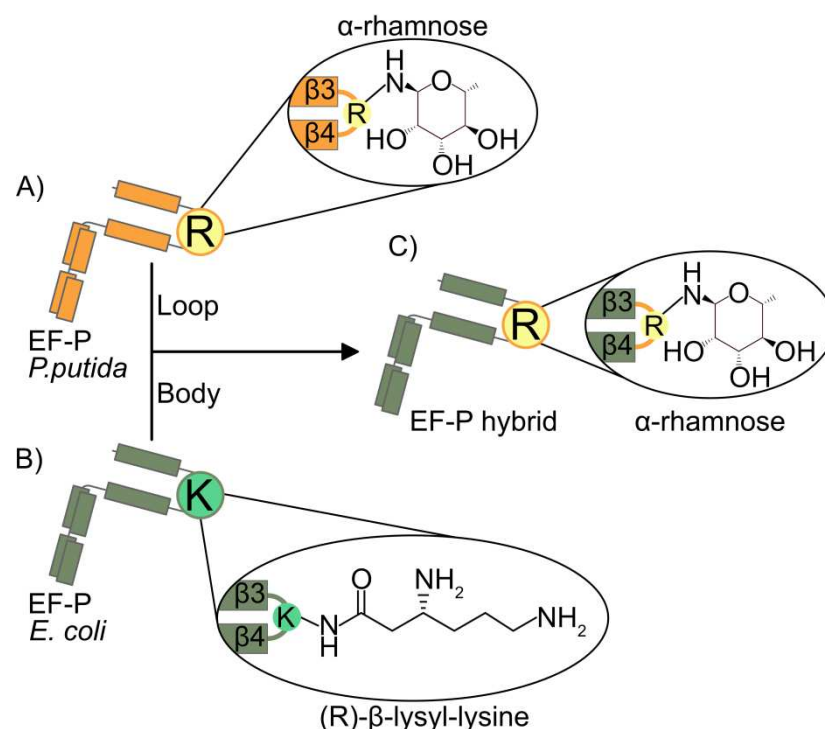
Figure 9. Natural and unnatural strategies for post-translational activation of *P. putida* and *E. coli* EF-P.
All EF-P homologues exhibit an unstructured loop region that is bracketed by the β-strands β3 and β4. This acceptor loop harbors an invariant amino acid that is post-translationally modified. Modified amino acids are shown in single letter code in yellow (R - arginine) and green (K - lysine) spheres. The acceptor loop regions are magnified. Covalent modifications are depicted as chemical structures. **A)** α-rhamnosylation of arginine on *P. putida* EF-P (orange). **B)** (R)-β-lysylation of lysine on *E. coli* EF-P (green). **C)** α-rhamnosylation of arginine on an *E. coli / P. putida* EF-P hybrid containing the *P. putida* acceptor loop and the *E. coli* body.

6.5   The role of glycosylation in EF-P evolution

The bacterial elongation factor EF-P is orthologous to the eukaryotic/archaeal e/aIF5A. The cognate modification enzyme deoxyhypusine synthase is strictly conserved in eukaryotes and archaea (125). By contrast, the modifications that activate the bacterial elongation factor are chemically diverse and species-specific. Beside arginine rhamnosylation two further modes of EF-P activation – β-lysylation of lysine e.g. in *E. coli* and 5-aminopentanolylation e.g. in *B. subtilis* – are known (116, 119, 147). Interestingly, minor sequential changes in *E. coli* EF-P (including the modification site) are sufficient to switch the mode of activation from lysine lysylation to arginine rhamnosylation. Specifically, the substitution of the lysine containing *E. coli* acceptor loop by the homologous *P. putida* loop is sufficient to enable *in vitro* rhamnosylation and *in vivo* activation of the resulting EF-P hybrid (Chapter 4) (Figure 9C). Thus, we provide first insights into the evolution of EF-P and its cognate modification system. However, the driving force behind this process is unknown.

Deletion of *efp* in *E. coli* severely impairs biosynthesis of polyproline-containing peptides and results in reduced growth (116, 117). These deficiencies can be fully complemented by heterologous co-expression of arginine type *efp* and the corresponding modification system *earP* (116). Consequently, the core function of ribosome rescue can be fulfilled by both EF-P isoforms, independent of the chemically distinct modifications. Nevertheless, miniscule variances between different EF-P types might provide an evolutionary advantage. Bacterial genomes encode for a species-specific assortment of polyproline-containing proteins that varies in number (125) and possibly also stalling motif composition. It is conceivable that different motifs form structurally different stalling complexes. Based on the structure of the modification these events could be rescued with varying efficiency. An elongation factor that is – for example due to its modification – more suitable to rescue critical stalling complexes could therefore displace another one. Thus, an adaptation of the modification to the polyproline-proteome could drive evolution.

EF-P recruitment to a stalled ribosome requires the simultaneous presence of prolyl-tRNA (148). Moreover, cryo-EM structures of ribosomes with bound EF-P suggest that interaction with the mRNA might also play a role in this process (149). As proline is encoded by four canonical codons (CCA, CCG, CCT, CCC) that are differently recognized by the three cognate tRNAs proK, proL and proM (150), the codon context might play a role in defining EF-P efficiency for certain stalling motifs. Future experiments will show, which of the differently modified elongation factors performs better in different amino acid and codon contexts. Taken together, this data will ultimately allow conclusions on the selective benefit of either modification type over the other and maybe even the engineering of bacterial strains or elongation factors with improved capabilities with respect to production of polyproline-containing peptides.

6.6 EarP-mediated arginine rhamnosylation and its application in synthetic biology

The comprehensive understanding of EarP-mediated arginine rhamnosylation lays the ground for the application of this cytosolic rhamnosyltransferase in synthetic biology and medicine (Chapters 2, 3 and 4). In a cellular context, directed modification of proteins could be used to fine-tune transcriptional or translational output by targeting corresponding regulatory proteins (15, 116). In a clinical context, selective glycosylation could be used to improve the pharmacokinetic properties of protein therapeutics (151).

The development of a synthetic glycosyltransferase, that could be tailored to modify a soluble protein of choice with a given carbohydrate, would therefore allow for a wide variety of applications. The prerequisite to generate such a synthetic glycosyltransferase is the possibility to modulate enzyme specificity with respect to both the donor and the acceptor substrate (152). Our data provides a structural and biochemical framework for the engineering of such a glycosynthase (Chapter 3). With the construction of the first unnatural acceptor for EarP we made an initial step in identifying a minimal structural recognition motif (Chapter 4). After further characterization, this structure could be implemented into other proteins or recreated using existing structural prerequisites in order to allow specific rhamnosylation by EarP. In combination with our highly sensitive anti-Arg[Rha], the modification could also be used as an affinity tag for protein purification (Chapters 2 and 3).

Complementarily, our data on binding of the nucleotide sugar (Chapter 3) provides the basis for rational mutagenesis of the TDP-β-L-rhamnose binding pocket in order to allow for enhanced donor substrate promiscuity or even a switch in specificity. Initially, the natural substrate specificity needs to be assessed in order to determine structural prerequisites for sugar binding. A subsequent switch of the nucleotide donor to UDP-GlcNAc could be used to synthesize N-actetylglucosaminyl-arginine. The ubiquitous eukaryotic initial linkage between N-actetylglucosamine and asparagine has been used as a precursor for synthesis of homogenous N-linked glycopeptides via transglycosylation (153). An engineered EarP variant could not only allow for direct synthesis of a mono-GlcNAc precursor but also provide the basis for an orthologous pathway. Here, not asparagine but arginine-linked glycans could be specifically glycosylated to yield clinically relevant glycoproteins such as humanized antibodies. If EarP could be further engineered to modify asparagine residues, the resulting variant could likely lead to even more applications.

The modular architecture of EarP and GT-B glycosyltransferases in general allows the generation of chimeric enzyme variants (154). Combination of the N-terminal EF-P binding domain of EarP with the C-terminal GlcNAc binding domain of a structurally related enzyme like MurG could potentially generate a functional glycosyltransferase with entirely new

properties. Similarly, other combinations could be engineered in order to design a modular platform for glycosylation of various target proteins with a wide variety of carbohydrates.

Many approaches in glycoengineering are focusing on the ingenious synthesis of novel lipid-linked oligosaccharides (LLOs) that can be transferred onto proteins using the natural substrate promiscuity of oligosaccharyltransferases (155, 156). While this approach has already been successfully applied, it faces certain constraints with respect to function and subcellular localization. As the name suggests, OSTs catalyze the transfer of oligosaccharides. Many applications are however depending on a monoglycosylated precursor that can be further modified by transglycosylation. Thus an additional step of polysaccharide trimming is necessary (153). Additionally, as OSTs are localized in the periplasm, both the donor LLOs and the target protein have to be translocated to the intermembrane space. The possibility of site-specific monoglycosylation of soluble proteins by a modified variant of the cytosolic enzyme EarP would therefore increase the range of possible targets and at the same time facilitate further processing by transglycosylation.

The virulence of several clinically relevant bacteria has been shown to be dependent on EarP activity (116, 157). The design of compounds that selectively inhibit this reaction might therefore be a promising route to develop novel antibiotics (116). Targeted inhibitor design depends on structural information and will therefore be facilitated by our data.

Taken together, our analysis has not only allowed comprehensive insights into the structural and biochemical mechanism of EF-P glycosylation but also made the rhamnosyltransferase EarP accessible for other fields of research including synthetic biology and medical application.

6.7 Outlook

The novel rhamnosyltransferase EarP was characterized with respect to its structure, substrate binding and catalytic activity. Future perspectives for research on this exceptional enzyme lie mostly in the application for synthetic biology and medical endeavors. Further characterization of the donor and acceptor substrate specificity will provide an understanding of the fundamental prerequisites that determine recognition and binding events. The synthesis of artificial mono-rhamnosylated and potentially cyclized acceptor peptides could greatly help in determining a minimal binding motif and the associated structural constraints in target recognition. Similarly, the chemical or enzymatic synthesis of nucleotide sugars that are structurally related to TDP-β-L-rhamnose will yield information on the natural selectivity of the glycosyltransferase. This information will in turn enable approaches to ultimately enhance or switch the spectrum of usable substrates in EarP-mediated protein glycosylation. In parallel, the feasibility of generating chimeric glycosyltransferases has to be investigated. In a first step, chimeras of sequentially distinct EarP variants will be constructed in order to determine optimal cleavage sites for hybrid-fusions. Thorough understanding of the structural prerequisites for EarP-mediated arginine glycosylation will ultimately also facilitate the rational design of specific inhibitors and thus enable the synthesis of a new class of species-specific antibiotics.

References for Chapters 1 and 6

1. **Gualerzi CO, Pon CL.** 1990. Initiation of mRNA translation in prokaryotes. Biochemistry **29:**5881-5889.
2. **Ramakrishnan V.** 2002. Ribosome structure and the mechanism of translation. Cell **108:**557-572.
3. **Hecht MH, Zarzhitsky S, Karas C, Chari S.** 2018. Are natural proteins special? Can we do that? Curr Opin Struct Biol **48:**124-132.
4. **Walsh CT, Garneau-Tsodikova S, Gatto GJ, Jr.** 2005. Protein posttranslational modifications: the chemistry of proteome diversifications. Angew Chem Int Ed Engl **44:**7342-7372.
5. **Cone JE, Del Río RM, Davis JN, Stadtman TC.** 1976. Chemical characterization of the selenoprotein component of clostridial glycine reductase: identification of selenocysteine as the organoselenium moiety. PNAS **73:**2659-2663.
6. **Hao B, Gong W, Ferguson TK, James CM, Krzycki JA, Chan MK.** 2002. A new UAG-encoded residue in the structure of a methanogen methyltransferase. Science **296:**1462-1466.
7. **Walsh C.** 2006. Posttranslational modification of proteins: Expanding nature's inventory. Roberts and Company Publishers, Greenwood, CO.
8. **Chin JW.** 2014. Expanding and reprogramming the genetic code of cells and animals. Annu Rev Biochem **83:**379-408.
9. **Liu CC, Schultz PG.** 2010. Adding new chemistries to the genetic code. Annu Rev Biochem **79:**413-444.
10. **Young DD, Young TS, Jahnz M, Ahmad I, Spraggon G, Schultz PG.** 2011. An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity. Biochemistry **50:**1894-1900.
11. **Cooley RB, Karplus PA, Mehl RA.** 2014. Gleaning unexpected fruits from hard-won synthetases: probing principles of permissivity in non-canonical amino acid-tRNA synthetases. Chembiochem **15:**1810-1819.
12. **Chin JW.** 2017. Expanding and reprogramming the genetic code. Nature **550:**53-60.
13. **Duan G, Walther D.** 2015. The roles of post-translational modifications in the context of protein interaction networks. PLoS Comput Biol **11:**e1004049.
14. **Johnson LN, Lewis RJ.** 2001. Structural basis for control by phosphorylation. Chem Rev **101:**2209-2242.
15. **Wells L, Whelan SA, Hart GW.** 2003. O-GlcNAc: a regulatory post-translational modification. Biochem Biophys Res Commun **302:**435-441.
16. **Müller MM.** 2018. Post-translational modifications of protein backbones: Unique functions, mechanisms, and challenges. Biochemistry **57:**177-185.
17. **Khoury GA, Baliban RC, Floudas CA.** 2011. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci Rep **1:**90.
18. **Brown CW, Sridhara V, Boutz DR, Person MD, Marcotte EM, Barrick JE, Wilke CO.** 2017. Large-scale analysis of post-translational modifications in *E. coli* under glucose-limiting conditions. BMC Genomics **18:**301.
19. **Su M-G, Weng JT-Y, Hsu JB-K, Huang K-Y, Chi Y-H, Lee T-Y.** 2017. Investigation and identification of functional post-translational modification sites associated with drug binding and protein-protein interactions. BMC Systems Biology **11:**132.
20. **Humphrey SJ, James DE, Mann M.** 2015. Protein phosphorylation: A major switch mechanism for metabolic regulation. Trends Endocrinol Metab **26:**676-687.
21. **Levene PA, Alsberg CL.** 1906. The cleavage products of Vitelling. J Biol Chem **2:**127-133.
22. **Hunter T.** 1995. Protein kinases and phosphatases: The Yin and Yang of protein phosphorylation and signaling. Cell **80:**225-236.
23. **Ubersax JA, Ferrell Jr JE.** 2007. Mechanisms of specificity in protein phosphorylation. Nat Rev Mol Cell Biol **8:**530.

24. **The UniProt Consortium.** 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res **45:**D158-D169.

25. **Nothaft H, Szymanski CM.** 2013. Bacterial protein N-glycosylation: New perspectives and applications. J Biol Chem **288:**6912-6920.

26. **Breton C, Snajdrova L, Jeanneau C, Koca J, Imberty A.** 2006. Structures and mechanisms of glycosyltransferases. Glycobiology **16:**29r-37r.

27. **Coutinho PM, Deleury E, Davies GJ, Henrissat B.** 2003. An evolving hierarchical family classification for glycosyltransferases. J Mol Biol **328:**307-317.

28. **Liang DM, Liu JH, Wu H, Wang BB, Zhu HJ, Qiao JJ.** 2015. Glycosyltransferases: mechanisms and applications in natural product development. Chem Soc Rev **44:**8350-8374.

29. **Breton C, Fournel-Gigleux S, Palcic MM.** 2012. Recent structures, evolution and mechanisms of glycosyltransferases. Curr Opin Struct Biol **22:**540-549.

30. **Bugg TD, Walsh CT.** 1992. Intracellular steps of bacterial cell wall peptidoglycan biosynthesis: enzymology, antibiotics, and antibiotic resistance. Nat Prod Rep **9:**199-215.

31. **Ha S, Walker D, Shi Y, Walker S.** 2000. The 1.9 A crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. Protein Sci **9:**1045-1052.

32. **Hu Y, Chen L, Ha S, Gross B, Falcone B, Walker D, Mokhtarzadeh M, Walker S.** 2003. Crystal structure of the MurG:UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. Proc Natl Acad Sci U S A **100:**845-849.

33. **Somerville C.** 2006. Cellulose synthesis in higher plants. Annu Rev Cell Dev Biol **22:**53-78.

34. **Buschiazzo A, Ugalde JE, Guerin ME, Shepard W, Ugalde RA, Alzari PM.** 2004. Crystal structure of glycogen synthase: homologous enzymes catalyze glycogen synthesis and degradation. Embo j **23:**3196-3205.

35. **Kornberg SR, Zimmerman SB, Kornberg A.** 1961. Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected *Escherichia coli*. J Biol Chem **236:**1487-1493.

36. **Vrielink A, Ruger W, Driessen HP, Freemont PS.** 1994. Crystal structure of the DNA modifying enzyme beta-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. Embo j **13:**3413-3422.

37. **Morera S, Lariviere L, Kurzeck J, Aschke-Sonnenborn U, Freemont PS, Janin J, Ruger W.** 2001. High resolution crystal structures of T4 phage beta-glucosyltransferase: induced fit and effect of substrate and metal binding. J Mol Biol **311:**569-577.

38. **Elshahawi SI, Shaaban KA, Kharel MK, Thorson JS.** 2015. A comprehensive review of glycosylated bacterial natural products. Chem Soc Rev **44:**7591-7697.

39. **Hu Y, Walker S.** 2002. Remarkable structural similarities between diverse glycosyltransferases. Chem Biol **9:**1287-1296.

40. **Campbell JA, Davies GJ, Bulone V, Henrissat B.** 1997. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem J **326 (Pt 3):**929-939.

41. **Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B.** 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res **42:**D490-495.

42. **Lairson LL, Henrissat B, Davies GJ, Withers SG.** 2008. Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem **77:**521-555.

43. **Henrissat B, Sulzenbacher G, Bourne Y.** 2008. Glycosyltransferases, glycoside hydrolases: surprise, surprise! Curr Opin Struct Biol **18:**527-533.

44. **Wu ZL, Ethen CM, Prather B, Machacek M, Jiang W.** 2011. Universal phosphatase-coupled glycosyltransferase assay. Glycobiology **21:**727-733.

45. **Breton C, Bettler E, Joziasse DH, Geremia RA, Imberty A.** 1998. Sequence-function relationships of prokaryotic and eukaryotic galactosyltransferases. J Biochem **123:**1000-1009.

46. **Breton C, Imberty A.** 1999. Structure/function studies of glycosyltransferases. Curr Opin Struct Biol **9:**563-571.

47. **Gloster TM.** 2014. Advances in understanding glycosyltransferases from a structural perspective. Curr Opin Struct Biol **28:**131-141.

48. **Wrabl JO, Grishin NV.** 2001. Homology between O-linked GlcNAc transferases and proteins of the glycogen phosphorylase superfamily. J Mol Biol **314:**365-374.

49. **Chen CI, Keusch JJ, Klein D, Hess D, Hofsteenge J, Gut H.** 2012. Structure of human POFUT2: insights into thrombospondin type 1 repeat fold and O-fucosylation. Embo j **31:**3183-3197.

50. **Lovering AL, de Castro LH, Lim D, Strynadka NC.** 2007. Structural insight into the transglycosylation step of bacterial cell-wall biosynthesis. Science **315:**1402-1405.

51. **Yuan Y, Barrett D, Zhang Y, Kahne D, Sliz P, Walker S.** 2007. Crystal structure of a peptidoglycan glycosyltransferase suggests a model for processive glycan chain synthesis. Proc Natl Acad Sci U S A **104:**5348-5353.

52. **Lizak C, Gerber S, Numao S, Aebi M, Locher KP.** 2011. X-ray structure of a bacterial oligosaccharyltransferase. Nature **474:**350-355.

53. **Sauvage E, Kerff F, Terrak M, Ayala JA, Charlier P.** 2008. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. FEMS Microbiol Rev **32:**234-258.

54. **Lizak C, Gerber S, Michaud G, Schubert M, Fan YY, Bucher M, Darbre T, Aebi M, Reymond JL, Locher KP.** 2013. Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation. Nat Commun **4:**2627.

55. **Ardevol A, Iglesias-Fernandez J, Rojas-Cervellera V, Rovira C.** 2016. The reaction mechanism of retaining glycosyltransferases. Biochem Soc Trans **44:**51-60.

56. **Ardevol A, Rovira C.** 2011. The molecular mechanism of enzymatic glycosyl transfer with retention of configuration: evidence for a short-lived oxocarbenium-like species. Angew Chem Int Ed Engl **50:**10897-10901.

57. **Monegal A, Planas A.** 2006. Chemical rescue of alpha3-galactosyltransferase. Implications in the mechanism of retaining glycosyltransferases. J Am Chem Soc **128:**16030-16031.

58. **Soya N, Fang Y, Palcic MM, Klassen JS.** 2011. Trapping and characterization of covalent intermediates of mutant retaining glycosyltransferases. Glycobiology **21:**547-552.

59. **Sinnott ML, Jencks WP.** 1980. Solvolysis of D-glucopyranosyl derivatives in mixtures of ethanol and 2,2,2-trifluoroethanol. JACS **102:**2026-2032.

60. **Sinnott ML.** 1990. Catalytic mechanism of enzymic glycosyl transfer. Chem Rev **90:**1171-1202.

61. **Lira-Navarrete E, Valero-Gonzalez J, Villanueva R, Martinez-Julvez M, Tejero T, Merino P, Panjikar S, Hurtado-Guerrero R.** 2011. Structural insights into the mechanism of protein O-fucosylation. PLoS One **6:**e25365.

62. **Bause E.** 1984. Model studies on N-glycosylation of proteins. Biochem Soc Trans **12:**514-517.

63. **Wild R, Kowal J, Eyring J, Ngwa EM, Aebi M, Locher KP.** 2018. Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. Science **359:**545-550.

64. **Lis H, Sharon N.** 1993. Protein glycosylation. Structural and functional aspects. Eur J Biochem **218:**1-27.

65. **Lafite P, Daniellou R.** 2012. Rare and unusual glycosylation of peptides and proteins. Nat Prod Rep **29:**729-738.

66. **Hebert DN, Molinari M.** 2012. Flagging and docking: dual roles for N-glycans in protein quality control and cellular proteostasis. Trends Biochem Sci **37:**404-410.

67. **Caramelo JJ, Parodi AJ.** 2015. A sweet code for glycoprotein folding. FEBS Lett **589:**3379-3387.

68. **Vembar SS, Brodsky JL.** 2008. One step at a time: endoplasmic reticulum-associated degradation. Nat Rev Mol Cell Biol **9:**944-957.

69. **O'Connor SE, Imperiali B.** 1996. Modulation of protein structure and function by asparagine-linked glycosylation. Chem Biol **3:**803-812.

70. **Hanson SR, Culyba EK, Hsu T-L, Wong C-H, Kelly JW, Powers ET.** 2009. The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. PNAS **106:**3131-3136.

71. **Eichler J, Koomey M.** 2017. Sweet new roles for protein glycosylation in prokaryotes. Trends Microbiol **25:**662-672.

72. **Valguarnera E, Kinsella RL, Feldman MF.** 2016. Sugar and spice make bacteria not nice: protein glycosylation and its influence in pathogenesis. J Mol Biol **428:**3206-3220.

73. **Karabasheva D, Cole NB, Donaldson JG.** 2014. Roles for trafficking and O-linked glycosylation in the turnover of model cell surface proteins. J Biol Chem **289:**19477-19490.

74. **Bard F, Chia J.** 2016. Cracking the glycome encoder: Signaling, trafficking, and glycosylation. Trends Cell Biol **26:**379-388.

75. **Kreppel LK, Blomberg MA, Hart GW.** 1997. Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. J Biol Chem **272:**9308-9315.

76. **Apweiler R, Hermjakob H, Sharon N.** 1999. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. Biochim Biophys Acta **1473:**4-8.

77. **Drake RR, Jones EE, Powers TW, Nyalwidhe JO.** 2015. Altered glycosylation in prostate cancer. Adv Cancer Res **126:**345-382.

78. **Kudelka MR, Ju T, Heimburg-Molinaro J, Cummings RD.** 2015. Simple sugars to complex disease - mucin-type O-glycans in cancer. Adv Cancer Res **126:**53-135.

79. **Neuberger A.** 1938. Carbohydrates in protein: The carbohydrate component of crystalline egg albumin. Biochem J **32:**1435-1451.

80. **Nothaft H, Szymanski CM.** 2010. Protein glycosylation in bacteria: sweeter than ever. Nat Rev Microbiol **8:**765-778.

81. **Sleytr UB.** 1975. Heterologous reattachment of regular arrays of glycoproteins on bacterial surfaces. Nature **257:**400-402.

82. **Sleytr UB, Thorne KJ.** 1976. Chemical characterization of the regularly arranged surface layers of *Clostridium thermosaccharolyticum* and *Clostridium thermohydrosulfuricum*. J Bacteriol **126:**377-383.

83. **Thibault P, Logan SM, Kelly JF, Brisson JR, Ewing CP, Trust TJ, Guerry P.** 2001. Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin. J Biol Chem **276:**34862-34870.

84. **Shen A, Kamp HD, Grundling A, Higgins DE.** 2006. A bifunctional O-GlcNAc transferase governs flagellar motility through anti-repression. Genes Dev **20:**3283-3295.

85. **Aas FE, Vik A, Vedde J, Koomey M, Egge-Jacobsen W.** 2007. *Neisseria gonorrhoeae* O-linked pilin glycosylation: functional analyses define both the biosynthetic pathway and glycan structure. Mol Microbiol **65:**607-624.

86. **Stimson E, Virji M, Makepeace K, Dell A, Morris HR, Payne G, Saunders JR, Jennings MP, Barker S, Panico M, et al.** 1995. Meningococcal pilin: a glycoprotein substituted with digalactosyl 2,4-diacetamido-2,4,6-trideoxyhexose. Mol Microbiol **17:**1201-1214.

87. **Castric P, Cassels FJ, Carlson RW.** 2001. Structural characterization of the *Pseudomonas aeruginosa* 1244 pilin glycan. J Biol Chem **276:**26479-26485.

88. **Vik A, Aas FE, Anonsen JH, Bilsborough S, Schneider A, Egge-Jacobsen W, Koomey M.** 2009. Broad spectrum O-linked protein glycosylation in the human pathogen *Neisseria gonorrhoeae*. Proc Natl Acad Sci U S A **106:**4447-4452.

89. **Ku SC, Schulz BL, Power PM, Jennings MP.** 2009. The pilin O-glycosylation pathway of pathogenic *Neisseria* is a general system that glycosylates AniA, an outer membrane nitrite reductase. Biochem Biophys Res Commun **378:**84-89.

90. **Szymanski CM, Yao R, Ewing CP, Trust TJ, Guerry P.** 1999. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. Mol Microbiol **32:**1022-1030.

91. **Linton D, Allan E, Karlyshev AV, Cronshaw AD, Wren BW.** 2002. Identification of N-acetylgalactosamine-containing glycoproteins PEB3 and CgpA in *Campylobacter jejuni*. Mol Microbiol **43:**497-508.

92. **Young NM, Brisson JR, Kelly J, Watson DC, Tessier L, Lanthier PH, Jarrell HC, Cadotte N, St Michael F, Aberg E, Szymanski CM.** 2002. Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. J Biol Chem **277:**42530-42539.

93. **Linton D, Dorrell N, Hitchen PG, Amber S, Karlyshev AV, Morris HR, Dell A, Valvano MA, Aebi M, Wren BW.** 2005. Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway. Mol Microbiol **55:**1695-1703.

94. **Glover KJ, Weerapana E, Chen MM, Imperiali B.** 2006. Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the Campylobacter jejuni N-linked glycosylation pathway. Biochemistry **45:**5343-5350.

95. **Weerapana E, Glover KJ, Chen MM, Imperiali B.** 2005. Investigating bacterial N-linked glycosylation: synthesis and glycosyl acceptor activity of the undecaprenyl pyrophosphate-linked bacillosamine. J Am Chem Soc **127:**13766-13767.

96. **Troutman JM, Imperiali B.** 2009. *Campylobacter jejuni* PglH is a single active site processive polymerase that utilizes product inhibition to limit sequential glycosyl transfer reactions. Biochemistry **48:**2807-2816.

97. **Alaimo C, Catrein I, Morf L, Marolda CL, Callewaert N, Valvano MA, Feldman MF, Aebi M.** 2006. Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. Embo j **25:**967-976.

98. **Kelly J, Jarrell H, Millar L, Tessier L, Fiori LM, Lau PC, Allan B, Szymanski CM.** 2006. Biosynthesis of the N-linked glycan in *Campylobacter jejuni* and addition onto protein through block transfer. J Bacteriol **188:**2427-2434.

99. **Feldman MF, Wacker M, Hernandez M, Hitchen PG, Marolda CL, Kowarik M, Morris HR, Dell A, Valvano MA, Aebi M.** 2005. Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. Proc Natl Acad Sci U S A **102:**3016-3021.

100. **Naegeli A, Aebi M.** 2015. Current approaches to engineering N-linked protein glycosylation in bacteria. Methods Mol Biol **1321:**3-16.

101. **Marshall RD.** 1972. Glycoproteins. Annu Rev Biochem **41:**673-702.

102. **Kowarik M, Young NM, Numao S, Schulz BL, Hug I, Callewaert N, Mills DC, Watson DC, Hernandez M, Kelly JF, Wacker M, Aebi M.** 2006. Definition of the bacterial N-glycosylation site consensus sequence. Embo j **25:**1957-1966.

103. **Schäffer C, Messner P.** 2017. Emerging facets of prokaryotic glycosylation. FEMS Microbiol Rev **41:**49-91.

104. **Grass S, Buscher AZ, Swords WE, Apicella MA, Barenkamp SJ, Ozchlewski N, St Geme JW, 3rd.** 2003. 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.

105. **Grass S, Lichti CF, Townsend RR, Gross J, St Geme JW, 3rd.** 2010. The *Haemophilus influenzae* HMW1C protein is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin. PLoS Pathog **6:**e1000919.

106. **Grass S, St Geme JW, 3rd.** 2000. Maturation and secretion of the non-typable *Haemophilus influenzae* HMW1 adhesin: roles of the N-terminal and C-terminal domains. Mol Microbiol **36:**55-67.

107. **Jacob-Dubuisson F, Locht C, Antoine R.** 2001. Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. Mol Microbiol **40:**306-313.

108. **Szymanski CM, Wren BW.** 2005. Protein glycosylation in bacterial mucosal pathogens. Nat Rev Microbiol **3:**225-237.

109. **Ribet D, Cossart P.** 2010. Post-translational modifications in host cells during bacterial infection. FEBS Lett **584:**2748-2758.

110. **St Geme JW, 3rd, Falkow S, Barenkamp SJ.** 1993. High-molecular-weight proteins of nontypable *Haemophilus influenzae* mediate attachment to human epithelial cells. Proc Natl Acad Sci U S A **90:**2875-2879.

111. **Singh DG, Lomako J, Lomako WM, Whelan WJ, Meyer HE, Serwe M, Metzger JW.** 1995. β-Glucosylarginine: a new glucose-protein bond in a self-glucosylating protein from sweet corn. FEBS Lett **376:**61-64.

112. **Konishi T, Ohnishi-Kameyama M, Funane K, Miyazaki Y, Konishi T, Ishii T.** 2010. An arginyl residue in rice UDP-arabinopyranose mutase is required for catalytic activity and autoglycosylation. Carbohydr Res **345:**787-791.

113. **Pearson JS, Giogha C, Ong SY, Kennedy CL, Kelly M, Robinson KS, Lung TW, Mansell A, Riedmaier P, Oates CV, Zaid A, Muhlen S, Crepin VF, Marches O, Ang CS, Williamson NA, O'Reilly LA, Bankovacki A, Nachbur U, Infusini G, Webb AI, Silke J, Strasser A, Frankel G, Hartland EL.** 2013. A type III effector antagonizes death receptor signalling during bacterial gut infection. Nature **501:**247-251.

114. **Li S, Zhang L, Yao Q, Li L, Dong N, Rong J, Gao W, Ding X, Sun L, Chen X, Chen S, Shao F.** 2013. Pathogen blocks host death receptor signalling by arginine GlcNAcylation of death domains. Nature **501:**242-246.

115. **Esposito D, Gunster RA, Martino L, El Omari K, Wagner A, Thurston TLM, Rittinger K.** 2018. Structural basis for the glycosyltransferase activity of the *Salmonella* effector SseK3. J Biol Chem **293:**5064-5078.

116. **Lassak J, Keilhauer EC, Furst M, Wuichet K, Godeke J, Starosta AL, Chen JM, Sogaard-Andersen L, Rohr J, Wilson DN, Haussler S, Mann M, Jung K.** 2015. Arginine-rhamnosylation as new strategy to activate translation elongation factor P. Nat Chem Biol **11:**266-270.

117. **Rajkovic A, Erickson S, Witzky A, Branson OE, Seo J, Gafken PR, Frietas MA, Whitelegge JP, Faull KF, Navarre W, Darwin AJ, Ibba M.** 2015. Cyclic rhamnosylated elongation factor P establishes antibiotic resistance in *Pseudomonas aeruginosa*. mBio **6:**e00823.

118. **Ude S, Lassak J, Starosta AL, Kraxenberger T, Wilson DN, Jung K.** 2013. Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. Science **339:**82-85.

119. **Bailly M, de Crecy-Lagard V.** 2010. Predicting the pathway involved in post-translational modification of elongation factor P in a subset of bacterial species. Biol Direct **5:**3.

120. **Kyrpides NC, Woese CR.** 1998. Universally conserved translation initiation factors. Proc Natl Acad Sci U S A **95:**224-228.

121. **Sumida T, Yanagisawa T, Ishii R, Yokoyama S.** 2010. Crystallization and preliminary X-ray crystallographic study of GenX, a lysyl-tRNA synthetase paralogue from *Escherichia coli*, in complex with translation elongation factor P. Acta Crystallogr Sect F Struct Biol Cryst Commun **66:**1115-1118.

122. **Navarre WW, Zou SB, Roy H, Xie JL, Savchenko A, Singer A, Edvokimova E, Prost LR, Kumar R, Ibba M, Fang FC.** 2010. PoxA, yjeK, and elongation factor P coordinately modulate virulence and drug resistance in *Salmonella enterica*. Mol Cell **39:**209-221.

123. **Yanagisawa T, Sumida T, Ishii R, Takemoto C, Yokoyama S.** 2010. A paralog of lysyl-tRNA synthetase aminoacylates a conserved lysine residue in translation elongation factor P. Nat Struct Mol Biol **17:**1136-1143.

124. **Kawai F, Grass S, Kim Y, Choi KJ, St Geme JW, 3rd, Yeo HJ.** 2011. Structural insights into the glycosyltransferase activity of the *Actinobacillus pleuropneumoniae* HMW1C-like protein. J Biol Chem **286:**38546-38557.

125. **Lassak J, Wilson DN, Jung K.** 2015. Stall no more at polyproline stretches with the translation elongation factors EF-P and IF-5A. Mol Microbiol **99:**219-235.

126. **Giraud M-F, Naismith JH.** 2000. The rhamnose pathway. Curr Opin Struct Biol **10:**687-696.

127. **Sengoku T, Suzuki T, Dohmae N, Watanabe C, Honma T, Hikida Y, Yamaguchi Y, Takahashi H, Yokoyama S, Yanagisawa T.** 2018. Structural basis of protein arginine rhamnosylation by glycosyltransferase EarP. Nat Chem Biol **14:**368-374.

128. **Wang S, Corcilius L, Sharp PP, Rajkovic A, Ibba M, Parker BL, Payne RJ.** 2017. Synthesis of rhamnosylated arginine glycopeptides and determination of the glycosidic linkage in bacterial elongation factor P. Chem Sci **8:**2296-2302.

129. **Kapitonov D, Yu RK.** 1999. Conserved domains of glycosyltransferases. Glycobiology **9:**961-978.

130. **Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW, Aebi M.** 2002. N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. Science **298:**1790-1793.

131. **Hurtado-Guerrero R, Davies GJ.** 2012. Recent structural and mechanistic insights into post-translational enzymatic glycosylation. Curr Opin Chem Biol **16:**479-487.

132. **Fuhrmann J, Clancy KW, Thompson PR.** 2015. Chemical biology of protein arginine modifications in epigenetic regulation. Chem Rev **115:**5413-5461.

133. **Marshall RD.** 1974. The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins. Biochem Soc Symp**:**17-26.

134. **Imperiali B, Shannon KL, Rickert KW.** 1992. Role of peptide conformation in asparagine-linked glycosylation. JACS **114:**7942-7944.

135. **Rust HL, Zurita-Lopez CI, Clarke S, Thompson PR.** 2011. Mechanistic studies on transcriptional coactivator protein arginine methyltransferase 1. Biochemistry **50:**3332-3345.

136. **Zhang R, Li X, Liang Z, Zhu K, Lu J, Kong X, Ouyang S, Li L, Zheng YG, Luo C.** 2013. Theoretical insights into catalytic mechanism of protein arginine methyltransferase 1. PLoS One **8:**e72424.

137. **Boels IC, Beerthuyzen MM, Kosters MH, Van Kaauwen MP, Kleerebezem M, De Vos WM.** 2004. Identification and functional characterization of the *Lactococcus lactis* rfb operon, required for dTDP-rhamnose Biosynthesis. J Bacteriol **186:**1239-1248.

138. **Maki M, Renkonen R.** 2004. Biosynthesis of 6-deoxyhexose glycans in bacteria. Glycobiology **14:**1r-15r.

139. **Caliot E, Dramsi S, Chapot-Chartier MP, Courtin P, Kulakauskas S, Pechoux C, Trieu-Cuot P, Mistou MY.** 2012. Role of the Group B antigen of *Streptococcus agalactiae*: a peptidoglycan-anchored polysaccharide involved in cell wall biogenesis. PLoS Pathog **8:**e1002756.

140. **van Sorge NM, Cole JN, Kuipers K, Henningham A, Aziz RK, Kasirer-Friede A, Lin L, Berends ETM, Davies MR, Dougan G, Zhang F, Dahesh S, Shaw L, Gin J, Cunningham M, Merriman JA, Hutter J, Lepenies B, Rooijakkers SHM, Malley R, Walker MJ, Shattil SJ, Schlievert PM, Choudhury B, Nizet V.** 2014. The classical lancefield antigen of group a *Streptococcus* is a virulence determinant with implications for vaccine design. Cell Host Microbe **15:**729-740.

141. **Helenius A, Aebi M.** 2004. Roles of N-linked glycans in the endoplasmic reticulum. Annu Rev Biochem **73:**1019-1049.

142. **Spiro RG.** 2002. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology **12:**43r-56r.

143. **Schoenhofen IC, McNally DJ, Vinogradov E, Whitfield D, Young NM, Dick S, Wakarchuk WW, Brisson JR, Logan SM.** 2006. Functional characterization of dehydratase/aminotransferase pairs from *Helicobacter* and *Campylobacter*: enzymes distinguishing the pseudaminic acid and bacillosamine biosynthetic pathways. J Biol Chem **281:**723-732.

144. **Erickson PR, Herzberg MC.** 1993. Evidence for the covalent linkage of carbohydrate polymers to a glycoprotein from *Streptococcus sanguis*. J Biol Chem **268:**23780-23783.

145. **Charbonneau M-È, Côté J-P, Haurat MF, Reiz B, Crépin S, Berthiaume F, Dozois CM, Feldman MF, Mourez M.** 2012. A structural motif is the recognition site for a new family of bacterial protein O-glycosyltransferases. Mol Microbiol **83:**894-907.

146. **Choi S, Choe J.** 2011. Crystal structure of elongation factor P from *Pseudomonas aeruginosa* at 1.75 A resolution. Proteins **79:**1688-1693.

147. **Hummels KR, Witzky A, Rajkovic A, Tollerson R, 2nd, Jones LA, Ibba M, Kearns DB.** 2017. Carbonyl reduction by YmfI in *Bacillus subtilis* prevents accumulation of an inhibitory EF-P modification state. Mol Microbiol **106:**236-251.

148. **Katoh T, Wohlgemuth I, Nagano M, Rodnina MV, Suga H.** 2016. Essential structural elements in tRNA(Pro) for EF-P-mediated alleviation of translation stalling. Nat Commun **7:**11657.

149. **Huter P, Arenz S, Bock LV, Graf M, Frister JO, Heuer A, Peil L, Starosta AL, Wohlgemuth I, Peske F, Novacek J, Berninghausen O, Grubmuller H, Tenson T, Beckmann R, Rodnina MV, Vaiana AC, Wilson DN.** 2017. Structural basis for polyproline-mediated ribosome stalling and rescue by the translation elongation factor EF-P. Mol Cell **68:**515-527.e516.

150. **Pavlov MY, Watts RE, Tan Z, Cornish VW, Ehrenberg M, Forster AC.** 2009. Slow peptide bond formation by proline and other N-alkylamino acids in translation. Proc Natl Acad Sci U S A **106:**50-54.

151. **Sinclair AM, Elliott S.** 2005. Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. J Pharm Sci **94:**1626-1635.

152. **McArthur JB, Chen X.** 2016. Glycosyltransferase engineering for carbohydrate synthesis. Biochem Soc Trans **44:**129-142.

153. **Schwarz F, Huang W, Li C, Schulz BL, Lizak C, Palumbo A, Numao S, Neri D, Aebi M, Wang LX.** 2010. A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. Nat Chem Biol **6:**264-266.

154. **Truman AW, Dias MV, Wu S, Blundell TL, Huang F, Spencer JB.** 2009. Chimeric glycosyltransferases for the generation of hybrid glycopeptides. Chem Biol **16:**676-685.

155. **Wacker M, Feldman MF, Callewaert N, Kowarik M, Clarke BR, Pohl NL, Hernandez M, Vines ED, Valvano MA, Whitfield C, Aebi M.** 2006. Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. Proc Natl Acad Sci U S A **103:**7088-7093.

156. **Faridmoayer A, Fentabil MA, Haurat MF, Yi W, Woodward R, Wang PG, Feldman MF.** 2008. Extreme substrate promiscuity of the *Neisseria* oligosaccharyl transferase involved in protein O-glycosylation. J Biol Chem **283:**34596-34604.

157. **Yanagisawa T, Takahashi H, Suzuki T, Masuda A, Dohmae N, Yokoyama S.** 2016. *Neisseria meningitidis* translation elongation factor P and its active-site arginine residue are essential for cell viability. PLoS One **11:**e0147907.

Supplemental information – Chapter 2

Supporting information

Resolving the *α*-Glycosidic Linkage of Arginine-Rhamnosylated Translation Elongation Factor P (EF-P) Triggers Generation of the First Arg[Rha] Specific Antibody **

Xiang Li,[†a] Ralph Krafczyk,[†b] Jakub Macošek,[c] Yu-Lei Li,[ad] Yan Zou,[a] Bernd Simon,[c] Xing Pan,[e] Qiu-Ye Wu,[a] Fang Yan,[d] Shan Li,[e] Janosch Hennig,[c] Kirsten Jung,[b] Jürgen Lassak*[b] and Hong-Gang Hu*[a]

a Department of Organic Chemistry, School of Pharmacy, Second Military Medical University, Shanghai 200433, China

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

  Center for Integrated Protein Science Munich, Ludwig-Maximilians-Universität München, Munich, Germany

c Structural and Computational Biology Unit, EMBL Heidelberg, Heidelberg 69117, Germany

d School of Pharmacy, Wei Fang Medical University, Shandong 261053, China

e Institute of infection and immunity, Taihe hospital, Hubei university of medicine, Shiyan, Hubei 442000

(China)

[†]: These authors contributed equally to this work.

*Email: juergen.lassak@lmu.de; huhonggang_fox@msn.com

Table of Contents

1. General Information

1.1 Materials

All reagents and solvents were purchased from Acros Organics, Alfa Aesar or Sinopharm Chemical Reagent Co. Ltd and were purified when necessary. THF was distilled from sodium/benzophenone ketyl before use. DMF was distilled under reduced pressure from sodium sulfate and stored in flask containing 4 Å molecular sieves. Et_3N and CH_2Cl_2 were distilled from calcium hydride immediately prior to use. All organic extracts were dried over sodium sulfate and concentrated under rotary evaporator. All other commercially obtained reagents and solvents were used directly without further purification. TLC was performed on plates pre-coated with silica gel 60 F_{254} (250 layer thickness). Flash column chromatographic purification of products was finished using forced-flow chromatography on Silica Gel (300-400 mesh). Visualization was accomplished with 5% (v/v) H_2SO_4 in EtOH, UV light, and/or phosphomolybdic acid (PMA) solution.

1.2 HPLC

Protein A chromatography was carried out on an ÄKTA Explorer chromatographic system from Amersham Biosciences with built-in UV, pH and conductivity detectors to monitor column effluent.

Analytical HPLC was run on a SHIMADZU (Prominence LC-20AD) instrument using an analytical column (Grace Vydac "Protein & Peptide C18", 250 X 4.6 mM, 5 μm particle size,

flow rate 1.0 mL/min, rt). Analytical injections were monitored at 214 nm, 254 nm. Semi preparative HPLC was run on a SHIMADZU (LC-6A) instrument using a semi preparative column (Grace Vydac "Peptide C18", 250 X 10 mM, 10 μm particle size, flow rate 4 mL/min). Solution A was 0.1% TFA in water, and solution B was 0.1% TFA in MeCN. Gradient A: A linear gradient of 10% to 10% B over 2 min, then a linear gradient of 10% to 90% B over 25 min. Gradient B: A linear gradient of 1% to 1% B over 3 min, then a linear gradient of 1% to 35% B over 25 min.

1.3 Mass spectrometry and NMR of small moleculars

^{1}H-NMR and ^{13}C-NMR spectra were obtained on a Bruker Avance 300 MHz or 600 MHz NMR Spectrometer. The chemical shifts of protons are given on the δ scale, ppm, with tetramethylsilane (TMS) as the internal standard. HR-Q-TOF-MS was measured on an Agilent 6538 UHD Accurate Mass Q-TOF LC/MS mass spectrometer.

2. Nuclear magnetic resonance spectroscopy

In order to determine the configuration of rhamnose when bound to EF-P, the rhamnosylated protein was expressed in M9 minimal medium with ^{13}C-Glucose and ^{15}N-ammonium chloride as sole carbon and nitrogen source, respectively. Triple resonance experiments[1] were used to assign backbone chemical shifts of the protein. The rhamnose moiety was assigned, using a sensitivity-enhanced ^{13}C-edited NOESY-HSQC with simultaneous evolution of ^{13}C and ^{15}N during t_2[2]. Three undecoupled, sensitivity-enhanced ^{13}C-HSQC[3] were acquired to obtain the $^1J_{CH}$ coupling of the anomeric carbon to its proton (H1'-C1'). One without ^{13}C decoupling in the ^{1}H direct dimension during acquisition to obtain the $^1J_{CH}$ coupling constant resolved in the ^{1}H dimension with 4096 points (160 ms acquisition time) for sufficient resolution to allow reliable fitting of the coupling constant. The second and third without ^{1}H decoupling during ^{13}C chemical shift evolution to resolve the $^1J_{CH}$ coupling in the ^{13}C indirect dimension was acquired with 3584 points (63.6 ms) to obtain sufficient resolution and with 600 points (10.7 ms) to not resolve ^{13}C-^{13}C coupling (see figure 2c). In order to acquire the ^{13}C dimension with large spectral width to include the methyl, aliphatic, sugar, and aromatic regions with sufficient carbon decoupling, but without sideband artifacts we employed low-power broadband heteronuclear decoupling based on optimal control theory[4]. All spectra were processed using NMRPipe[5] with a gaussian window function and zero filling. Peak detection and fitting of the coupling constant has been used with in-built functions in NMRPipe

(nlinLS) using gaussian fitting. The coupling constants derived from fitting the three spectra are consistent.

3. Chemical synthesis

3.1 2,3,4-Tri-O-acetyl-6-deoxy-α-L-mannopyranosyl chloride (3)

Acetyl chloride (60.0 mL, 843.6 mmol) was added dropwise into a round flask containing L-Rhamnose (**2**, 25.0 g, 152.4 mmol). The mixture was stirred for 48 h under Ar. Dichloromethane (300 mL) was added and the organic layer was washed with water (3*100 mL) and then saturated sodium bicarbonate (3*100 mL). The organic layer was dried over sodium sulfate and concentrated. The residue was purified by column chromatography eluting with 4/1 petroleum ether/ethyl acetate to give **3** as a white foam (39.9 g, 85 %). ^{1}H NMR (300 MHz, CDCl$_3$): δ 5.91 (d, J = 0.8 Hz, 1 H), 5.56 (dd, J = 10.2 Hz, 4.2 Hz, 1 H), 5.37 (q, J = 1.7 Hz, 1 H), 5.15 (t, J = 10.1 Hz, 1 H), 4.17- 4.08 (m, 1 H), 2.15 (s, 3 H), 2.06 (s, 3 H), 1.98 (s, 3 H), 1.26 (d, J = 6.4 Hz, 3H). ^{13}C NMR (75 MHz, CDCl$_3$): δ 169.9, 169.8, 169.7, 89.0, 71.9, 70.3, 69.4, 67.7, 20.8, 20.8, 20.6, 17.1.

3.2 2,3,4-Tri-O-acetyl-6-deoxy-α-L-mannopyranosyl isothiocyanate (4)

A mixture of tetrabutylammonium iodide (9.18 g, 25.0 mmol), KSCN (4.85 g, 50.0 mmol), and molecular sieve (4 Å, 20 g) in dry acetonitrile (200 mL) was stirred at room temperature under argon for 2 h. Then compound **3** (7.70 g, 25.0 mmol) was added to the solution and the mixture was refluxed for another 3 h. Then the mixture was filtered, and the filtrate was concentrated *under vacuum*. The residue was purified by column chromatography with 4/1 petroleum ether/ethyl acetate to yield compound **4** (5.80 g, 70 %) as a white solid. ^{1}H NMR

(300 MHz, CDCl$_3$): δ 5.46 (d, J = 1.0 Hz, 1 H), 5.29- 5.24 (m, 2 H), 5.10- 5.03 (m, 1 H), 4.01- 3.96 (m, 1 H), 2.15 (s, 3 H), 2.06 (s, 3 H), 1.99 (s, 3 H), 1.26 (d, J = 6.2 Hz, 3 H). ^{13}C NMR (75 MHz, CDCl$_3$): δ 170.8, 170.6, 170.6, 169.2, 143.3, 82.8, 77.2, 74.0, 71.7, 67.9, 61.7, 56.1, 23.3, 20.7, 20.6, 20.6. HR-Q-TOF-MS: calcd. for C$_{13}$H$_{17}$NNaO$_7$S$^+$ [M+Na]$^+$ m/z, 354.0618; found, 354.0620.

3.3 N-(2,3,4-Tri-O-acetyl-6-deoxy-α-L-mannopyranos -1-yl) thiourea (5)

Gaseous ammonia was passed through a solution of **4** (3.31 g, 10 mmol) in anhydrous THF (50 mL) for 1 h. Compound **5** was obtained as a white solid after removal of solvent *in vacuum* (3.3 g, 99 %) and used in the next step without further purification. ^{1}H NMR (300 MHz, CDCl$_3$): δ 7.72 (s, 1 H), 6.86 (s, 2 H), 5.45-5.29 (m, 3 H), 5.08 (t, J = 9.5 Hz, 1 H), 3.91-3.86 (m, 1 H), 2.14 (s, 3 H), 2.03 (s, 3 H), 1.99 (s, 3 H), 1.22 (d, J = 6.2 Hz, 3 H). ^{13}C NMR (75 MHz, CDCl$_3$): δ 184.4, 170.1 (2 C), 167.0, 80.9, 70.5, 68.7, 67.8, 65.9, 20.9, 20.8, 20.8, 17.4. HR-Q-TOF-MS: calcd. for C$_{13}$H$_{20}$N$_2$NaO$_7$S$^+$ [M+Na]$^+$ m/z, 371.0883; found, 371.0887.

3.4 1-(tert-Butoxycarbonyl)-3-(2,3,4-Tri-O-acetyl-6-deoxy-α-L-mannopyranos -1-yl) -2-ethylisothiourea (6)

A solution of **5** (2.8 g, 8.0 mmol) and EtI (0.76 mL, 9.6 mmol) in 40 mL of anhydrous CH$_3$OH was stirred under reflux for 4 h. The solvent was removed *in vacuo* and the residue was desolved in dry DCM (50 mL) by adding Et$_3$N (2.24 mL, 16.0 mmol), (Boc)$_2$O (1.96 g,

9.0 mmol) and catalytic amount of DMAP. After the mixture was stirred at room temperature overnight, it was washed with water and brine. The organic layer was dried by Na_2SO_4 and concentrated under vacuum. The residue was purified by flash chromatography eluting with 5:1 petroleum ether/acetone and further crystallized from ether to yield compound **6** (2.86 g, 75 %) as a white solid. ^{1}H NMR (600 MHz, CDCl$_3$): δ 5.39 (s, 1 H), 5.17 (s, 1 H), 5.04- 5.01 (m, 2 H), 3.62- 3.57 (m, 1 H), 3.13- 3.02 (m, 2 H), 2.24 (s, 3 H), 2.05 (s, 3 H), 1.98(s, 3 H), 1.45 (s, 9 H), 1.30- 1.25 (m, 6H). ^{13}C NMR (150 MHz, CDCl$_3$): δ 170.3, 170.2, 169.9, 80.1, 79.3, 72.1, 71.1, 70.0, 69.4, 28.1 (3 C), 25.5, 20.8 (2 C), 20.6, 17.5, 13.9. ESI MS: calcd. for $C_{20}H_{33}N_2O_9S^+$ [M+H]$^+$ m/z, 477.1901; found, 477.1909.

3.5 Synthesis of the title hapten glycopeptide 1

After a mixture of 2-Cl-trityl-Cl resin (400.0 mg, active Cl, 0.20 mmol), Fmoc-Leu-OH (354.0 mg, 1.0 mmol) and DIEA (365.0 µL) in DMF (5 mL) was shaken on a vortex mixer at rt overnight, the resin was filtered off and washed several times with MeOH, DCM and DMF. The Leu-linked resin was then used for the construction of full length peptide/glycopeptides. The protocols employed were: deprotection of Fmoc with 20% piperidine in DMF (15 min) and peptide coupling using 5 eq. of amino acid, 4.5 eq. of HCTU and 10 eq. of DIEA. All coupling reactions were set to perform 1 h at rt. Amino acids Fmoc-Gly-OH, Fmoc-Orn(Aloc)-OH, Fmoc-Gly-OH, Fmoc-Cys-OH, and AcOH were sequentially installed to construct peptide after removal of allyloxycarbonyl group on the resin. Then, a solution of compound **6**

(3 eq.), Et_3N (10 eq.) and $AgNO_3$ (3 eq.) in 5 mL of anhydrous DMF was added into the resin. After the mixture was shaken for 8 h in the dark, the resultant mixture was filtered and the resin was washed thoroughly with DMF, MeOH, and DCM to compound **8**. To remove the acetyl groups from the glycopeptide on the resin, the peptide-loaded resin was treated with a mixture of NH_2-NH_2 in DMF (5 %, 10 mL) for 10 h at rt. The resin was filtered off and washed thoroughly with DMF, H_2O, MeOH, and DCM. Then, the peptide-loaded resin was treated with a mixture of TIPS/TFA (5:95, 10 mL) for 2 h at rt. The resin was filtered off and washed with TFA. The washings were combined and condensed in vacuum to give crude **1**. The crude product was dissolved in water and purified by HPLC (conditions: Grace Vydac "Peptide C18", 250 X 10 mM, 10 μm particle size, suitable ratio of acetonitrile-0.1%TFA in water-0.1%TFA, 4 mL/min) to give **1** (18.0 mg, isolated yield 27 %) as a white solid. ^1H NMR (600 MHz, DMSO-$d6$) : δ 8.29-7.40 (signal of amide protons), 5.33 (s, 1 H), 4.84 (d, J = 9.2 Hz, 1H), 4.34- 4.21 (m, 3H), 3.75- 3.73 (m, 4H), 3.67 (s, 1H), 3.28- 3.25 (m, 3 H), 3.17- 3.12 (m, 3H), 2.79- 2.76 (m, 1 H), 2.71- 2.67 (m, 1 H), 2.46 (t, J = 6.4 Hz, 1H), 1.89 (s, 3 H), 1.76- 1.70 (m, 1 H), 1.64- 1.55 (m, 1 H), 1.55- 1.45 (m, 5H), 1.13 (d, J = 6.1 Hz, 3 H), 0.89 (d, J = 6.6 Hz, 3 H), 0.84 (d, J = 6.6 Hz, 3 H). ^{13}C NMR (150 MHz, DMSO-d_6): δ 173.9, 171.3, 170.4, 169.8, 168.7, 168.5, 155.4, 78.2, 73.6, 72.8, 71.2, 69.9, 55.4, 52.1, 50.2, 42.1, 41.5, 10.1, 29.0, 25.8, 24.6, 24.2, 22.8, 22.5, 21.3, 17.7. HR-Q-TOF-MS: calcd. for $C_{27}H_{49}N_8O_{11}S^+$ [M+H]$^+$ m/z, 693.3236; found 693.3246.
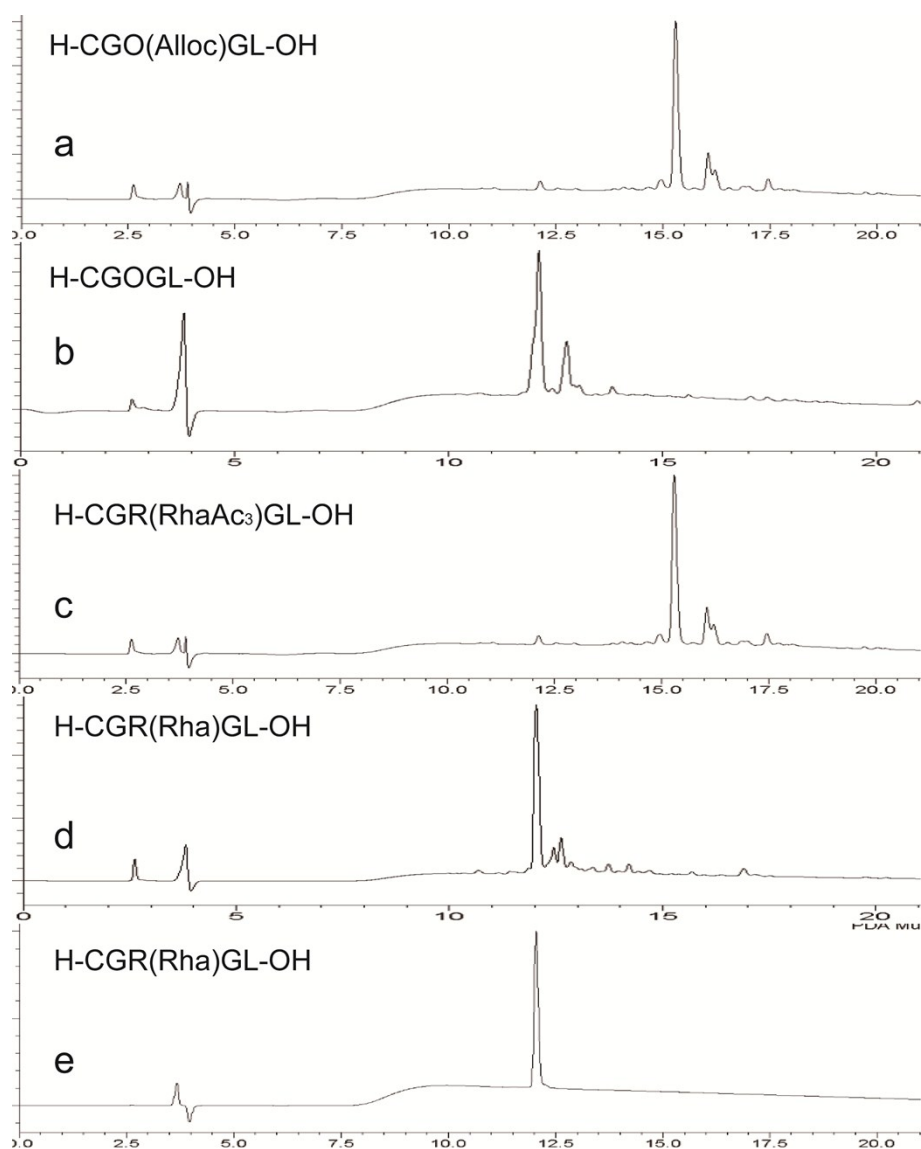
Fig S1 Analytical HPLC traces of synthetic intermediates and title compounds about gylcopeptide **1** respectively. All products were detected after TIPS/TFA cleavage of the resin. a) Full protected peptide b); Peptide **7**; c) Glycosylation of amino group of Orn with **6**/TEA/AgNO$_3$/DMF; d) NH$_2$NH$_2$ mediated removing of Ac groups, e) Analytic trace of the purified gylcopeptide **1**. Analytical HPLC was run on a SHIMADZU (Prominence LC-20AD) instrument using an analytical column (Grace Vydac "Protein & Peptide C18", 250 X 4.6 mM, 5 μm particle size, flow rate 1.0 mL/min, rt). Solution A was 0.1% TFA in water, and solution B was 0.1% TFA in MeCN. Gradient for **a-e**: A linear gradient of 10% to 10% B over 2 min, then a linear gradient of 10% to 90% B over 25 min.

3.6 X-ray crystallography of compound 5

Single crystal of compound **5** was obtained by slow evaporation of dichloromethane/n-hexane solution of **5** at room temperature. The X-ray single crystal diffraction data for **5** was collected on Bruker APEX DUO diffractometers with Mo Kα radiation (λ = 0.71073 Å) at 298 ± 2 K in the ω-2θ scanning mode. The structures were solved by direct methods using the SHELXS-97 program[6] and refined by full-matrix least-squares techniques (SHELXL-97) on F^2.

Anisotropic thermal parameters were assigned to all non-hydrogen atoms. The organic hydrogen atoms were generated geometrically. CCDC-1469830 contains the supplementary crystallographic data for this paper. These data can be obtained free of charge from the Cambridge Crystallographic Data Center via www.ccdc.cam.ac.uk/data_request/cif.
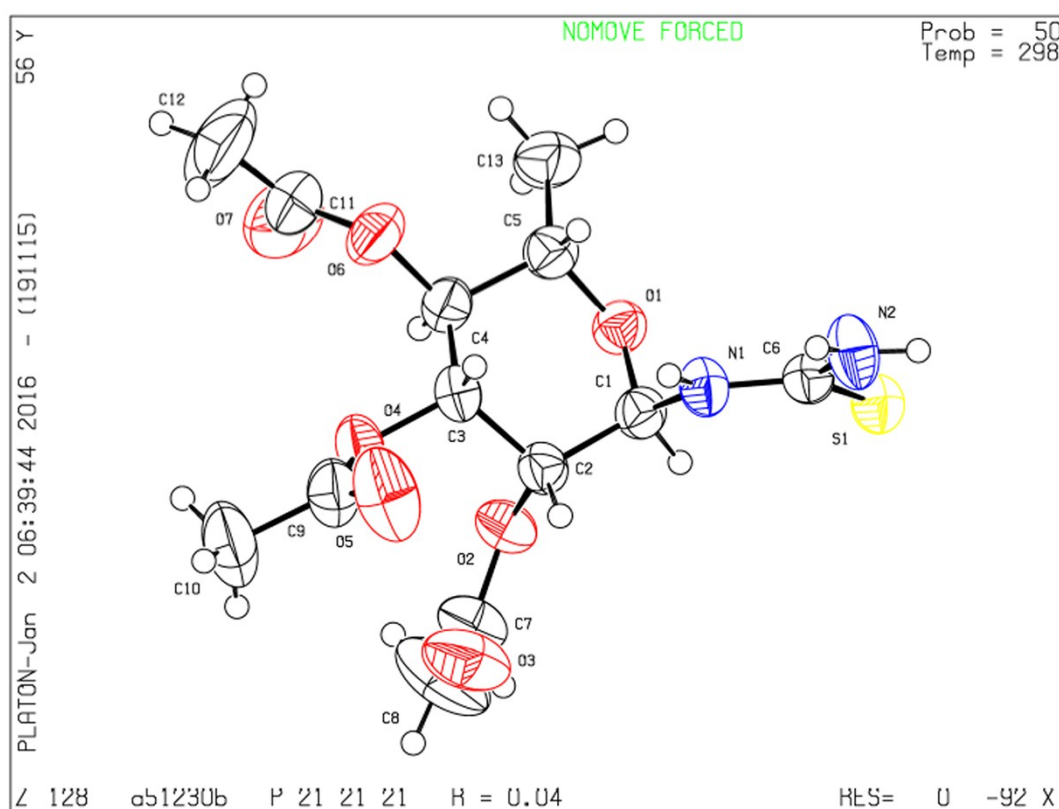
Fig S2 Single-crystal structure of Compound 5.

4. Methods for antibody generation and purification

4.1 Conjugation of glycopeptides to BSA

Glycopeptides were conjugated with Bovine serum albumin (BSA) for inoculation and ELISA analysis, respectively, using Sulfo-GMBS as the crosslinker by following the manufacturer's instruction. BSA was mixed with indicated peptides at a molecular ratio of 1:1.

4.2 Rabbit immunization

Two rabbits (male, 2- 2.5 Kg) were immunized three times at three-week intervals. P1(Rha)-BSA was injected together with the complete and incomplete form of Freund's adjuvant for the 1st and 2nd/3rd immunization, respectively. For each immunization 2 mg peptide was used. Serum samples were collected prior to immunization (pre-bleed) and 1 week after the final immunization for subsequent analyses.

4.3 Affinity purification of antibody

Firstly, the crude anti sera were purified over Protein A Sepharose 4B fast flow. Protein A chromatography was carried out on a 1.1 cm D × 10 cm H column. The column was equilibrated with 25 mM PBS, 150 mM sodium chloride, pH 7.2 buffer, washed with equilibration buffer, and eluted with 100 mM glycine-HCl, pH 2.8. The eluate was collected and neutralized with 1 M tris base. Turbidity and protein concentration were determined following neutralization of the elution pool. Next our anti sera were further purified with BSA Sepharose 4B Fast Flow & naked peptide-BSA Sepharose 4 Fast Flow. Affinity resin consisting of BSA and naked peptide-BSA immobilized onto agarose were prepared by covalently attaching BSA and naked peptide-BSA to agarose beads using CNBr-activated Sepharose 4B (GE Healthcare) according to the manufacturer's instructions. Eluting procedure was same as that of the Protein A purification. Fractions containing rhamnosylarginine antibody were determined by SDS-PAGE and ELISA analysis.

4.4 ELISA of crude anti-Arg-Rha anti sera and purified antibody

BSA or BSA-peptides (5 μg/mL) were dissolved in coating buffer (0.032 M Na_2CO_3 and

0.068 M NaHCO$_3$, pH 9.6). 100 μL of each sample containing 0.5 μg of protein was added to a Nunc-Immuno Maxisorp 96-well plate. The plate was incubated at 4°C overnight on a nutator. The wells were then washed three times with wash buffer (phosphate buffer solution supplemented with 0.05% Tween-20) and blocked with 10% Fetal Bovine Serum (FBS) in the wash buffer (200 μL/well) for 120 min at 37°C. After removing the blocking solution, crude anti sera (including the neutralized anti sera) or purified antibody diluted in wash buffer were added to the wells (100 μL/well) and incubated for 60 min at 37°C. The neutralized crude anti sera were prior incubated with BSA for 60 min at 37°C before adding to the wells. The wells were then washed three times with the wash buffer. Goat-anti rabbit-HRP secondary antibody (1:5,000 dilution) was added to the wells (100 μL/well) and incubated for 60 min at 37°C. After extensive wash, 100 μL of fresh ELISA substrate solution (4.86 mL of 0.1 M citric acid and 5.14 mL of 0.2 M Na$_2$HPO$_4$ supplemented with 4 mg o-phenylenediamine and 15 μL of H$_2$O$_2$) was added to each well and incubated for 5-10 min at room temperature followed by addition of 50 μL of 2 M H$_2$SO$_4$ to quench the reaction. The absorbance at 490 nm was measured on a BioRad microplate reader.
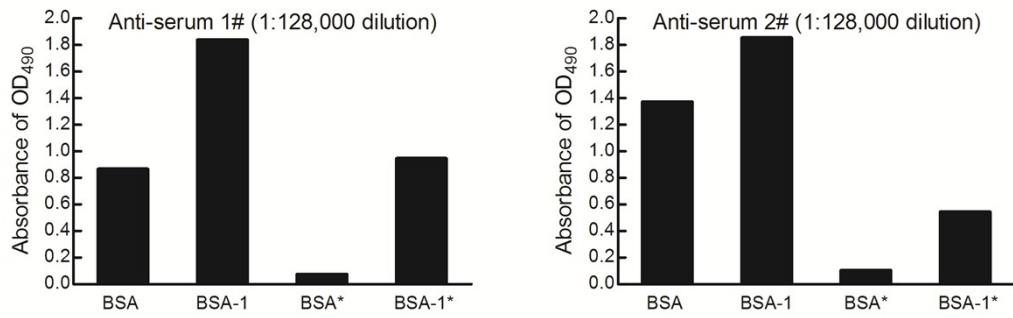
Fig S3 Anti-Arg^Rha crude sera immunized by BSA-glycoconjugate can recognize the arginine rhamnosylated BSA with high affnity. ELISA analysis of two batches of anti-sera immunized by glycopeptide **1**. Anti-serum 1# and anti-serum 2# were diluted 128,000 fold and subjected to indirect ELISA experiments against BSA-glycoconjugate or independent BSA. BSA-1: BSA-Arg^Rha, *: The anti-sera were neutralized with BSA when detecting immune reactivity. Absorbance at 490 nm is shown.
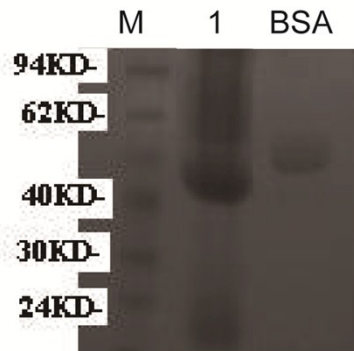
Fig S4 SDS-PAGE of the crude antibody. M: Marker, 1: Crude Anti-serum：20ul (4 mg/ml), BSA：20ug.
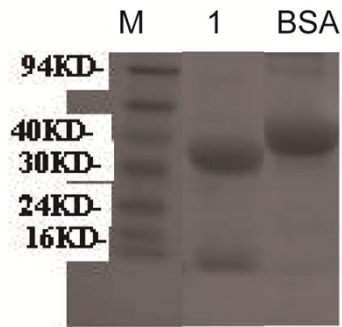
Fig S5 SDS-PAGE of the antibody after affinity purification. M: Marker, 1: Anti-Arg-Rha

Rabbit IgG after specific purification：20ul (4 mg/ml), BSA：20ug.
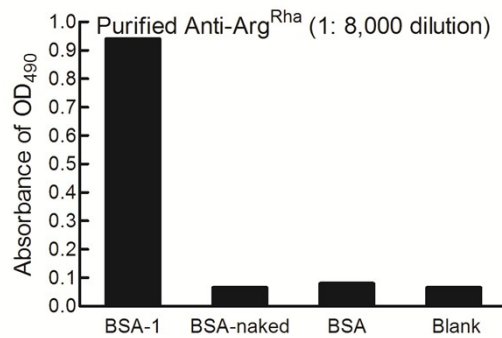
Fig S6 Purified anti-Arg^{Rha} antibody immunized by BSA-glycoconjugate can recognize the

BSA-glycoconjugate with high affinity and specificity. purified anti-Arg^{Rha} antibody were

diluted 8,000 fold and subjected to indirect ELISA experiments against BSA-glycoconjugate,

BSA-non-glycoconjugate and independent BSA. BSA-1: BSA-Arg^{Rha} (BSA-glycoconjugate),

BSA-naked: BSA-Arg (BSA-non-glycoconjugate). Absorbance at 490 nm is shown.

5. Specificity of the anti-*Arg^Rha* for EF-P^Rha

5.1 Bacterial strains and growth conditions

Strains and plasmids used in this study are listed in **Table S1**. *P. aeruginosa, S. oneidensis* MR-1 and *E. coli* were routinely grown in Lysogeny broth (LB) according to Miller at 30°C (for *S. oneidensis*) and 37°C (for *E. coli* and *P. aeruginosa*). When required, media were solidified by using 1.5% (w/v) agar. The optical density at 600 nm (OD600) of cultures was monitored. If necessary, media were supplemented with 30 µg/ml chloramphenicol, 50 µg/ml kanamycin sulfate, and/or 100 µg/ml ampicillin sodium salt. For promoter induction from plasmids pET SUMO, paCYC-DUET1 and derivatives 1 mM Isopropyl-*β*-D-thiogalactopyranosid (IPTG) was added to liquid medium.

5.2 Construction of plasmids for protein overproduction

Genes and gene fragments to be overexpressed were amplified from template *S. oneidensis* genomic DNA. The resulting PCR products for *efp* and *earP* were ligated into pET SUMO (Invitrogen) and pACYC-DUET1 respectively resulting in an N-terminal SUMO-tag fusion for EF-P (pET SUMO-*efp*$_{S.o.}$) and a His$_6$-tag fusion for EarP (paCYC-Duet1-e*arP*$_{S.o.}$). Routinely, chemically competent *E. coli* as strain for high-efficiency transformation[7].

5.3 Production and purification of unmodified and rhamnosylated EF-P

N-terminal His$_6$-SUMO-tagged EF-P$_{S.o.}$ was overproduced in *E. coli* BL21 (DE3) by addition of 1 mM IPTG to exponentially growing cells. After induction, cells were incubated at 16 °C

overnight. Rhamnosylated EF-P$_{S.o.}$(EF-P^{Rha}) was produced by co-overexpression of paCYC-Duet1-e$arP_{S.o.}$. Cells were lysed by sonication and His$_6$-tagged protein was purified using Ni-NTA (Qiagen) and 250 mM imidazole. The His$_6$-SUMO-tag was removed by incubation with 1 u/mg His$_6$-ulp1[8] for 16 h and pure EF-P variants were collected in the flowthrough after repeated metal chelate affinity chromatography. For use in later experiments and NMR spectroscopy, proteins were dialyzed against storage buffer. Protein concentrations were determined as described by Bradford[9].

5.4 Immunodetection analysis of S. oneidensis EF-P

Electrophoretic separation of proteins was carried out using SDS-PAGE as described by Laemmli[10]. Separated proteins were visualized in-gel using 0.5 % (v/v) 2-2-2-trichloroethanol [11] and transferred onto nitrocellulose membrane by vertical Western blotting. Antigens were detected by 0.2 µg/ml Anti-EF-P or 2 µg/ml or dilutions of Anti-Arg^{Rha} respectively. Primary antibodies (rabbit) were targeted by 0.2 µg/ml alkaline phosphatase conjugated Anti-RABBIT IgG (H&L) (GOAT) antibody (Rockland) and target proteins were visualized by addition of substrate solution (50 mM sodium carbonate buffer, pH 9.5, 0.001% (w/v) Nitro-Blue-Tetrazolium, 0,045% (w/v) 5-Bromo-4-chloro-3-indolylphosphate).

To assess the specificity of Anti-Arg^{Rha} towards rhamnosyl arginine, 0.5 µg of purified, unmodified and modified (rhamnosylated) EF-P (EF-P^{Rha}) were subjected to SDS-PAGE and Western Blot analysis as described above and targeted by 0.2 µg/ml Anti-EF-P and Anti-

Arg^{Rha} respectively.

Minimal Anti-Arg^{Rha} concentrations for detection of EF-P were determined by employing 0.5 µg of EF-P^{Rha} and decreasing amounts of Anti-Arg^{Rha} (2 µg/ml, 1 µg/ml, 0.4 µg/ml, 0.2 µg/ml, 0.1 µg/ml, 0.04 µg/ml, 0.02 µg/ml) for detection.

Antigen detection limits were determined by subjecting decreasing amounts of EF-P^{Rha} (0.5 µg, 0.25 µg, 0.125 µg, 62.5 ng, 31.25 ng, 15.6 ng, 7.8 ng, 3.9 ng) to SDS-PAGE and Western Blot analysis as described above and detection with 2 µg/ml and 0.2 µg/ml Anti-Arg^{Rha} respectively.

To investigate cross-reactivity of Anti-Arg^{Rha} against free L-arginine, L-fucose or L-rhamnose, varying concentrations of EF-P^{Rha} (0.015 µM, 0.15 µM, 1.5 µM 15 µM) and putative competitors (15 µM, 1 mM 5 mM, 15 mM) were preincubated with Anti-Arg^{Rha} prior to EF-P^{Rha} detection.

In vivo detection of EF-P and EF-P^{Rha} was carried out using total cell lysates of *Pseudomonas aeruginosa* PAO1 wildtype, *Shewanella oneidensis* MR-1 wildtype as well as *S. oneidensis* mutant strains carrying markerless inframe deletions[12] in *efp* (Δ*efp*), *earP* (Δ*earP*) and *rmlC* (Δ*rmlC*). Cells were grown to an OD600 of 1. About 15 µg total protein were subjected to SDS-PAGE and Western Blot analysis as described above. EF-P and EF-P^{Rha} were detected using 0.2 µg/ml Anti-EF-P and 2 µg/ml Anti-Arg^{Rha} respectively.

Table S1: Strains and plasmids used for EF-P detection

Strain	Genotype	Purpose	Reference
E.coli DH5αλpir	*recA*1 *gyrA* (*lacIZYA-argF*) (80d *lac* [*lacZ*] M15) *pir* RK6		[14]
E. coli BL21 (DE3)	*fhuA*2 [lon] *ompT gal* (λ DE3) [dcm] Δ*hsdS* λ DE3 = λ s*BamH*Io Δ*EcoR*I-B int::(*lacI*::P*lac*UV5::T7 gene1) i21 Δ*nin*5	Overproduction of EF-P	[15]
P. aeruginosa PAO1	*Pseudomonas aeruginosa* PAO1 wildtype		
S. oneidensis MR-1	*Shewanella oneidensis* MR-1 wildtype	*in vivo* detection of EF-P	[16]
S. oneidensis MR-1 Δ*efp*	Δ*efp* (SO_2328)	*in vivo* detection of EF-P	[13]
S. oneidensis MR-1 Δ*earP*	Δ*earP* (SO_2329)	*in vivo* detection of EF-P	[13]
S. oneidensis MR-1 Δ*rmlC*	Δ*rmlC* (SO_3160)	*in vivo* detection of EF-P	[13]
Plasmid	**Features**		
pET SUMO	Expression vector, Kan^r cassette, pBR322 origin, SUMO coding sequence, *lacI* coding sequence, ROP coding sequence, IPTG inducible *lac* operator		Invitrogen
pET SUMO-*efp*_S.o.	C-terminal His_6-Tag *efp* from *S. oneidensis* in pET SUMO		This study
paCYC-Duet1	Expression vector, Cam^r cassette, P15A replicon, *lacI* coding sequence, IPTG inducible lac operator		Novagen
paCYC-Duet1-e*arP*_S.o.	*earP* from *S. oneidensis* in paCYC-Duet1		This study
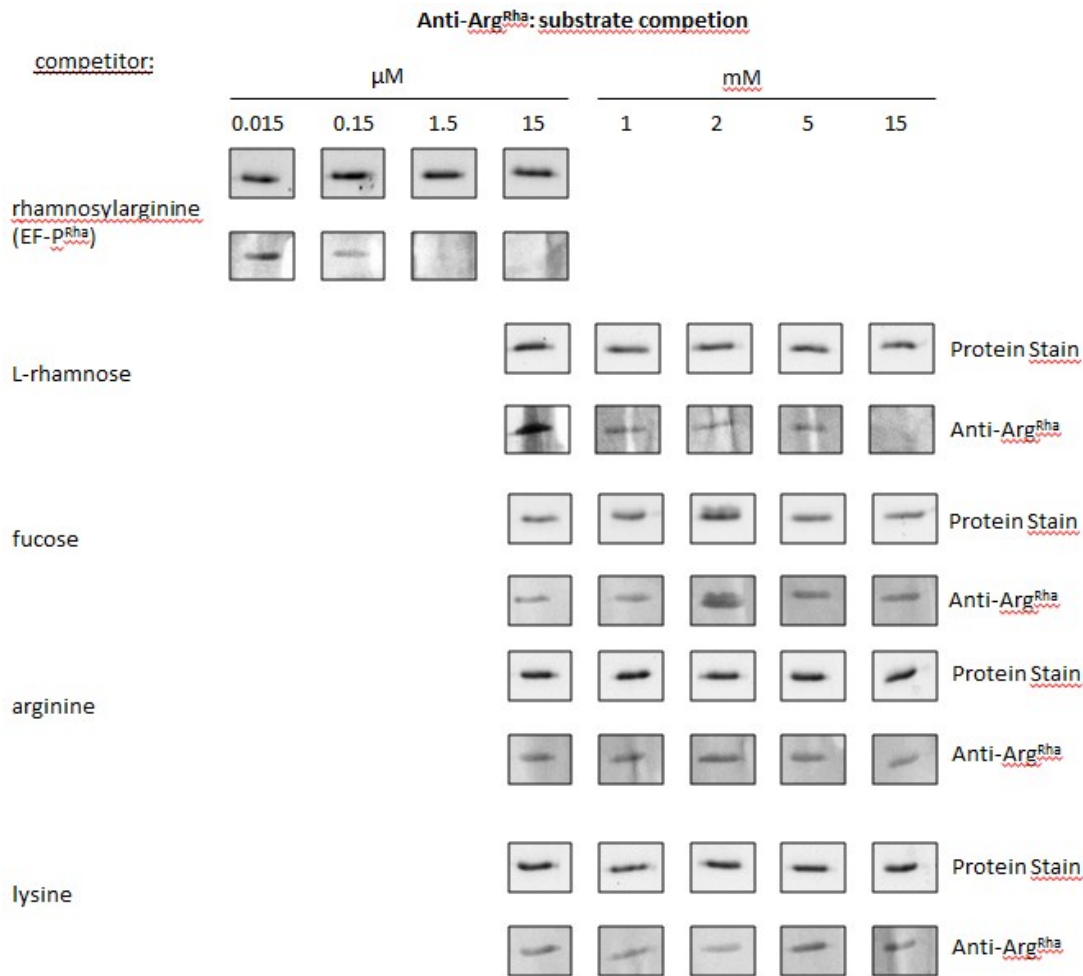
Fig S7 Detection of purified EF-P with Anti-Arg^{Rha}. Cross-reactivity analysis of Anti-Arg^{Rha} against L-rhamnose, L-fucose, L-arginine or L-lysine. 0.5 µg of purified EF-P^{Rha} were subjected to SDS-PAGE and subsequent Western Blot analysis using 0.2 µg/ml Anti-Arg^{Rha}. Anti-Arg^{Rha} was incubated with varying concentrations of EF-P^{Rha}, L-rhamnose, L-fucose, L-arginine and L-lysine. Buffer only served as control.

6. NMR spectrum

¹H-NMR of compound 1 (DMSO-*d₆*)
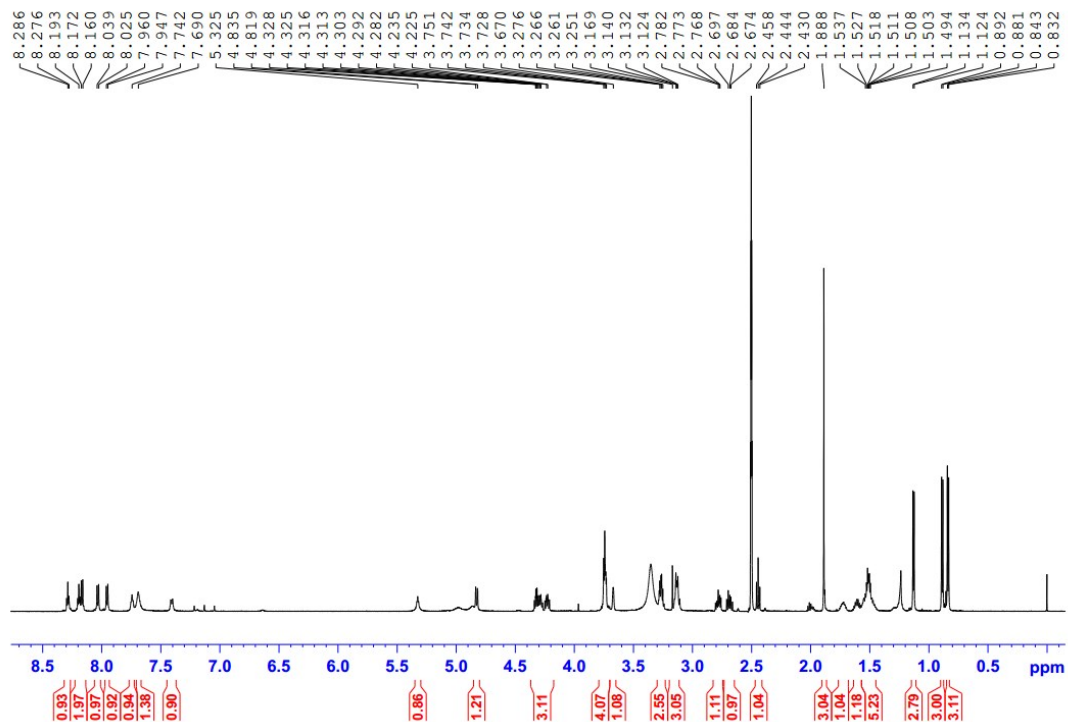
¹³C-NMR of compound 1 (DMSO-*d₆*)

HR-Q-TOF-MS of compound 1

Fragmentor Voltage	Collision Energy	Ionization Mode
150	0	ESI

+ESI Scan (0.099-0.198 min, 13 Scans) Frag=150.0V LX-PEPPHA-1.d

693.8246 (M+H)+

x10 5

Counts vs. Mass-to-Charge (m/z)
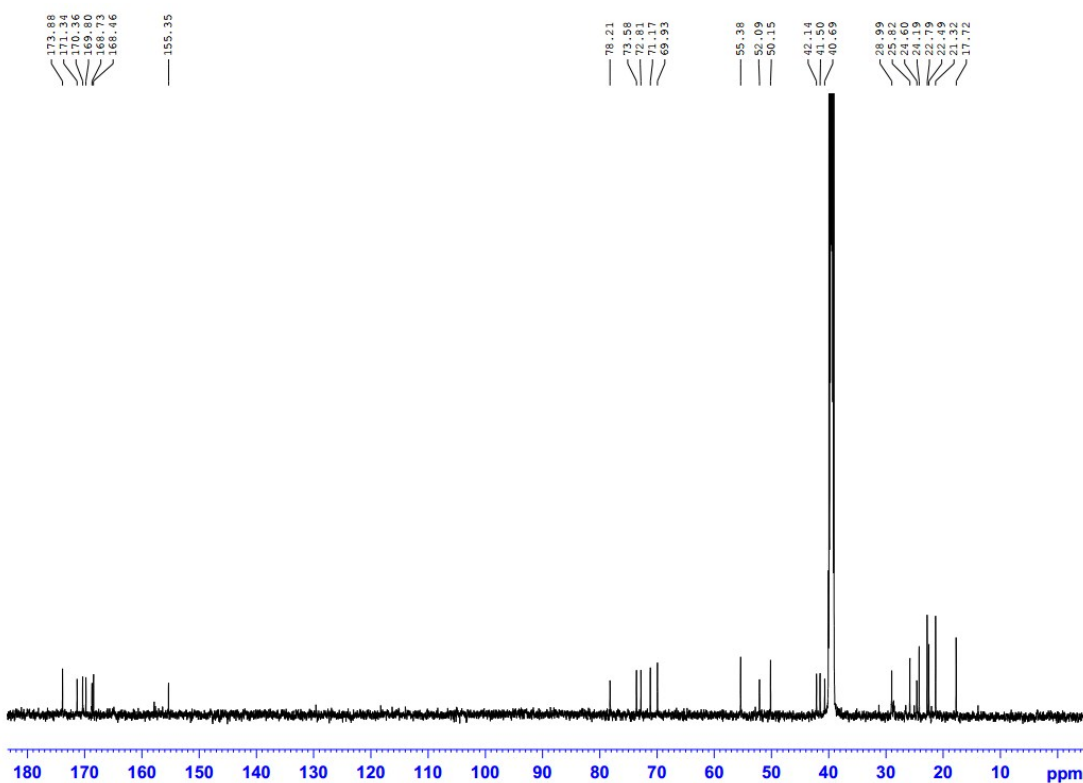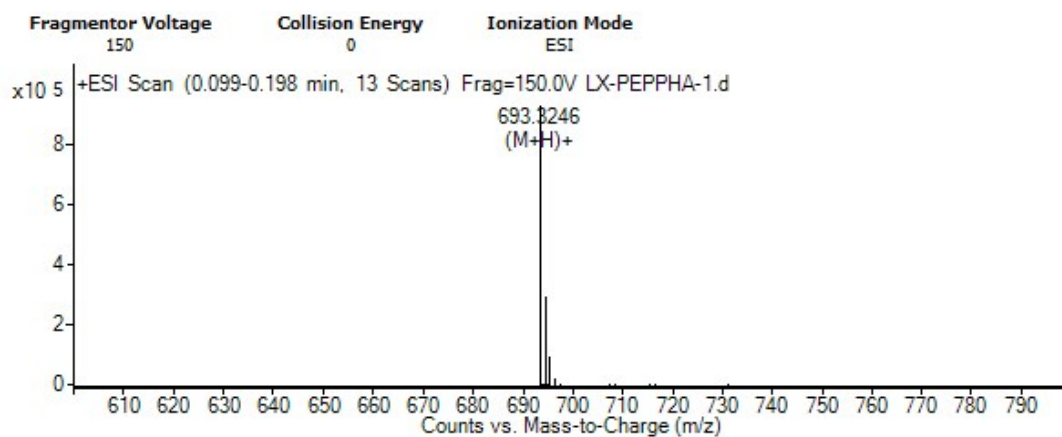
¹H-NMR of compound 3 (CDCl₃)
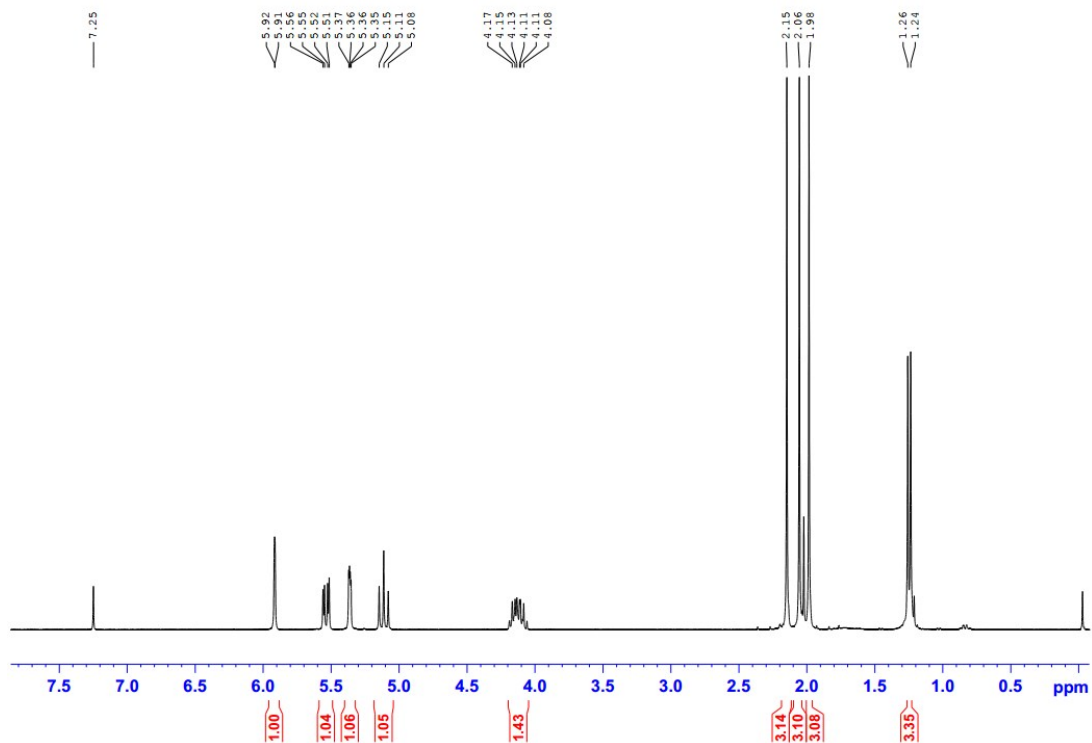
^{13}C-NMR of compound 3 (CDCl$_3$)
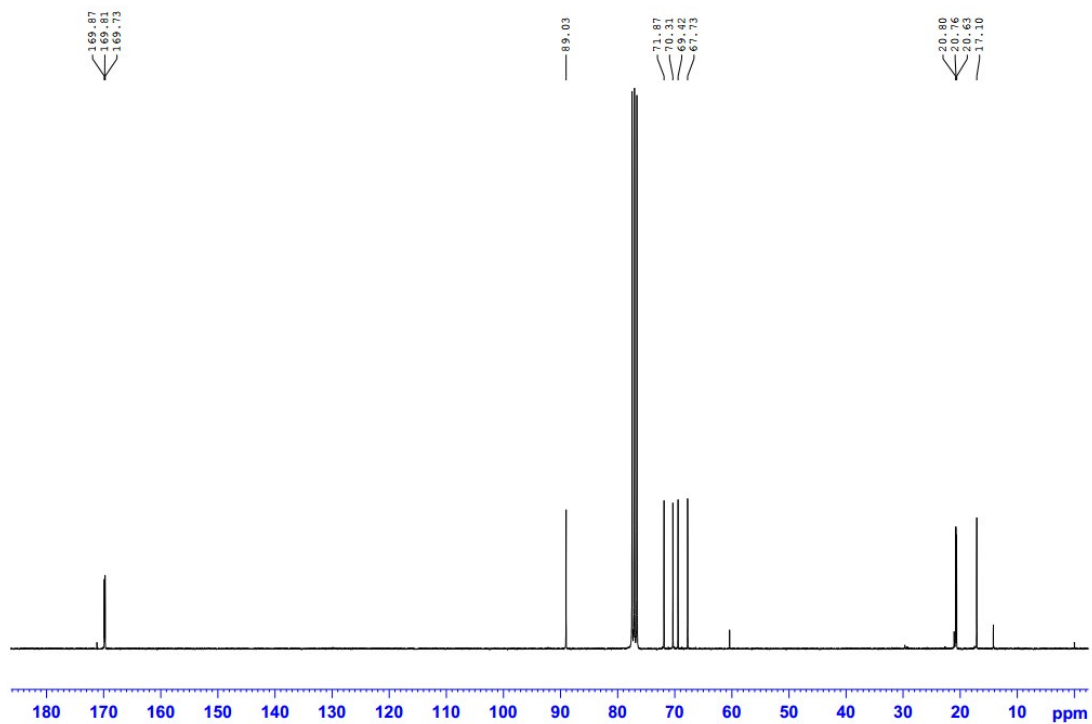
^{1}H-NMR of compound 4 (CDCl$_3$)

^{13}C-NMR of compound 4 (CDCl$_3$)

169.84, 169.81, 143.16, 82.80, 70.17, 69.88, 69.65, 68.33, 20.79, 20.76, 20.62, 17.27

HR-Q-TOF-MS of compound 4

Fragmentor Voltage 120   Collision Energy 0   Ionization Mode ESI

x10 4   + Scan (0.078-0.200 min, 16 Scans) SL-2.d

349.1067 (M+NH4)+

354.0620 (M+Na)+

Counts vs. Mass-to-Charge (m/z)

^{1}H-NMR of compound 5 (CDCl$_3$)

7.72
6.86
5.45 5.35 5.32 5.29 5.08 5.05 5.02
3.91 3.89 3.86
2.14 2.04 1.99
1.22 1.20

8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 ppm

0.94 2.15 2.92 1.29 1.00 3.17 2.88 3.00 3.05

^{13}C-NMR of compound 5 (CDCl$_3$)

184.38
170.13 169.98
80.91
70.51 68.67 67.80 65.86
20.85 20.82 20.75 17.37

190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 ppm

HR-Q-TOF-MS of compound 5

Fragmentor Voltage
120

Collision Energy
0

Ionization Mode
ESI

x10 5 + Scan (0.219 min) SL-3.d

349.1064
(M+H)+

371.0887
(M+Na)+

Counts vs. Mass-to-Charge (m/z)

^{1}H-NMR of compound 6 (CDCl$_3$)

^{13}C-NMR of compound 6 (CDCl$_3$)

190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 ppm

HR-Q-TOF-MS of compound 6

| Fragmentor Voltage | Collision Energy | Ionization Mode |
| 120 | 0 | ESI |

x10 6 + Scan (0.513 min) SL-4 .d

477.1909
(M+H)+

499.1726

455 460 465 470 475 480 485 490 495 500 505 510 515
Counts vs. Mass-to-Charge (m/z)
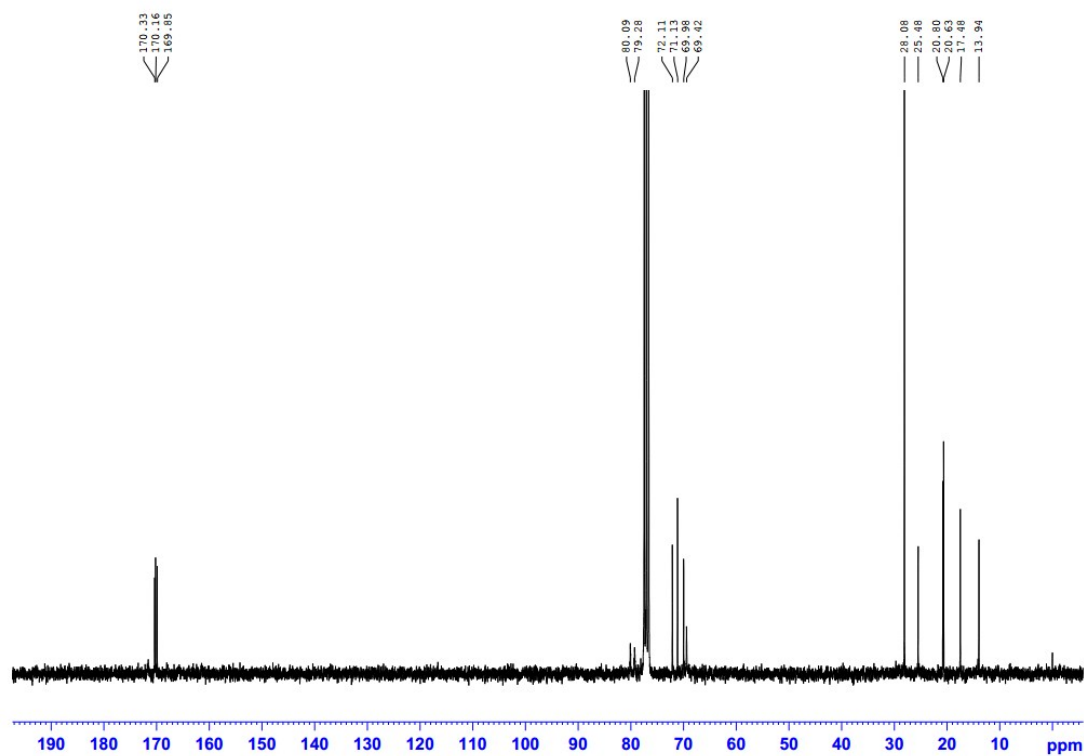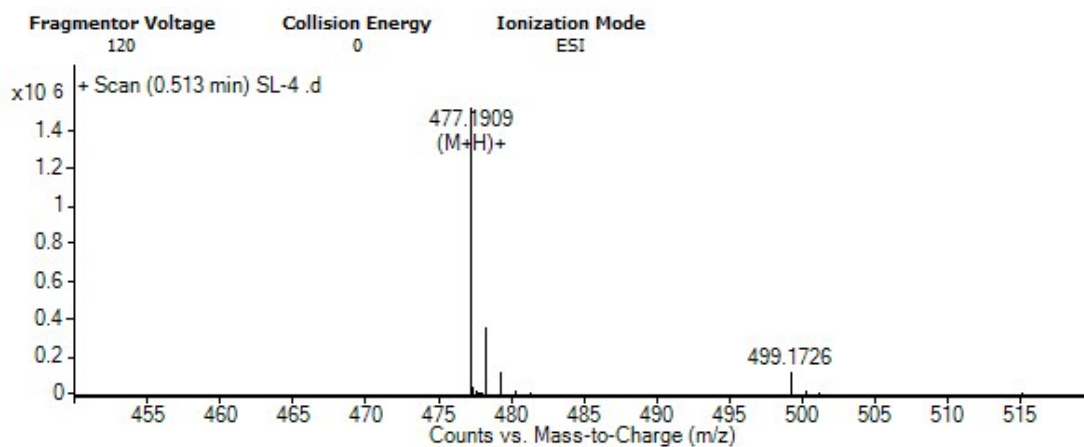
Table S2 Crystal data and structure refinement for compound **5**.

Identification code	a51230b
Empirical formula	$C_{13}H_{20}N_2O_7S$
Formula weight	348.37
Temperature	298(2) K
Wavelength	0.71073 Å
Crystal system	Orthorhombic
Space group	P 21 21 21
Unit cell dimensions	a = 8.984(5) Å $\alpha = 90°$.
	b = 10.648(6) Å $\beta = 90°$.
	c = 18.894(10) Å $\gamma = 90°$.
Volume	1807.5(17) Å³
Z	4
Density (calculated)	1.280 Mg/m³
Absorption coefficient	0.213 mm⁻¹
F (000)	736
Crystal size	0.660 x 0.200 x 0.150 mm
Theta range for data collection	2.156 to 26.979 °
Index ranges	-10<=h<=11, -13<=k<=13, -18<=l<=23
Reflections collected	8825
Independent reflections	3845 [R (int) = 0.0627]
Completeness to theta = 25.242°	99.7 %
Absorption correction	Semi-empirical from equivalents
Max. and min. transmission	1.000 and - 0.009
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	3845 / 8 / 224
Goodness-of-fit on F^2	0.841
Final R indices [I>2 sigma(I)]	R1 = 0.0410, wR2 = 0.0809
R indices (all data)	R1 = 0.0624, wR2 = 0.0860
Absolute structure parameter	0.03(6)
Extinction coefficient	n/a
Largest diff. peak and hole	0.146 and -0.193 e.Å⁻³

7. References

[1]   M. Sattler, J. Schleucher, C. Griesinger, *Prog. Nucl. Magn. Reson. Spectrosc.***1999**, 34, 93-158.

[2]   M. Sattler, M. Maurer, J. Schleucher, C. Griesinger, *J. Biomol. NMR* **1995**, *5*, 97-102.

[3]   L. Kay, P. Keifer, T. Saarinen, *J. Am. Chem. Soc.* **1992**, *114*, 10663-10665.

[4]   F. Schilling, L. R. Warner, N. I. Gershenzon, T. E. Skinner, M. Sattler, S. J. Glaser, *Angew. Chem.* **2014***, 126,* 4564–4568; *Angew. Chem.Int. Ed.* **2014**, *53*, 4475-4479.

[5]   F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277-293.

[6]   G. M. Sheldrick, *Acta. Crystallogr. A.* **2008**, *64*, 112-122.

[7]   H. Inoue, H. Nojima, H. Okayama, *Gene* **1990**, *96*, 23-28.

[8]   A. L. Starosta, J. Lassak, L. Peil, G. C. Atkinson, C. J. Woolstenhulme, K. Virumae, A. Buskirk, T. Tenson, J. Remme, K. Jung, D. N. Wilson, *Cell Rep.* **2014**, *9*, 476-483.

[9]   M. M. Bradford, *Anal. Biochem.* **1976**, *72*, 248-254.

[10]   U. K. Laemmli, *Nature* **1970**, *227*, 680-685.

[11]   C. L. Ladner, J. Yang, R. J. Turner, R. A. Edwards, *Anal. Biochem.* **2004**, *326*, 13-20.

[12]   J. Lassak, A. L. Henche, L. Binnenkade, K. M. Thormann, *Appl. Environ. Microbiol.* **2010**, *76*, 3263-3271.

[13]   J. Lassak, E. C. Keilhauer, M. Furst, K. Wuichet, J. Godeke, A. L. Starosta, J. M. Chen, L. Sogaard-Andersen, J. Rohr, D. N. Wilson, S. Haussler, M. Mann, K. Jung, *Nat. Chem. Biol.* **2015**, *11*, 266-270.

[14]   D. R. Macinga, M. M. Parojcic, P. N. Rather, *J. Bacteriol.* **1995**, *177*, 3407-3413.

[15]   F. W. Studier, B. A. Moffatt, *J. Mol. Biol.* **1986**, *189*, 113-130.

[16]   K. Venkateswaran, D. P. Moser, M. E. Dollhopf, D. P. Lies, D. A. Saffarini, B. J. MacGregor, D. B. Ringelberg, D. C. White, M. Nishijima, H. Sano, J. Burghardt, E. Stackebrandt, K. H. Nealson, *Int. J. Syst. Evol. Microbiol.* **1999**, *49 Pt 2*, 705-724.

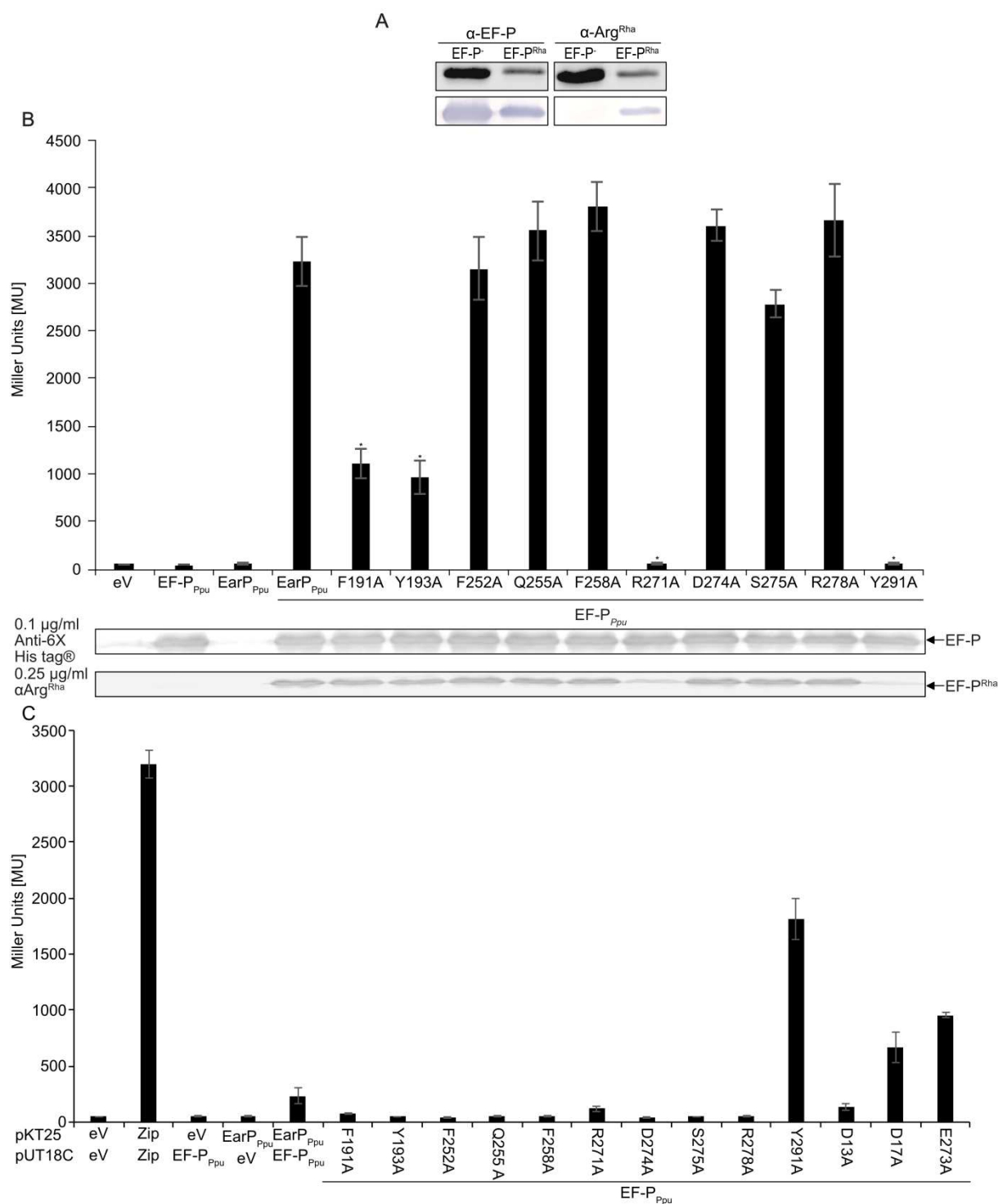**Supplemental information – Chapter 3**

Supplemental figures

Figure S1. (A) Specificity of anti-Arg^{Rha} antibodies. One point five micrograms of unmodified EF-P (EF-P$^-$) and 0.5 µg modified EF-P (EF-P^{Rha}) were subjected to SDS-PAGE and transferred to a nitrocellulose membrane by horizontal Western blotting. The membrane was cut, and the halves were used to detect EF-P$_{Ppu}$ using 0.2 µg/ml anti-EF-P and EF-P^{Rha} using 0.25 µg/ml anti-Arg^{Rha} antibodies. (B) *In vivo* functionality analysis of EarP$_{Ppu}$ single-amino-acid exchange variants. (Top) β-Galactosidase activity in *E. coli* MG1655 P$_{cadBA}$::*lacZ* Δ*efp* upon expression of EF-P$_{Ppu}$ and EarP$_{Ppu}$ and coexpression of EF-P$_{Ppu}$ with EarP$_{Ppu}$ and single-amino-acid substitution variants. Cells were incubated under *cadBA*-inducing conditions (LB, pH 5.8) at 30°C o/n ([17]). Data are mean values from three independent replicates ± standard deviations. Asterisks indicate significance. (Middle) The presence of heterologously expressed His$_6$-tagged EF-P$_{Ppu}$ was verified using 0.1 µg/ml anti-His$_6$ tag (Abcam, Inc.). (Bottom) Rhamnosylation of EF-P$_{Ppu}$ in MG1655 cells coexpressing EF-P$_{Ppu}$ and its EarP$_{Ppu}$ single-amino-acid exchange variants was assessed by Western blotting and detection of EF-P^{Rha} using 0.25 µg/ml anti-Arg^{Rha}.
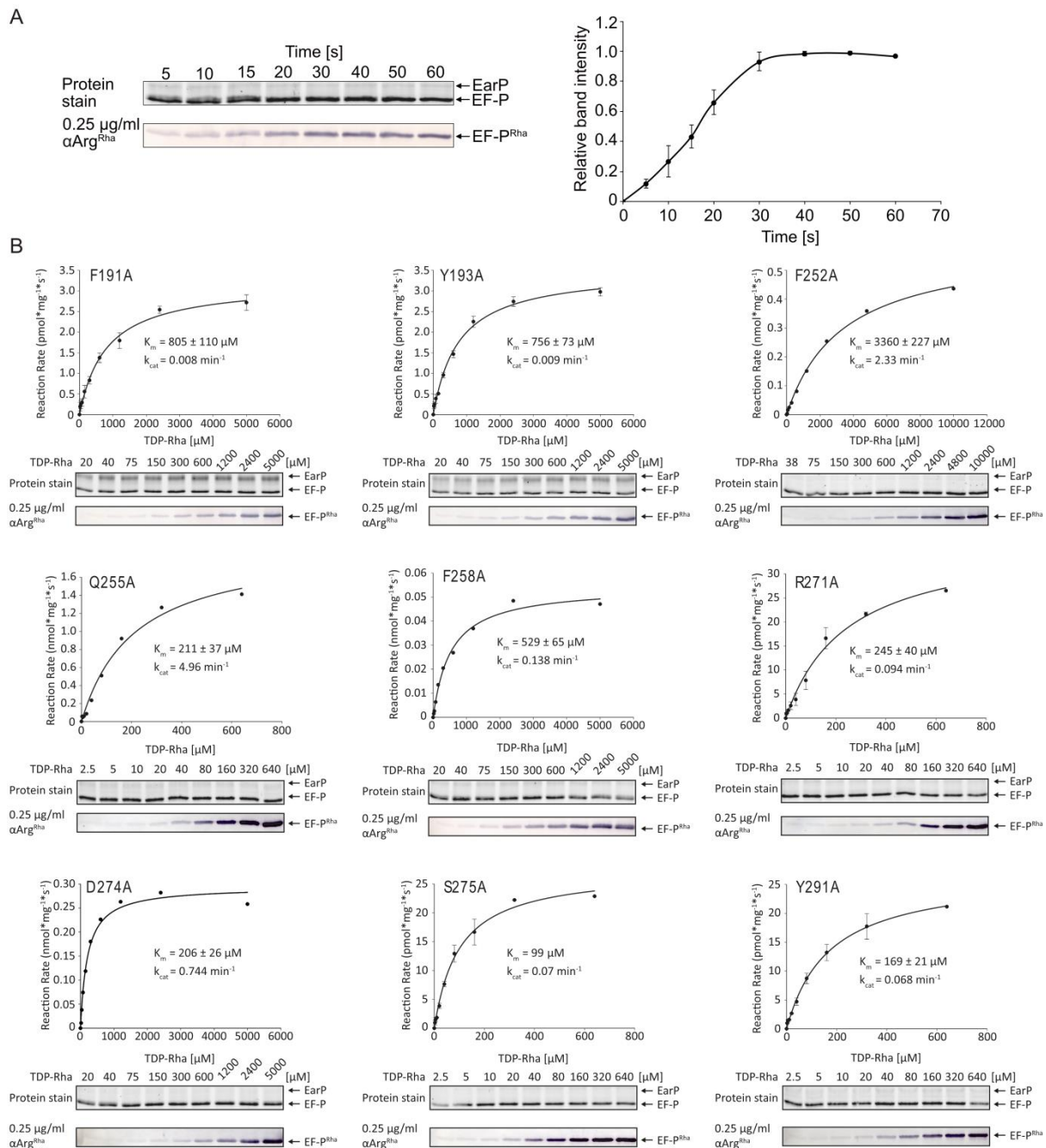
Figure S2. (A) Time-dependent *in vitro* rhamnosylation of EF-P$_{Ppu}$ by EarP$_{Ppu}$. Fixed amounts of EF-P$_{Ppu}$ (2.5 µM), EarP$_{Ppu}$ (0.1 µM), and TDP-Rha (500 µM) were incubated at 30°C in 100 mM NaP$_i$, pH 7.6, for various amounts of time. Measurements were performed in technical duplicates. Standard errors are shown. A representative SDS-PAGE gel and Western blot used for generation of the time course curve are shown. Rhamnosylated EF-P$_{Ppu}$ (EF-P^{Rha}) was detected using 0.25 µg/ml anti-Arg^{Rha}. Band intensities were quantified using ImageJ ([76]). (B) TDP-Rha saturation curves of EarP$_{Ppu}$ single-amino-acid exchange variants. Fixed amounts of EF-P$_{Ppu}$ (2.5 µM) and the respective EarP$_{Ppu}$ single-amino-acid exchange variants (0.1 µM) were incubated with various concentrations of TDP-Rha at 30°C in 100 mM NaP$_i$, pH 7.6. In the cases of the F191A^{EarP} and Y1931A^{EarP} variants, 0.5 µM concentrations were used. Suitable incubation times were determined by time course experiments prior to determination of kinetic parameters (data not shown). Representative SDS-PAGE gels and Western blots used for generation of the TDP-Rha saturation curves are depicted underneath the corresponding graphs. Rhamnosylated EF-P$_{Ppu}$ was detected after Western blotting using 0.25 µg/ml anti-Arg^{Rha}. Band intensities were quantified using ImageJ ([76]). K_m and k_{cat} were calculated using SigmaPlot. Standard errors are shown when measurements were performed in technical duplicates
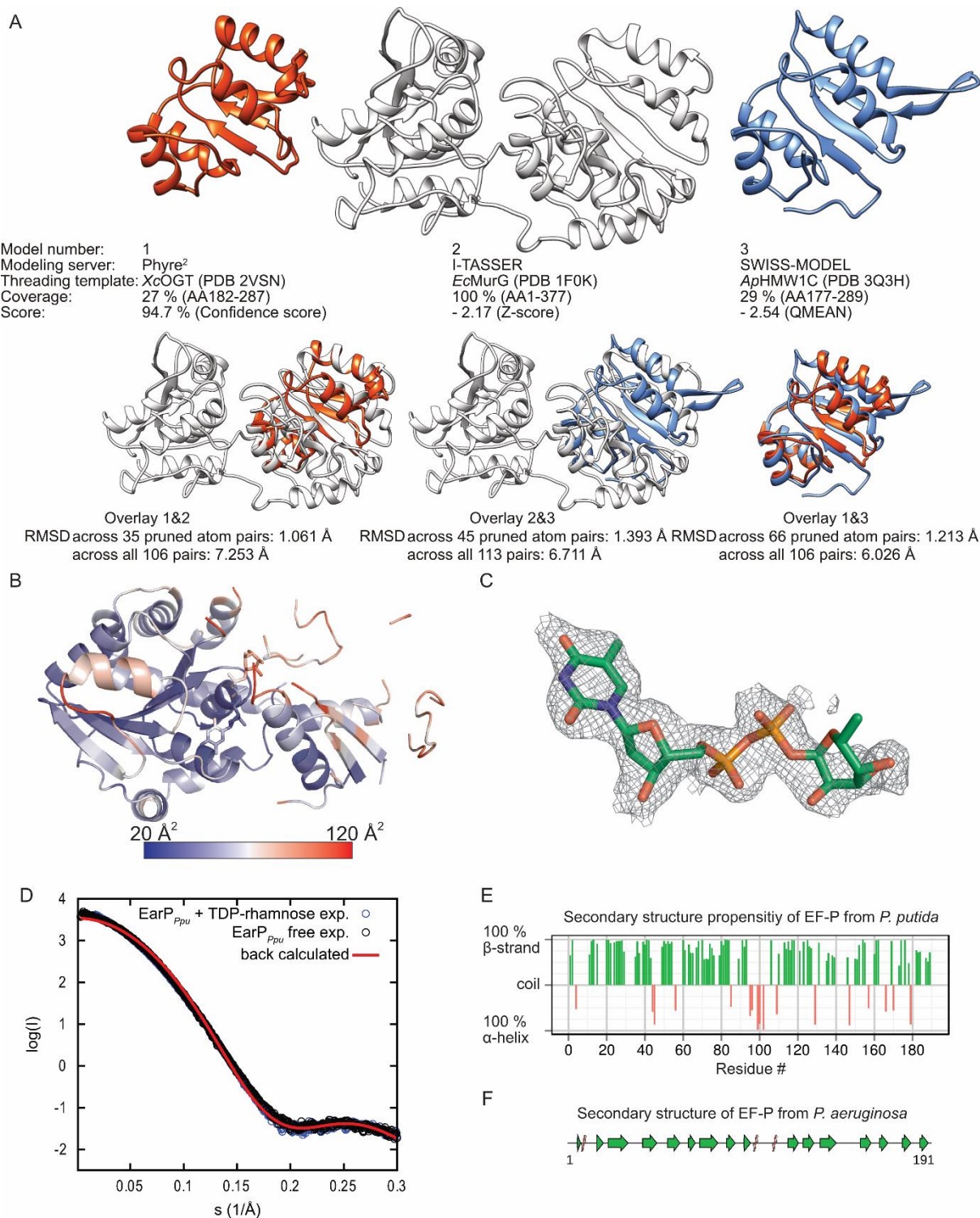
Figure S3. (A) Fold recognition models of EarP from *P. putida*. (Top) Structural models of EarP from *P. putida* were generated using Phyre[2] (orange), the I-TASSER server (white), and SWISS-MODEL (blue). Model coverage and scoring is shown underneath the respective model structures. (Bottom) Overlay of structural models from Phyre[2] and I-TASSER (left), I-TASSER and SWISS-MODEL (middle), and SWISS-MODEL and Phyre[2] (right). Root mean square deviations of atomic positions are shown underneath the respective structural overlays. All illustrations and overlays were generated using UCSF Chimera and the UCSF Chimera MatchMaker ([82]). (B) B-factors are plotted on the crystal structure of EarP. Several regions in the N-domain show significantly higher B-factors than in the C domain, suggesting the mobile nature of this region. (C) Electron density of TDP-Rha in the donor binding pocket. TDP-Rha is depicted as sticks. Sigma-Aldrich weighted 2mFo-DFc map of TDP-Rha contoured at 1 σ is depicted as grey mesh. (D) SAXS analysis of free EarP and EarP bound to TDP-Rha. The two experimental small-angle X-ray scattering curves of free EarP (black) and EarP plus TDP-Rha (blue) do not exhibit any significant changes. Thus, it can be concluded that there are no major conformational changes upon ligand binding. The crystal structure and monomeric states of the protein are also validated in solution, as the back-calculated scattering densities (red line) fit well with the experimental data (χ^2 = 1.96). The deviation at the

low Q range is due to aggregation and/or interparticle interaction. (E) Secondary structure of EF-P from *P. putida*. The secondary structures of individual amino acids are indicated as a propensity to form either α-helix (red) or β-strand (green). The amino acids with a propensity to adopt a random coil or lacking information about secondary structure were assigned zero values in the plot. The propensity values were obtained from C_α, C_β, and C' secondary chemical shifts using CcpNmr Analysis ([106]). (F) Secondary structure of EF-P from *P. aeruginosa* (PDB accession number 3OYY) ([45]). Red ribbons indicate α-helices; green arrows indicate β-strands.
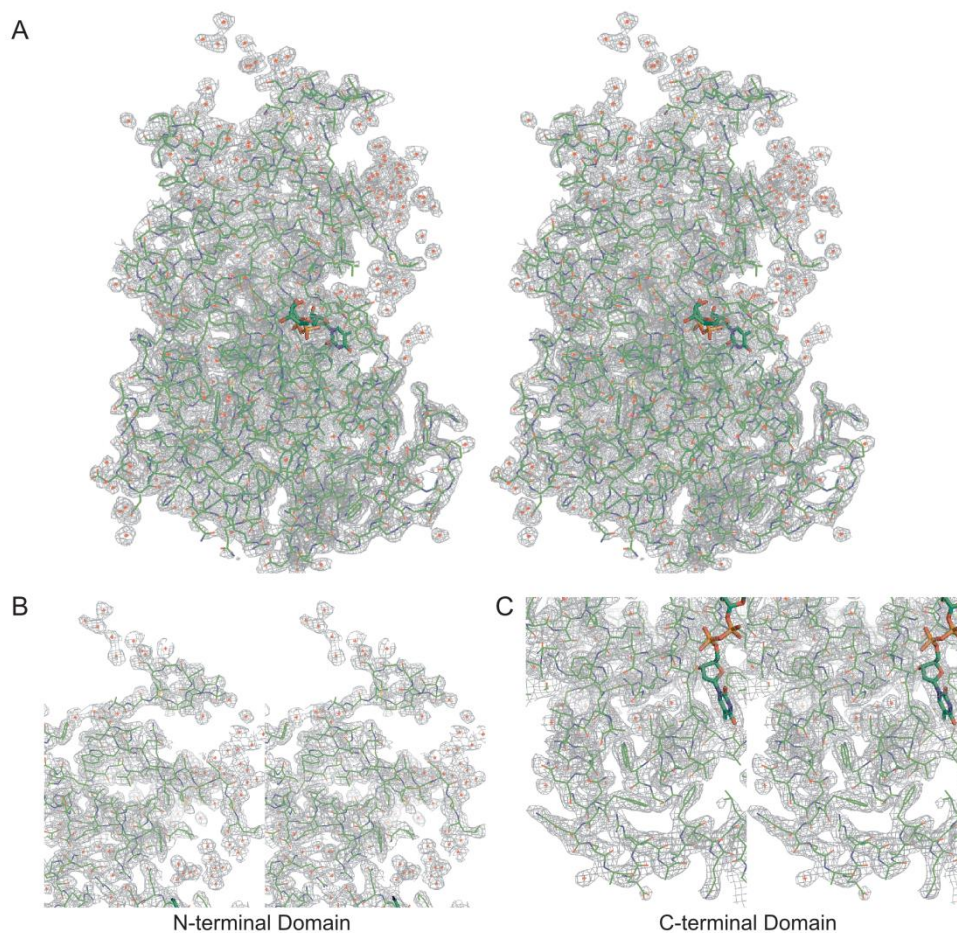
Figure S4. (A) Stereo view of the EarP–TDP-Rha complex. Electron density is shown. (B and C) Electron densities for the N- and C-terminal domains, respectively. The N-terminal domain shows several disconnected blobs of electron density currently modeled with water, and some regions are entirely missing.
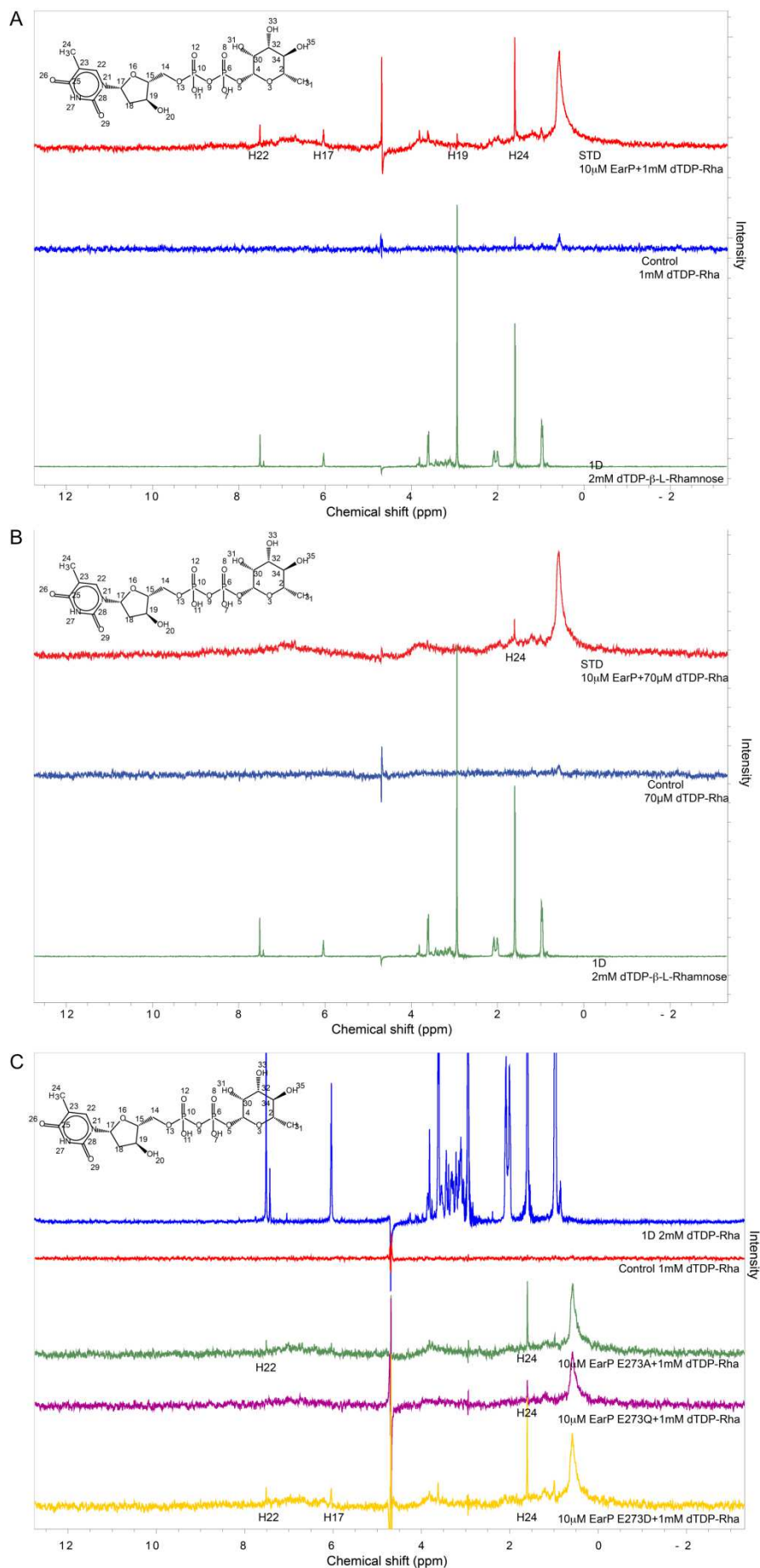
Figure S5. (A) STD NMR at a 1:100 ratio of protein to ligand shows several signals from the TDP but no signal from the sugar region of the TDP-Rha, thus confirming the crystal structure where the sugar is solvent exposed, and does not show extensive contacts with the protein. (B) STD NMR with a low excess of TDP-Rha (1:7 ratio of protein to ligand) clearly shows the signal for the methyl group of the thymidine ring, confirming the interaction between TDP-Rha and EarP under conditions used for SAXS data collection. (C) STD NMR experiments with E273 variants showing that the mutation of E273 to alanine, glutamine, or aspartate does not affect the binding of TDP-Rha to EarP.
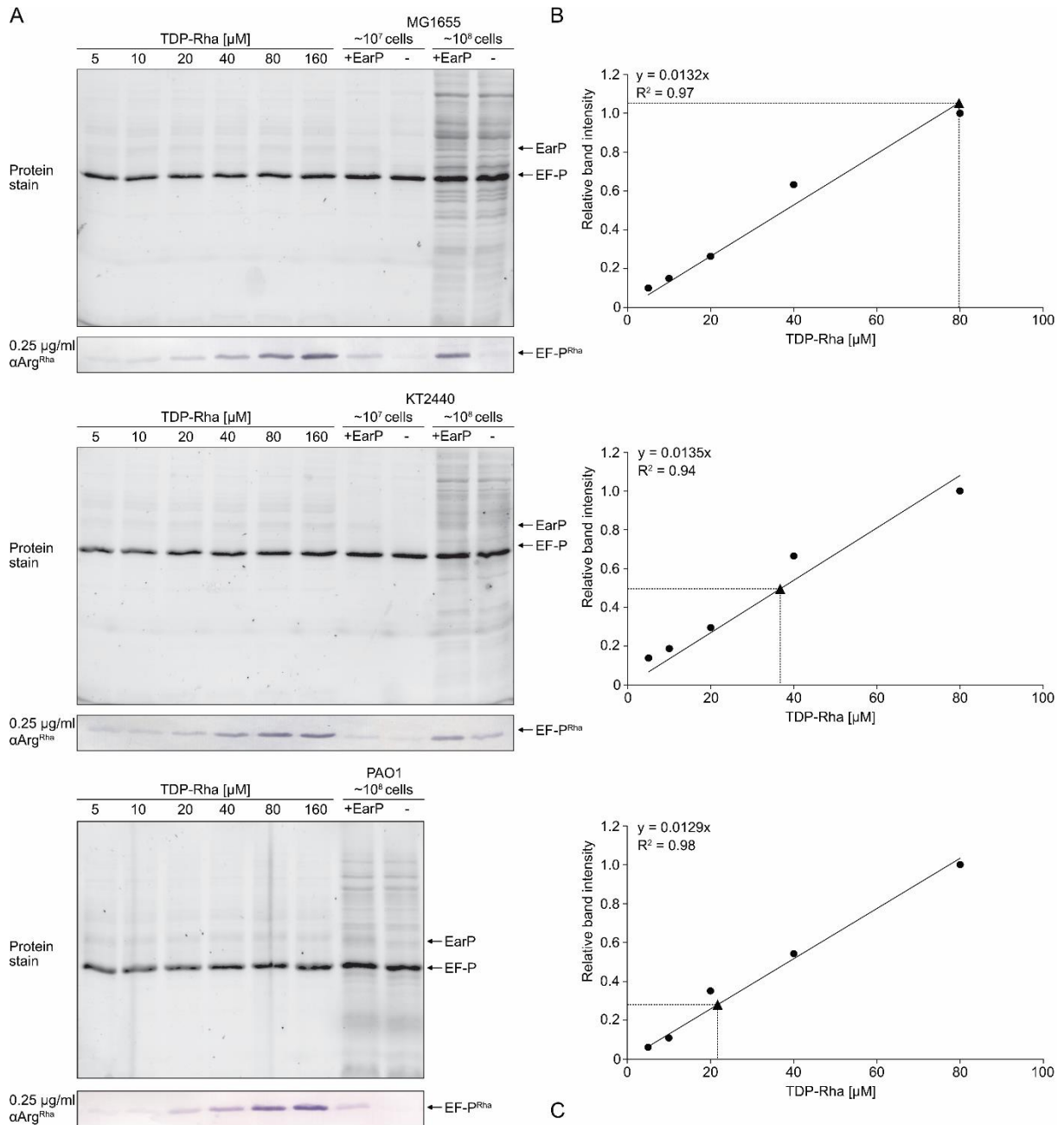
Figure S6. (A) SDS-PAGE gels and Western blots for determination of the intracellular TDP-Rha concentration in *E. coli* MG1655 (top), *P. putida* KT2440 (middle), and *P. aeruginosa* (bottom). Purified EF-P$_{Ppu}$ (2.5 µM) and EarP$_{Ppu}$ (0.1 µM) were incubated with increasing concentrations of TDP-Rha or lysates from ~10^7 and ~10^8 cells (sample volume, 20 µl; 100 mM NaP$_i$ [pH 7.6]; 20 s; 30°C). Samples containing cell lysates were incubated in the presence and absence of EarP$_{Ppu}$. Rhamnosylated EF-P (EF-P^{Rha}) was detected using 0.25 µg/ml anti-Arg^{Rha}. For *E. coli* and *P. aeruginosa*, experiments were performed as biological triplicates. (B) Representative calibration curves calculated from relative Western blot band intensities in samples containing increasing TDP-Rha concentrations. Band intensities were quantified using ImageJ ([76]). TDP-Rha concentrations in samples containing ~10^8 lysed cells (triangles) were calculated from the slope of the calibration curve and are indicated by dashed lines. Band intensities from samples containing no EarP$_{Ppu}$ were subtracted from those containing EarP$_{Ppu}$ to correct for background signal. (C) Equations used to calculate average TDP-Rha concentrations in single cells. For calculation of average TDP-Rha concentrations in *E. coli* cells, an average cell volume of 3.9 fl was assumed ([84]). For calculation of average TDP-Rha concentrations in *P. putida* and *P. aeruginosa* cells, an average cell volume of 2.1 fl was assumed ([85]).
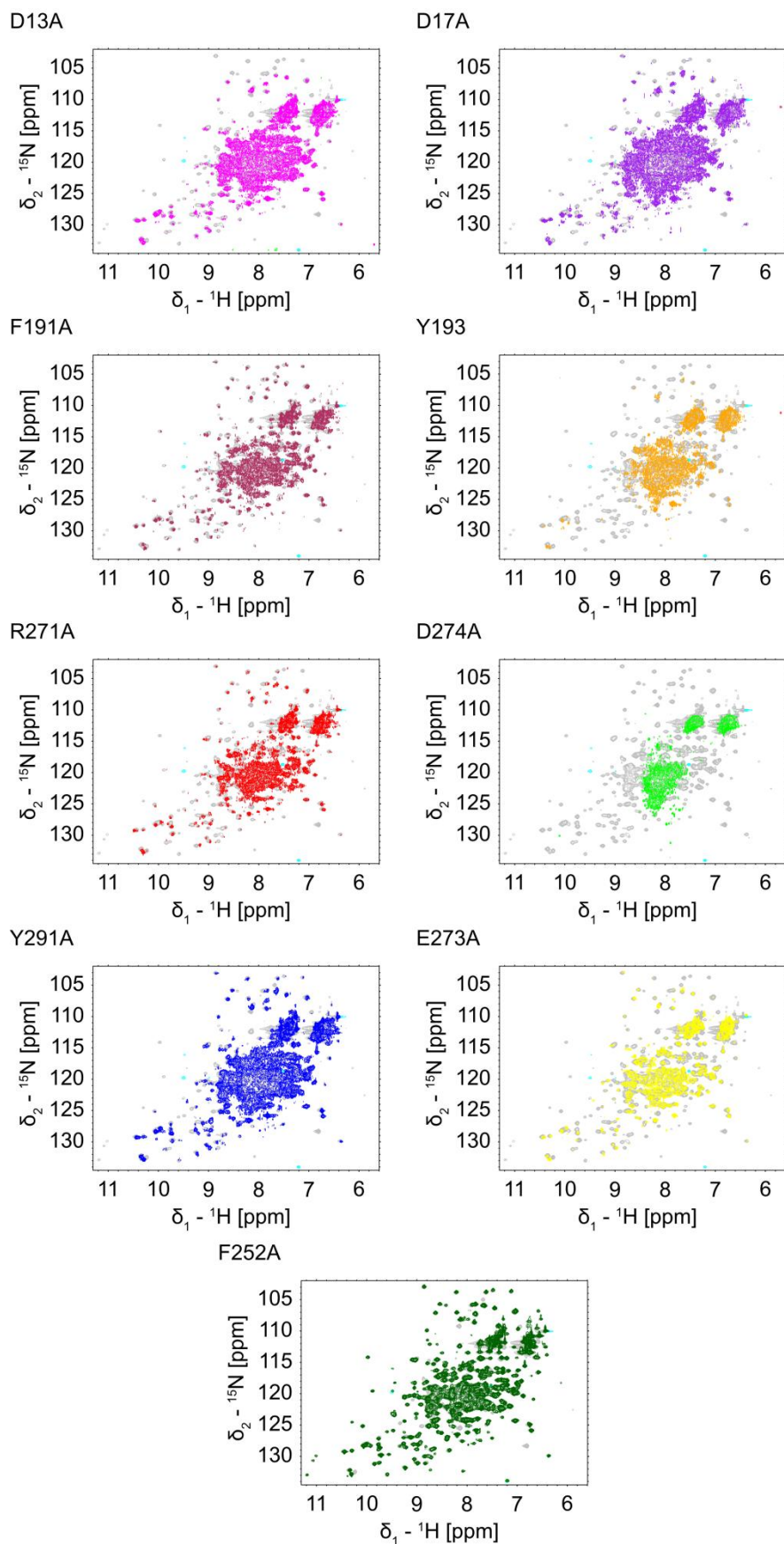
Figure S7. Assessing folding state of EarP$_{Ppu}$ variants. To check whether a given variant is properly folded, a ^{1}H-^{15}N HSQC of ^{15}N-labeled variants was recorded (colored) and overlaid with a spectrum of wild-type protein (grey). Note that the mutant spectra were recorded with a small number of scans. Spectra of all variants except the D274A variant overlap well with the wild-type spectrum and show well-dispersed peaks (an indicator of a stable secondary structure), confirming that the single-amino-acid substitutions do not influence proper folding of EarP. In contrast, the D274A variant is in large part unfolded.

Supplemental information – Chapter 4

Supplemental figures
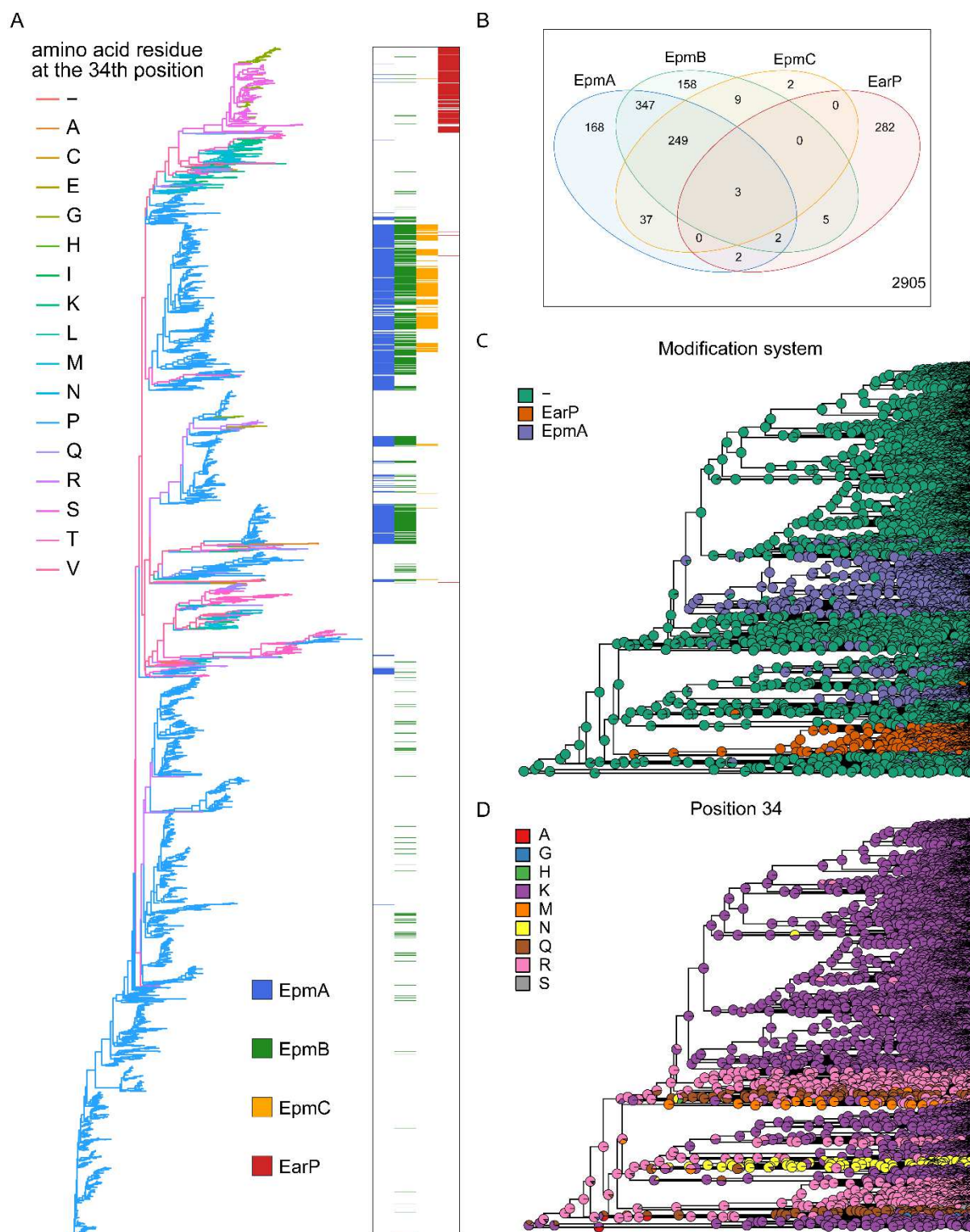
Figure S1, related to Figure 1 & 2 (A) Phylogenetic tree built from the EF-P KOW-like N-domain I. The tree is colored according to the 34th position in the multiple sequence alignment and annotated according to the co-occurrence with EpmABC and EarP. (B) Venn diagram of modification proteins that co-occur with EF-P/IF5A. (C) (D) Phylogenetic trees of the EF-P KOW-like N-domain I, with reconstructed state of the modification system (C) and the 34th position (D).
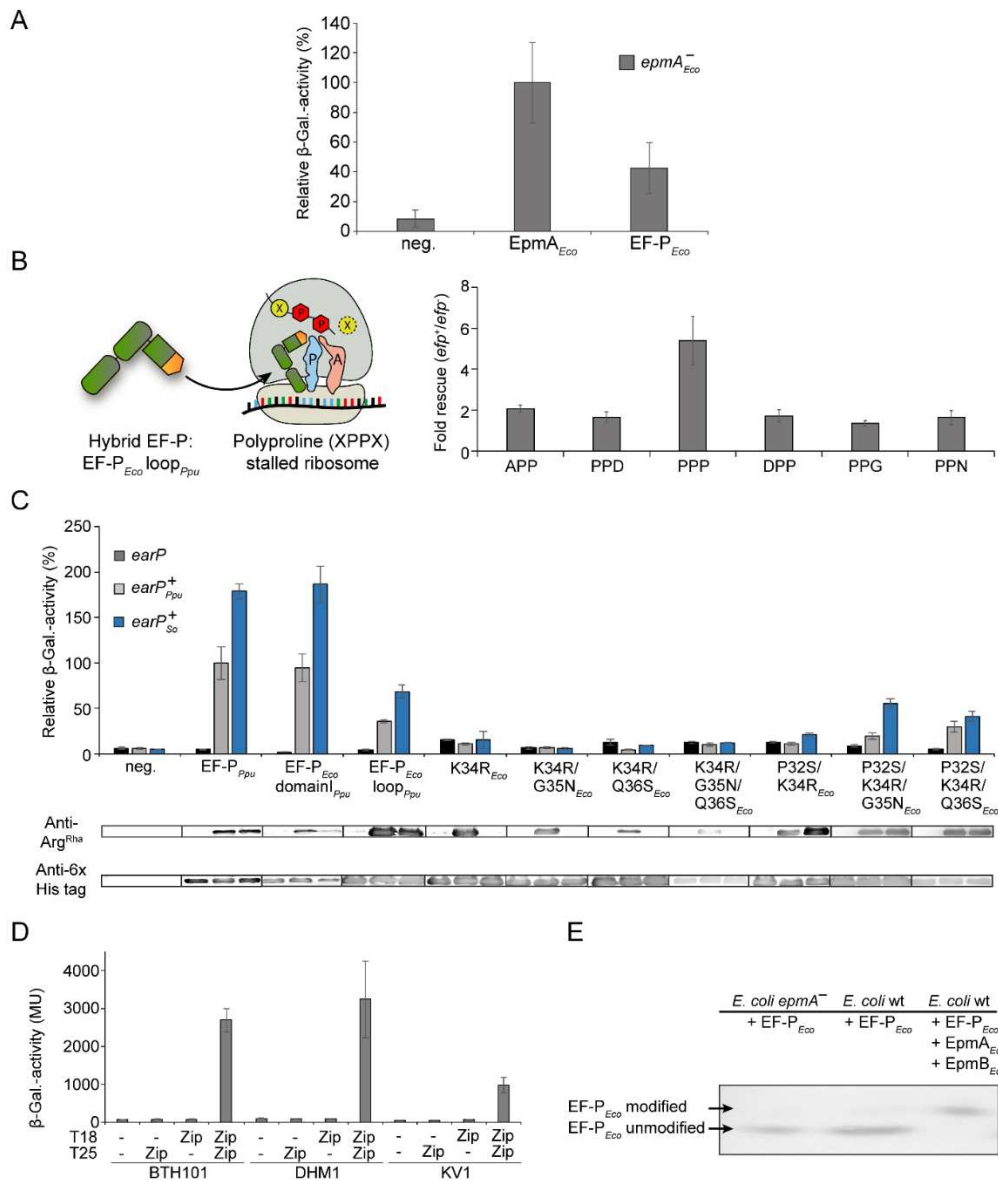
Figure S2, related to Figure 3 & 4 (A) β-galactosidase activity in the ΔepmA reporter strain (*E. coli* MG-CL-12-yjeA) harboring either a plasmid borne copy of *epmA* (pBAD33-*epmA*) or *efp* (pBAD24-*efp*$_{Eco}$) (B) Effect of the hybrid EF-P$_{Eco}$ loop$_{Ppu}$ on different polyproline containing stalling motifs. Measurements were performed in *E. coli* Δ*efp* cells (JW4107), harboring plasmid encoded different stalling motifs followed by the *lacZ* reporter (pBBR1MCS-3 XPPX *lacZ*). (C) *In vivo* rhamnosylation and functionality analysis of hybrid EF-Ps coexpressed with EarP from *P. putida* (light grey) or EarP from *Shewanella oneidensis* MR-1 (blue). The measurements were performed in the reporter strain were *efp* was deleted (*E. coli* MG-CR-efp-KanS). β-galactosidase activity is given in Miller units and reflects the degree of EF-P functionality. Production and rhamnosylation of EF-P was verified by Western blot analysis. (D) Single colonies of *E. coli* BTH101, DHM1 and KV1 cotransformed with variants of pUT18C (T18) and pKT25 (T25) were inoculated in 2 ml of LB medium supplemented with 50 µg/ml kanamycin sulfate and 100 µg/ml carbenicillin and grown at 30 °C under agitation over night. On the next day the cells were harvested by centrifugation, and the β-galactosidase activities were determined. (E) Isoelectric focusing of *E. coli* EF-P, either overproduced in *E. coli* epmA⁻ (BW25113-epmA), in which *epmA* is chromosomally deleted, or in *E. coli* wild type cells (BW25113). Furthermore, EF-P was overproduced in combination with its PTM proteins EpmA and EpmB in *E. coli* wild type (BW25113). Production of EF-P was verified by Western blot analysis.
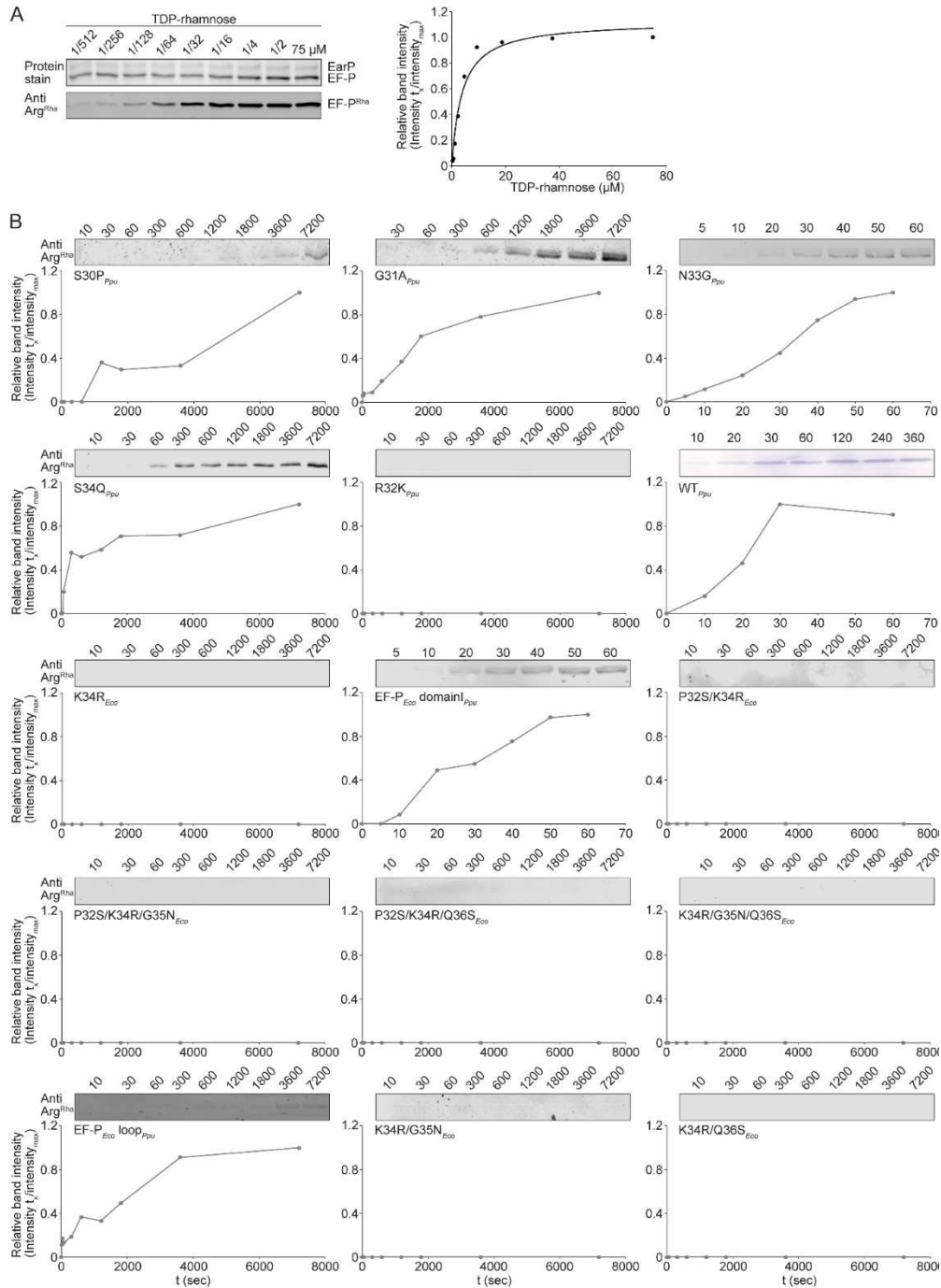
Figure S3, related to Figure 3 & 4 (A) Left: Analysis of EarP$_{Ppu}$ kinetic parameters. 0.5 µg of EF-P$_{Ppu}$ and 0.05 µg of EarP$_{Ppu}$ were subjected to SDS-PAGE after *in vitro* rhamnosylation (see methods) for 20 seconds at varying TDP-rhamnose concentrations. Proteins were transferred to nitrocellulose membrane by horizontal Western blotting. Rhamnosylated EF-P$_{Ppu}$ was detected using 0.25 µg/ml of Anti-Arg[Rha]. Right: TDP-rhamnose saturation curve of EarP$_{Ppu}$. Band intensities on nitrocellulose membrane were quantified using ImageJ and relative band intensities were plotted against TDP-rhamnose concentration. (B) Timecourse analysis of various EF-P$_{Ppu}$ and EF-P$_{Eco}$ variants. 0.5 µg of EF-P$_{Ppu}$ and 0.05 µg of EarP$_{Ppu}$ were subjected to SDS-PAGE after *in vitro* rhamnosylation at a TDP-rhamnose concentration of 50 µM for varying timespans. Band intensities on nitrocellulose membrane were quantified using ImageJ and relative band intensities were plotted against time.
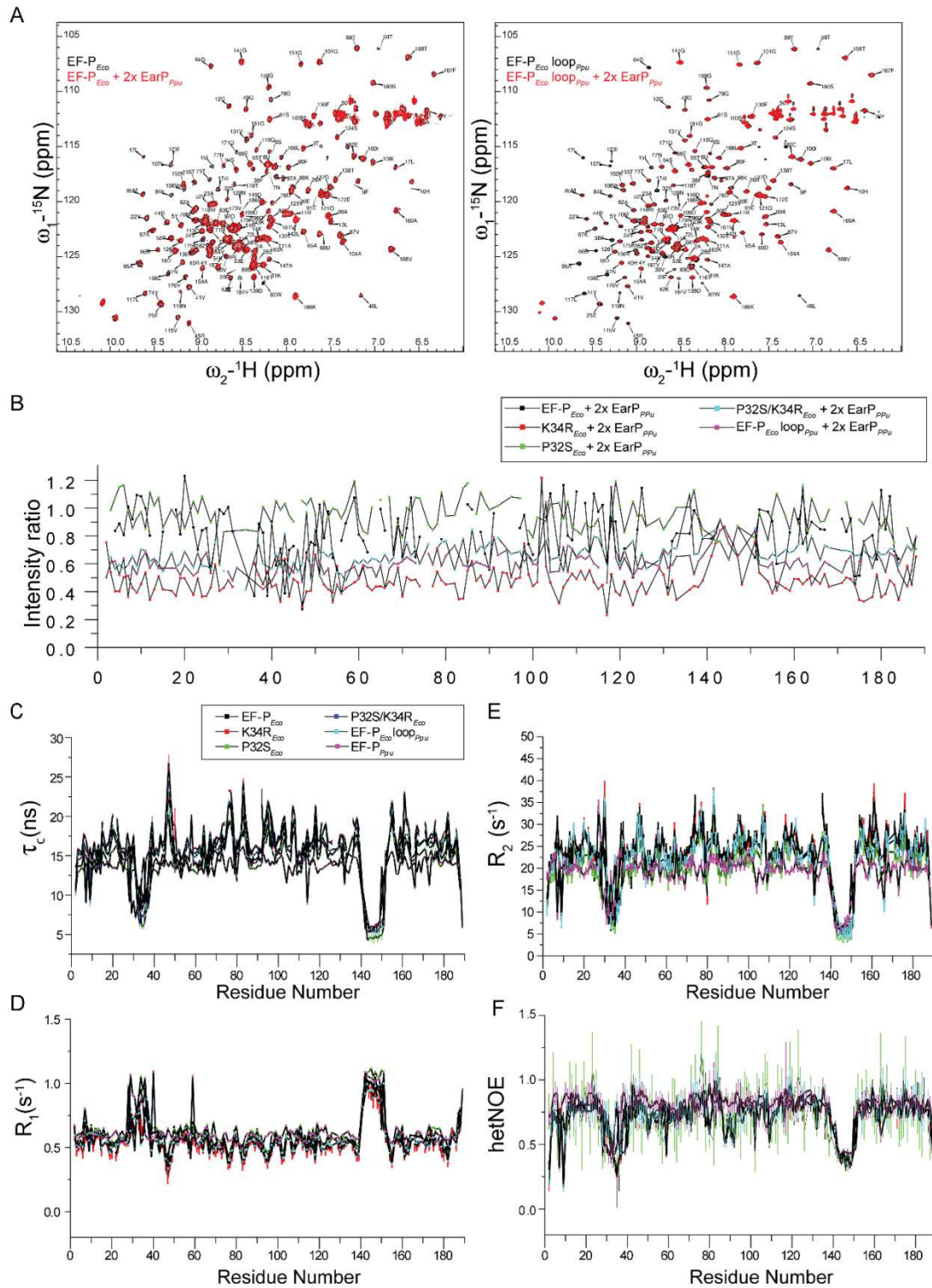
137

Figure S4, related to Figure 4 (A) ^{1}H-^{15}N HSQC titrations of EF-P$_{Eco}$ and EF-P$_{Eco}$ loop$_{Ppu}$ with EarP$_{Ppu}$ is shown along with the backbone assignments. Both proteins show a decrease in intensity upon titration with EarP$_{Ppu}$ indicating their interaction with EarP$_{Ppu}$. (B) Intensity ratio of all the EF-P$_{Eco}$ mutants on titration with EarP$_{Ppu}$ is shown. (C) The correlation time (τ_c) for EF-P$_{Eco}$ and its variants along with (D) R1, (E) R2 rates and (F) hetNOE are shown.

138

Supplemental tables

Table S1: Distribution of EF-P, EpmA, EpmB, EpmC and EarP homologs
Supplemental online material

Table S2: Strains used in this study

Strain	Genotype	Reference
DH5αλ*pir*	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 ΔlacZYA-argF* U169 φ80d*lacZ*ΔM15 λpir	(1)
LMG194	F⁻ *ΔlacX74 galE galK thi rpsL ΔphoA* (PvuII) *Δara714 leu*::Tn*10*	(2)
BL21(DE3)	F⁻ *ompT gal dcm lon hsdS_B*(rB⁻ mB⁻) λ(DE3)	(3)
DHM1	F⁻ *cya-854 recA1 endA1 gyrA96* (*NalR*) *thi1 hsdR17 spoT1 rfbD1 glnV44(AS)*	(4)
BTH101	F⁻ *cya-99 araD139 galE15 galK16 rpsL1* (*StrR*) *hsdR2 mcrA1 mcrB1* additional *relA1* mutation reported by (Battesti and Bouveret, 2012)	Euromedex
JW4107	Δ*(araD-araB)567*, Δ*lacZ4787*(::rrnB-3), λ⁻, *rph-1*, Δ*(rhaD-rhaB)568*, Δ*efp-772::kan, hsdR514*	(5)
KV1	MG1655 *rpsL150 ΔcyaA* P_*lac*::*luxCDABE-lacZ*	This study
LF1	MG1655 *rpsL150* P_*lac*::*rpsL-neo-kan::lacZ*^Δ1–100 *bp;Kan^R Strp^S*	(6)
MG-CR-*efp*	MG1655 Δ*lacZ*::*tet rpsL150 efp::npt ΔcadBA* P_*cadBA*::*lacZ*	(7)
MG-CR-*efp*-KanS	MG1655 Δ*lacZ*::*tet rpsL150 Δefp ΔcadBA* P_*cadBA*::*lacZ*	This study
MG-CR-*efp*-epmA-KanR	MG1655 Δ*lacZ*::*tet rpsL150 Δefp ΔcadBA epmA::npt* P_*cadBA*::*lacZ*	This study
MG-CL-12-*yjeA*	MG1655 Δ*lacZ*::*tet rpsL150 yjeA::npt ΔcadBA cadBA::lacZ*	(8)
BW25113	Δ*(araD-araB)567*, Δ*lacZ4787*(::rrnB-3), λ-, rph-1, Δ(*rhaD-rhaB)568*, hsdR514	(9)
BW25113-*epmA*	Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, rph-1, Δ(rhaD-rhaB)568, hsdR514, ΔepmA	This study
JW4116	F⁻, Δ*(araD-araB)567*, Δ*lacZ4787*(::rrnB-3), λ⁻, *rph-1*, Δ*(rhaD-rhaB)568*, Δ*poxA782::kan, hsdR514*	(5)

Kan^R: kanamycin resistance, Strep^S: streptomycin sensitive

139

Table S3: Plasmids used in this study

Plasmid	Features	Reference
Plasmids for strain construction		
pRED/ET® Amp	λ-RED recombinase in pBAD24; Amp^R	GeneBridges, Germany
FRT-PGK-gb2-neo-FRT template DNA	PCR-template (plasmid DNA) for generating a FRT-flanked PGK-gb2-neo cassette, Kan^R	GeneBridges, Germany
709-FLPe, amp	Plasmid for removal of FRT flanked resistance cassette, Supplier ID: A106	GeneBridges, Germany
pBAD/HisA-Lux	Contains the *luxCDABE* operon from *Photorhabdus luminescens*	(10)
Plasmids for mutational analysis of the *E. coli* loop and for overproduction		
pBAD24	Amp^R-cassette, pBBR322 origin, *araC* coding sequence, *ara* operator	(2)
pBAD24-*efp_Eco*	C-terminal His$_6$-tagged *E. coli efp* amplified from pBAD33-*efp_E.c.*-His6 (Lassak et al., 2015) using P1/P2	this study
pBAD24-*efp_Eco* P32G	C-terminal His$_6$-tagged *E. coli efp* substitution variant P32G. Overlap PCR fragment was amplified from pBAD24-*efp_Eco* using P1/P16 and P2/P15	this study
pBAD24-*efp_Eco* P32S	C-terminal His$_6$-tagged *E. coli efp* substitution variant P32S. Overlap PCR fragment was amplified from pBAD24-*efp_Eco* using P1/P18 and P2/P17	this study
pBAD24-*efp_Eco* G33A	C-terminal His$_6$-tagged *E. coli efp* substitution variant G33A. Overlap PCR fragment was amplified from pBAD24-*efp_Eco* using P1/P20 and P2/P19	this study
pBAD24-*efp_Eco* G33S	C-terminal His$_6$-tagged *E. coli efp* substitution variant G33S. Overlap PCR fragment was amplified from pBAD24-*efp_Eco* using P1/P22 and P2/P21	this study
pBAD24-*efp_Eco* K34A	C-terminal His$_6$-tagged *E. coli efp* substitution variant K34A. Amplified from pBAD33-*efp_E.c.*-His6-K34A (Lassak et al., 2015) using P1/P2	this study
pBAD24-*efp_Eco* K34M	C-terminal His$_6$-tagged *E. coli efp* substitution variant K34M. Overlap PCR fragment was amplified from pBAD24-efpEco using P1/P24 and P2/P23	this study
pBAD24-*efp_Eco* K34N	C-terminal His$_6$-tagged *E. coli efp* substitution variant K34N. Overlap PCR fragment was amplified from pBAD24-*efp_Eco* using P1/P26 and P2/P25	this study
pBAD24-*efp_Eco* K34Q	C-terminal His$_6$-tagged *E. coli efp* substitution variant K34Q. Overlap PCR fragment was amplified from pBAD24-*efp_Eco* using P1/P28 and P2/P27	this study
pBAD24-*efp_Eco* K34R	C-terminal His$_6$-tagged *E. coli efp* substitution variant K34R. Amplified from pBAD33-*efp_E.c.*-His6-K34R (Lassak et al., 2015) using P1/P2	this study
pBAD24-*efp_Eco* G35N	C-terminal His$_6$-tagged *E. coli efp* substitution variant G35N. Overlap PCR fragment was amplified from pBAD24-*efp_Eco* using P1/P30 and P2/P29	this study
pBAD24-*efp_Eco* Q36S	C-terminal His$_6$-tagged *E. coli efp* substitution variant Q36S. Overlap PCR fragment was amplified from pBAD24-*efp_Eco* using P1/P32 and P2/P31	this study

Plasmids for mutational analysis of the *P. putida* loop and for overproduction		
pBAD24-*efp*_{*Ppu*}	C-terminal His$_6$-tagged *P. putida efp* amplified from *Pseudomonas putida* KT2440 using P1/P2	this study
pBAD24-*efp*_{*Ppu*} K29A	C-terminal His$_6$-tagged *efp P. putida* substitution variant K29A. Overlap PCR fragment was amplified from pBAD24-*efp*_{*Ppu*} using P1/P36 and P2/P35	this study
pBAD24-*efp*_{*Ppu*} K29R	C-terminal His$_6$-tagged *P. putida efp* substitution variant K29R. Overlap PCR fragment was amplified from pBAD24-*efp*_{*Ppu*} using P1/P37 and P2/P35	this study
pBAD24-*efp*_{*Ppu*} S30A	C-terminal His$_6$-tagged *P. putida efp* substitution variant S30A. Overlap PCR fragment was amplified from pBAD24-*efp*_{*Ppu*} using P1/P38 and P2/P35	this study
pBAD24-*efp*_{*Ppu*} S30G	C-terminal His$_6$-tagged *P. putida efp* substitution variant S30G. Overlap PCR fragment was amplified from pBAD24-*efp*_{*Ppu*} using P1/P39 and P2/P35	this study
pBAD24-*efp*_{*Ppu*} S30P	C-terminal His$_6$-tagged *P. putida efp* substitution variant S30P. Overlap PCR fragment was amplified from pBAD24-*efp*_{*Ppu*} using P1/P40 and P2/P35	this study
pBAD24-*efp*_{*Ppu*} G31A	C-terminal His$_6$-tagged *P. putida efp* substitution variant G31A. Overlap PCR fragment was amplified from pBAD24-*efp*_{*Ppu*} using P1/P41 and P2/P35	this study
pBAD24-*efp*_{*Ppu*} G31S	C-terminal His$_6$-tagged *P. putida efp* substitution variant G31S. Overlap PCR fragment was amplified from pBAD24-*efp*_{*Ppu*} using P1/P42 and P2/P35	this study
pBAD24-*efp*_{*Ppu*} R32K	C-terminal His$_6$-tagged *P. putida efp* substitution variant R32K. Overlap PCR fragment was amplified from pBAD24-*efp*_{*Ppu*} using P1/P44 and P2/P43	this study
pBAD24-*efp*_{*Ppu*} N33D	C-terminal His$_6$-tagged *P. putida efp* substitution variant N33D. Overlap PCR fragment was amplified from pBAD24-*efp*_{*Ppu*} using P1/P46 and P2/P45	this study
pBAD24-*efp*_{*Ppu*} N33G	C-terminal His$_6$-tagged *P. putida efp* substitution variant N33G. Overlap PCR fragment was amplified from pBAD24-*efp*_{*Ppu*} using P1/P47 and P2/P45	this study
pBAD24-*efp*_{*Ppu*} S34A	C-terminal His$_6$-tagged *P. putida efp* substitution variant S34A. Overlap PCR fragment was amplified from pBAD24-*efp*_{*Ppu*} using P1/P48 and P2/P45	this study
pBAD24-*efp*_{*Ppu*} S34Q	C-terminal His$_6$-tagged *P. putida efp* substitution variant S34Q. Overlap PCR fragment was amplified from pBAD24-*efp*_{*Ppu*} using P1/P49 and P2/P45	this study
pBAD24-*efp*_{*Ppu*} A35S	C-terminal His$_6$-tagged *P. putida efp* substitution variant A35S. Overlap PCR fragment was amplified from pBAD24-*efp*_{*Ppu*} using P1/P50 and P2/P45	this study
pBAD33	Cm^R-cassette, p15A origin, *araC* coding sequence, *ara* operator	(2)
pBAD33 PP1857-His6	C-terminal His6-Tag *earP* version from *P. putida* KT2440	(11)

Plasmids for cross modification/actication and overproduction		
pBAD24-*efp_Eco* domainI-*efp_Ppu*	C-terminal His$_6$-tagged *E. coli efp* where the first 65 amino acids (domainI) were substituted by the first 65 amino acids from *P. putida efp*. Overlap PCR fragment was amplified from pBAD24-*efp_Ppu* using P1/P52 and pBAD24-*efp_Eco* P2/P51	this study
pBAD24-*efp_Eco* loop-*efp_Ppu*	C-terminal His$_6$-tagged *E. coli efp* P32S K34R G35N Q36S multiple amino acid substitution variant, corresponding to EF-P of *E. coli* carrying the acceptor loop of EF-P from *P. putida.* Overlap PCR fragment was amplified from pBAD24-*efp_Eco* using P1/P54 and P2/P53	this study
pBAD24-*efp_Eco* P32S K34R	C-terminal His$_6$-tagged *E. coli efp* substitution variant P32S K34R. Overlap PCR fragment was amplified from pBAD24-*efp_Eco* K34R using P1/P56 and P2/P55	this study
pBAD24-*efp_Eco* K34R G35N	C-terminal His$_6$-tagged *E. coli efp* substitution variant K34R G35N. Overlap PCR fragment was amplified from pBAD24-*efp_Eco* K34R using P1/P58 and P2/P57	this study
pBAD24-*efp_Eco* K34R Q36S	C-terminal His$_6$-tagged *E. coli efp* substitution variant K34R Q36S. Overlap PCR fragment was amplified from pBAD24-*efp_Eco* K34R using P1/P60 and P2/P59	this study
pBAD24-*efp_Eco* P32S K34R G35N	C-terminal His$_6$-tagged *E. coli efp* substitution variant P32S K34R G35N. Overlap PCR fragment was amplified from pBAD24-*efp_Eco* P32S K34R using P1/P62 and P2/P61	this study
pBAD24 -*efp_Eco* P32S K34R Q36S	C-terminal His$_6$-tagged *E. coli efp* substitution variant P32S K34R Q36S. Overlap PCR fragment was amplified from pBAD24-*efp_Eco* P32S K34R using P1/P64 and P2/P63	this study
pBAD24-*efp_Eco* K34R G35N Q36S	C-terminal His$_6$-tagged *E. coli efp* substitution variant K34R G35N Q36S. Overlap PCR fragment was amplified from pBAD24-*efp_Eco* K34R using P1/P66 and P2/P65	this study
pBAD33-*earP_So*	*earP* from *Shewanella oneidensis* MR-1 amplified from pBAD24-*earP_S.o.*-His$_6$ (Lassak et al., 2015) using P1/P67	this study
Plasmids for XPPX assay		
p3LC-TL30-APP	p3LC-TL30 + sequence encoding Ala-Pro-Pro	(12)
p3LC-TL30-PPD	p3LC-TL30 + sequence encoding Pro-Pro-Asp	(12)
p3LC-TL30-PPP	p3LC-TL30 + sequence encoding Pro-Pro-Pro	(Ude et al., 2013)
p3LC-TL30-DPP	p3LC-TL30 + sequence encoding Asp-Pro-Pro	(12)
p3LC-TL30-PPG	p3LC-TL30 + sequence encoding Pro-Pro-Gly	(12)
p3LC-TL30-PPN	p3LC-TL30 + sequence encoding Pro-Pro-Asn	(12)
Plasmids used in the reporter strain MG-CL-12-yjeA		
pBAD33-*epmA*	*epmA* from *E. coli*, amplified using P72/P75	this study

142

	Plasmids for protein overproduction (NMR)		
pET SUMO	pBR322 origin, *lacI*, T7 *lac* promoter, N-terminal His$_6$ tag, SUMO coding sequence, Kan^R, Supplier ID: K300-01	Invitrogen
pET SUMO-*efp$_{Eco}$*	C-terminal genetic fusion of *efp* from *E. coli* to His$_6$-SUMO-tag. Amplified from pBAD24-*efp$_{Eco}$* using P68/P69	this study
pET SUMO-*efp$_{Ppu}$*	C-terminal genetic fusion of *efp* from *P. putida* KT2440 to His$_6$-SUMO-tag	(11)
pET SUMO-*efp$_{Eco}$* P32S	C-terminal genetic fusion of *efp* from *E. coli* substitution variant P32S to His$_6$-SUMO-tag. Overlap PCR fragment was amplified from pET SUMO-*efp$_{Eco}$* using P5/P18 and P6/P17	this study
pET SUMO-*efp$_{Eco}$* K34R	C-terminal genetic fusion of *efp* from *E. coli* substitution variant K34R to His$_6$-SUMO-tag. Overlap PCR fragment was amplified from pET SUMO-*efp$_{Eco}$* using P5/P71 and P6/P70	this study
pET SUMO-*efp$_{Eco}$* P32S K34R	C-terminal genetic fusion of *efp* from *E. coli* substitution variant P32S K34R to His$_6$-SUMO-tag. Overlap PCR fragment was amplified from pET SUMO-*efp$_{Eco}$* using P5/P56 and P6/P55	this study
pET SUMO-*efp$_{Eco}$* loop$_{Ppu}$	C-terminal genetic fusion of *efp* from *E. coli* substitution variant P32S K34R G35N Q36S to His$_6$-SUMO-tag. Overlap PCR fragment was amplified from pET SUMO-*efp$_{Eco}$* using P5/P54 and P6/P53	this study
	Plasmids for bacterial two-hybrid (BTH)		
pUT18-*zip*	N-terminal genetic fusion of the leucine zipper from GCN4 to the T18 fragment of CyaA	Euromedex
pKT25-*zip*	C-terminal genetic fusion of the leucine zipper from GCN4 to the T25 fragment of CyaA	Euromedex
pUT18C-PP1858	C-terminal genetic fusion of *efp* from *Pseudomonas putida t* to the T18 fragment of CyaA	(11)
pKT25-PP1857	C-terminal genetic fusion of *earP* from *Pseudomonas putida* to the T25 fragment of CyaA	(11)
pUT18C-*efp$_{Eco}$*	C-terminal genetic fusion of *efp* from *Escherichia coli* to the T18 fragment of CyaA. PCR fragment was amplified from pBAD24-*efp$_{Eco}$* using P80/P81	this study
pUT18C-*efp$_{Eco}$* K34R	C-terminal genetic fusion of *efp* from *Escherichia coli* to the T18 fragment of CyaA. K34R single amino acid exchange variant. PCR fragment was amplified from pBAD24-*efp$_{Eco}$* K34R using P80/P81	this study
pUT18C-*efp$_{Eco}$* P32S K34R	C-terminal genetic fusion of *efp* from *Escherichia coli* to the T18 fragment of CyaA. P32S K34R double amino acid exchange variant. PCR fragment was amplified from pBAD24-*efp$_{Eco}$* P32S K34R using P80/P81	this study
pUT18C-*efp$_{Eco}$* loop$_{Ppu}$	C-terminal genetic fusion of *efp* from *E. coli* to the T18 fragment of CyaA. P32S K34R G35N Q36S multiple amino acid exchange variant, corresponding to EF-P of *E. coli* carrying the acceptor loop of EF-P from *P. putida*. PCR fragment was amplified from pBAD24-*efp$_{Eco}$* loop$_{Ppu}$ using P80/P81	this study

Plasmids for isoelectric focusing		
pBAD33-*efp*-His₆	C-terminal His₆-Tag *efp* version from *E. coli* into pBAD33	(7)
pBAD33-*efp*-His₆-*epmAB*	C-terminal His₆-tagged *efp* and *epmAB from E. coli.* Overlap PCR fragment was amplified using P73/P76, P72/P75 and P74/P77	this study

Amp^R, Cm^R, Kan^R: ampicillin, chloramphenicol, kanamycin resistance.

Table S4: Primers used in this study

Identifier	Oligonucleotide	Sequence (5' - 3')	Restriction site	Reference
Primers for sequencing and cloning				
P1	Seq33 fw	GGC GTC CAC ACT TTG CTA TGC		(13)
P2	pBAD HisA rev	CAG TTC CCT ACT CTC GCA TG		(13)
P3	*epmA* chk fw	TAG GTA CAA CAG TAT AGT CTG ATG GAT AA		this study
P4	*epmA* chk rev	TGA GGC ATG AAA CCA TCC TTC ATT TC		this study
P5	T7 Prom	TAA TAC GAC TCA CTA TAG G		
P6	T7 Term	TAT GCT AGT TAT TGC TCA G		
Primers for strain construction of KV1				
P7	LacI-583-fw	GTC TGC GTC TGG CTG GCT GGC ATA		(6)
P8	LuxC-OL-rev	TAG TGC CCA TAG CTG TTT CCT GTG TGA AAT TGT TAT CC		this study
P9	LuxC-OL-fw	GGA AAC AGC TAT GGG CAC TAA AAA AAT TTC ATT CAT TAT TAA CGG		this study
P10	LuxE-OL-sRBS-lacZ-rev	AAT GTA CCT CCT TAC TTT ATT TAT TGT ATT TGT TTA GCT ATC AAA CGC TTC GGT TAA GCT C		this study
P11	OL-sRBS-lacZ-fw	ACA AAT ACA ATA AAT AAA GTA AGG AGG TAC ATT ATG ACC ATG ATT ACG GAT TCA CTG GCC G		this study
P12	lacZ 500bp anti	CGA CTG TCC TGG CCG TAA CCG ACC		(6)
P13	delta CyaA fw	GTT GGC GGA ATC ACA GTC ATG ACG GGT AGC AAA TCA GGC GAT ACG TCT TGA ATT AAC CCT CAC TAA AGG GCG		this study
P14	delta CyaA rev	TCC GCT AAG ATT GCA TGC CGG ATA AGC CTC GCT TTC GGG CAC GTT CAT CAT AAT ACG ACT CAC TAT AGG GCT C		this study

		Primers *E. coli* loop mutation and overproduction constructs			
P15	EF-P *Eco* P32G fw	GTA AAA GGC GGT AAA GGC CAG G			this study
P16	EF-P *Eco* P32G rev	CCT GGC CTT TAC CGC CTT TTA C			this study
P17	EF-P *Eco* P32S fw	CGT AAA ATC GGG TAA AGG CCA GG			this study
P18	EF-P *Eco* P32S rev	CCT GGC CTT TAC CCG ATT TTA CG			this study
P19	EF-P *Eco* G33A fw	TAA AAC CGG CGA AAG GCC AGG			this study
P20	EF-P *Eco* G33A rev	CCT GGC CTT TCG CCG GTT TTA			this study
P21	EF-P *Eco* G33S fw	GTA AAA CCG TCG AAA GGC CAG G			this study
P22	EF-P *Eco* G33S rev	CCT GGC CTT TCG ACG GTT TTA C			this study
P23	EF-P *Eco* K34M fw	GTA AAA CCG GGT ATG GGC CAG GCA TTT			this study
P24	EF-P *Eco* K34M rev	AAA TGC CTG GCC CAT ACC CGG TTT TAC			this study
P25	EF-P *Eco* K34N fw	GTA AAA CCG GGT AAC GGC CAG GCA TTT			this study
P26	EF-P *Eco* K34N rev	AAA TGC CTG GCC GTT ACC CGG TTT TAC			this study
P27	EF-P *Eco* K34Q fw	GTA AAA CCG GGT CAG GGC CAG GCA TTT			this study
P28	EF-P *Eco* K34Q rev	AAA TGC CTG GCC CTG ACC CGG TTT TAC			this study
P29	EF-P *Eco* G35N fw	CCG GGT AAA AAC CAG GCA TTT GC			this study
P30	EF-P *Eco* G35N rev	GCA AAT GCC TGG TTT TTA CCC GG			this study
P31	EF-P *Eco* Q36S fw	GGG TAA AGG CAG CGC ATT TGC			this study
P32	EF-P *Eco* Q36S rev	GCA AAT GCG CTG CCT TTA CCC			this study

Primers *P. putida* loop mutation and overproduction constructs				
P33	NheI-NRBS-PP_1858-fw	GCA CTA GCT AGC CGC GGC CTC GAT TTT TAT AAA TCC	*Nhe*I	this study
P34	XbaI-PP_1858-GS-His6-rev	CGT CTA GAT TAG TGA TGG TGA TGG TGA TGC GAG CCC TTC TTG GAG CGG CCT TTG AA	*Xba*I	this study
P35	EF-P *Ppu* 29 30 31 OL fw	CGT AAC AGC GCG ATC ATG AAG ACC		this study
P36	EF-P Ppu K29A OL rev	CAT GAT CGC GCT GTT ACG GCC CGA CGC GGT GAA CTC AGC		this study
P37	EF-P Ppu K29R OL rev	CAT GAT CGC GCT GTT ACG GCC CGA GCG GGT GAA CTC AGC		this study
P38	EF-P *Ppu* S30A OL rev	CAT GAT CGC GCT GTT ACG GCC CGC CTT GGT GAA CTC AGC		this study
P39	EF-P *Ppu* S30G OL rev	CAT GAT CGC GCT GTT ACG GCC GCC CTT GGT GAA CTC AGC		this study
P40	EF-P *Ppu* S30P OL rev	CAT GAT CGC GCT GTT ACG GCC CGG CTT GGT GAA CTC AGC		this study
P41	EF-P *Ppu* G31A OL rev	CAT GAT CGC GCT GTT ACG CGC CGA CTT GGT GAA CTC		this study
P42	EF-P *Ppu* G31S OL rev	CAT GAT CGC GCT GTT ACG GCT CGA CTT GGT GAA CTC		this study
P43	EF-P *Ppu* R32K fw	ACC AAG TCG GGC AAG AAC AGC GCG ATC		this study
P44	EF-P *Ppu* R32K rev	GAT CGC GCT GTT CTT GCC CGA CTT GGT		this study
P45	EF-P *Ppu* 33 34 35 OL fw	ATC ATG AAG ACC AAG CTG AAG AAC CTG		this study
P46	EF-P *Ppu* N33D OL rev	CTT CAG CTT GGT CTT CAT GAT CGC GCT ATC ACG GCC CGA CTT		this study
P47	EF-P *Ppu* N33G OL rev	CTT CAG CTT GGT CTT CAT GAT CGC GCT GCC ACG GCC CGA CTT		this study
P48	EF-P *Ppu* S34A OL rev	CTT CAG CTT GGT CTT CAT GAT CGC CGC GTT ACG GCC CGA		this study
P49	EF-P *Ppu* S34Q OL rev	CTT CAG CTT GGT CTT CAT GAT CGC CTG GTT ACG GCC CGA		this study
P50	EF-P *Ppu* A35S OL rev	CTT CAG CTT GGT CTT CAT GAT GCT GCT GTT ACG GCC		this study

Primers cross modification/activation and overproduction				
P51	EF-P *Eco* domainI *Ppu* OL fw	AAG CTG GAC GAC GTG ATC CTG GAT ATG AAC CTG ACT TAC CTG		this study
P52	EF-P *Eco* domainI *Ppu* OL rev	CAG GTA AGT CAG GTT CAT ATC CAG GAT CAC GTC GTC CAG CTT		this study
P53	EF-P *Eco* loop *Ppu* OL fw	AAG TCG GGC CGT AAC AGC GCG TTT GCT CGC GTT AAA CTG CGT		this study
P54	EF-P *Eco* loop *Ppu* OL rev	CGC GCT GTT ACG GCC CGA CTT TAC GAA TTC ACT CGC TTC AAC		this study
P55	EF-P *Eco* P32S K34R OL fw	GAA TTC GTA AAA AGC GGT CGC GGC CAG GCA TTT		this study
P56	EF-P Eco P32S K34R OL rev	AAA TGC CTG GCC GCG ACC GCT TTT TAC GAA TTC		this study
P57	EF-P *Eco* K34R G35N OL fw	CCG GGT CGC AAC CAG GCA TTT GCT CG		this study
P58	EF-P *Eco* K34R G35N OL rev	CGA GCA AAT GCC TGG TTG CGA CCC GG		this study
P59	EF-P *Eco* K34R Q36S OL fw	CGC GGC AGC GCA TTT GCT CGC GTT A		this study
P60	EF-P *Eco* K34R Q36S OL rev	TAA CGC GAG CAA ATG CGC TGC CGC G		this study
P61	EF-P *Eco* P32S K34R G35N OL fw	AGC GGT CGC AAC CAG GCA TTT GCT CG		this study
P62	EF-P *Eco* P32S K34R G35N OL rev	CGA GCA AAT GCC TGG TTG CGA CCG CT		this study
P63	EF-P *Eco* P32S K34R Q36S OL fw	AGC GGT CGC GGC AGC GCA TTT GCT CG		this study
P64	EF-P *Eco* P32S K34R Q36S OL rev	CGA GCA AAT GCG CTG CCG CGA CCG CT		this study
P65	EF-P *Eco* K34R G35N Q36S OL fw	GGG TCG CAA CAG CGC ATT TG		this study
P66	EF-P *Eco* K34R G35N Q36S OL rev	CAA ATG CGC TGT TGC GAC CC		this study
P67	EarP So rev	GCG GTA CCC GAT TTT CTA TTT CAG CGC AGC AT	*Kpn*I	this study
Primers pET SUMO constructs				
P68	EF-P *Eco*-SUMO-fw	ATG GCA ACG TAC TAT AGC AAC GAT TTT		this study
P69	EF-P *Eco*-SUMO-rev	TTA GTG ATG GTG ATG GTG ATG GCT		this study
P70	EF-P *Eco* K34R OL fw	GTA AAA CCG GGT CGC GGC CAG GCA TTT		this study
P71	EF-P *Eco* K34R OL rev	AAA TGC CTG GCC GCG ACC CGG TTT TAC		this study

Primers isoelectric focusing and pBAD33-*epmA*				
P72	*efp-yjeA*-OL-XbaI-fw	CGT GAA GTA ATC TAG ATT GTC AAA AAC TGG AGA TTT AAC TAT GAG C	*Xba*I	this study
P73	*efp-yjeA*-OL-XbaI-rev	TTT TTG ACA ATC TAG ATT ACT TCA CGC GAG AGA CGT ATT CA	*Xba*I	this study
P74	*yjeA-yjeK*-OL-PstI-fw	GCA TAA CTG CAG GGT AGC TAA GCC ACA AAA TGG CG	*Pst*I	this study
P75	*yjeA-yjeK*-OL-PstI-rev	GCT ACC CTG CAG TTA TGC CCG GTC AAC GCT AAA G	*Pst*I	this study
P76	SacI-RBS-*efp*-fw	GCG ATG AGC TCA ATT AAC AAA TTT CAG AGG GCC TTA TGG	*Sac*I	this study
P77	SphI-*yjeK*-rev	GCA TCG CAT GCT TAC TGC TGG CGT AGC TGG AG	*Sph*I	this study
Primer for strain construction of MG-CR-efp-epmA-KanR				
P78	*epmA* fw	CAC CGC TGT TTG ATT CCT GCG T		this study
P79	empA rev	GCT ACA GAA TGG CGC TTA TCA CG		this study
Primer for for bacterial two-hybrid (BTH)				
P80	XbaI-EF-P*Eco*FW	GTA TCG TCT AGA GGC AAC GTA CTA TAG CAA CGA TTT TCG TG	*Xba*I	this study
P81	XmaI-EF-P*Eco*Rev	GTA TCG CCC GGG ACT TCA CGC GAG AGA CGT ATT CAC C	*Xma*I	this study

Supplemental Experimental Procedures

Molecular biological techniques

All kits and enzymes were used according to manufacturers instructions. Plasmid DNA was isolated using the Hi Yield® Plasmid Mini Kit from Süd Laborbedarf. DNA fragments were purified from agarose gels using the Hi Yield® Gel/PCR DNA fragment extraction 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.

SDS-PAGE and Western blotting

Cells were subjected to 12 % (w/v) sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) as described by Laemmli (14). To visualize and confirm protein separation, 2,2,2-trichloroethanol was incorporated into the polyacrylamide gels (15) and detected within a Gel DocTM EZ gel documentation system (Bio-Rad). Afterwards the proteins were transferred onto nitrocellulose membranes (Whatman) which were then subject for immunoblotting. The membranes were in a first step incubated either with 0.1 µg/mL Anti-His$_6$ antibody (Abcam, Inc.) to detect EF-P, or with 0.25 µg/ml Anti-Arg^{Rha} antibody (11) to detect rhamnosylation of EF-P. These primary antibodies (rabbit) were then targeted with 0.2 µg/ml Anti-rabbit phosphatase-conjugated secondary antibody (Rockland). Localization was visualized by adding development solution (50 mM sodium carbonate buffer, pH 9.5, 0.01% (w/v) p-nitroblue tetrazolium chloride (NBT) and 0.045% (w/v) 5-bromo-4-chloro-3-indolyl-phosphate (BCIP)).

Growth conditions

E. coli cells were routinely grown in Miller modified LB (16-18) at 37 °C aerobically under agitation, if not indicated otherwise. When required, media were solidified by using 1.5% (w/v) agar. The medium was supplemented when indicated with antibiotics at the following concentrations: 100 µg/ml ampicillin sodium salt, 50 µg/ml kanamycin sulfate, 30 µg/ml chloramphenicol, or 15 µg/ml tetracycline hydrochloride. Plasmids carrying the P$_{BAD}$ promoter (2) were induced with L-arabinose at a final concentration of 0.2% (w/v).

β-galactosidase activity assay

Whenever the plasmid based reporter system pBBR1MCS-3 XPPX *lacZ* was used, cells were grown over night (o/n) in 100 mM sodium-phosphate buffered Miller modified LB (pH 5.8), aerobically under agitation at 37 °C. Whenever the *E. coli* ΔepmA reporter strain (MG-CL-12-yjeA (8)) was used, cells were grown in potassium buffered KE minimal medium (19) pH 5.8, supplemented with 10 mM lysine, 0.2% glycerol and antibiotics in the appropriate concentrations. In both cases, the cells were harvested by centrifugation on the next day, and the β-galactosidase activities were determined as described (20) and are given in Miller units (MU) (21).

Protein overproduction and purification for NMR studies

To obtain labeled proteins for NMR studies, overproductions were performed in M9 glucose minimal medium (18) containing either 15N-labeled ammonium chloride alone (pET-SUMO-*efp_Eco*K34R, pET-SUMO-*efp_Eco* P32S, pET-SUMO-*efp_Eco* P32S K34R, pET-SUMO-*efp_Ppu*), or 15N-labeled ammonium chloride in combination with 13C labeled glucose (pET-SUMO-*efp_Eco*, pET-SUMO-*efp_Eco* loop_Ppu). The overproduction of these N-terminally His6-SUMO tagged hybrid EF-P variants was induced in E. *coli* BL21 (DE3) by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; Sigma Aldrich) during exponential growth. Till the induction point the cells were grown at 37 °C, after IPTG induction the temperature was shifted to 18 °C and the cells were grown o/n. On the next day, the cells were harvested by centrifugation. The resulting pellet was resuspended on ice in dialysis buffer 1 (100 mM Na2HPO4/NaH2PO4, pH 6.5, 1 mM DTT). Cells were lysed using a continuous-flow cabinet from Constant Systems Ltd. at 1.35 kb, in combination with sonication. The resulting lysate was clarified by centrifugation for 40 minutes at 4 °C at 39,810 x g. The His6-SUMO tagged proteins were purified in a first step using (Ni-NTA; Qiagen) according to the manufacturers 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 dialysis buffer 1 (100 mM Na2HPO4/NaH2PO4, pH 6.5, 1 mM DTT buffer). Afterwards, the His6-SUMO tag was cleaved off using His6-Ulp1 protease (22), followed by a second Ni-NTA purification step to remove the His6-SUMO tag itself as well as the His6 tagged Ulp1 protease. As a final step, the purified protein was dialyzed again o/n at 4 °C in dialysis buffer 1 (100 mM Na2HPO4/NaH2PO4, pH 6.5, 1 mM DTT).

C-terminally His$_6$-tagged EarP$_{Ppu}$ for NMR interaction studies was overproduced in E. *coli* LMG194 cells harboring a pBAD33-*earP_Ppu* plasmid in Miller modified LB at 37 °C. During exponential growth, 0.2% (w/v) L-arabinose was added to induce the protein production. After induction, the temperature was shifted to 18 °C, and the cells were grown o/n. On the next day, the cells were harvested by centrifugation. The resulting pellet was resuspended on ice in dialysis buffer 2 (100 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$, pH 7.5, 50 mM NaCl, 5 mM DTT). Cell lysis, clarification of the lysate and the first Ni-NTA purification step was performed as described above. In a final step, the purified protein was dialyzed o/n in dialysis buffer 2 (100 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$, pH 7.5, 50 mM NaCl, 5 mM DTT) to remove imidazole from the purification step.

Protein overproduction and purification for *in vitro* studies

To obtain EF-P variants for in *vitro* studies, overproductions were performed in E. *coli* LMG194 cells, grown in Miller modified LB, harboring the following C-terminally His$_6$-tagged EF-P constructs:

EF-P$_{Ppu}$: pBAD24-efp_{Ppu} S30P, pBAD24-efp_{Ppu} G31A, pBAD24-efp_{Ppu} G31S, pBAD24-efp_{Ppu} R32K, pBAD24-efp_{Ppu} N33G, pBAD24-efp_{Ppu} S34Q

EF-P$_{Eco}$: pBAD24-efp_{Eco} K34R, pBAD24-efp_{Eco} P32S K34R, pBAD24-efp_{Eco} K34R G35N, pBAD24-efp_{Eco} K34R Q36S, pBAD24-efp_{Eco} P32S K34R G35N, pBAD24-efp_{Eco} P32S K34R Q36S, pBAD24-efp_{Eco} K34R G35N Q36S

Furthermore, C-terminally His$_6$ tagged EarP$_{Ppu}$ was overproduced in *E. coli* LMG194 harboring pBAD33-$earP_{Ppu}$.

To overproduce proteins, cells with the corresponding plasmids were grown at 37 °C, and during exponential growth, 0.2% (w/v) L-arabinose was added to induce protein production. After induction, the temperature was shifted to 18 °C, and the cells were grown o/n. On the next day, the cells were harvested by centrifugation and the resulting pellet was resuspended in buffer 1 (100 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$, pH 7.5, 50 mM NaCl). Cells were than lysed by sonication and the resulting cell lysate was clarified by centrifugation for 40 minutes at 4 °C at 39,810 x *g*. The His$_6$ tagged proteins were then purified using Ni-NTA beads (Qiagen) according to the manufacturers instructions, whereby 20 mM imidazole was used for washing and 250 mM imidazole for elution of the His$_6$ tagged proteins. Subsequently, imidazole was removed by dialysis o/n at 4 °C in buffer 1 (100 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$, pH 7.5, 50 mM NaCl), followed by a second dialysis step at the next morning for 5 hours, again in buffer 1 (100 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$, pH 7.5, 50 mM NaCl). The resulting proteins were than used for *in vitro* rhamnosylation assays.

Determination of kinetic parameters

Kinetic parameters were determined by varying TDP-rhamnose concentrations while keeping concentrations of EarP$_{Ppu}$ (0.1 µM) and unmodified EF-P$_{Ppu}$ (2.5 µM) constant. A mixture of EarP$_{Ppu}$ and unmodified EF-P$_{Ppu}$ was equilibrated to 30 °C in 100 mM NaP$_i$ pH 7.6. The reaction was started by the addition of TDP-rhamnose and was stopped after 20 seconds of incubation at 30 °C by the addition of one volume 2x Laemmli buffer (14) and incubation at 95 °C for 5 minutes. Samples were subjected to SDS-PAGE and rhamnosylated EF-P$_{Ppu}$ was detected using an Anti-Arg^{Rha} antibody (11). A secondary FITC coupled Anti-rabbit antibody (Abcam, UK) was used to visualize rhamnosylation in a LI-COR Odyssey CLx. Band intensities were quantified using ImageJ (23). K_m values were determined by fitting relative band intensities to the Michaelis-Menten equation using SigmaPlot. The K_m of 5 µM TDP-rhamnose was determined using commercially available substrate (Carbosynth, UK). Previously, we determined a K_m of 50 µM using biochemically synthesized TDP-rhamnose (11). After rigorous assessment of this discrepancy we found that contaminations with ammonium acetate were responsible for a miscalculation of the TDP-rhamnose concentration in stock solutions. We further excluded that these contaminations had an effect on the *in vitro* rhamnosylation reaction.

In vitro rhamnosylation

In vitro rhamnosylation of EF-P$_{Eco}$ and EF-P$_{Ppu}$ variants were conducted in 100 mM NaPi pH 7.6 containing 50 mM NaCl. A master mix containing 25 µM of the corresponding EF-P variant and 100 µM TDP-β-L-rhamnose was prepared in a reaction tube and divided into 10 µl aliquots. The reaction was started by addition of 10 µl of 0.5 µM EarP solution and stopped after distinct time intervals by addition of 20 µl 2x Laemmli buffer and immediate heating to 95 °C in an Eppendorf ThermoMixer for 2 minutes. All samples were diluted by a factor of 10 in 1x Laemmli buffer and 20 µl (corresponding to 0.5 µg of EF-P) were subjected to SDS-PAGE and Western blotting. Rhamnosylated EF-P was detected and visualized using a polyclonal rabbit Anti-Arg^{Rha} and a fluorescence labeled Anti-rabbit antibody respectively. Band intensities were determined using ImageJ (23). Relative rhamnosylation rates were calculated by plotting the normalized linear range (intensity t_x/intensity$_{max}$) of the timecourse experiments and determining the slope of the resulting graphs.

Bacterial two-hybrid analysis

Protein-protein interactions were detected using the bacterial adenylate cyclase two-hybrid system kit (Euromedex) according to the product manual. The *E. coli* KV1 strain used in this study was generated by start to stop deletion of the *cyaA* gene from *E. coli* LF1 and subsequent incorporation of the *lux* operon at the *lac* locus by using described methods (6). Chemically competent *E. coli* KV1 cells were cotransformed with pKT25-EarP$_{Ppu}$ and/or variants of pUT18C coding for an EF-P isoform (EF-P$_{Ppu}$, EF-P$_{Eco}$, EF-P$_{Eco}$ loop$_{Ppu}$, EF-P$_{Eco}$ domainI$_{Ppu}$) that was to be tested for interaction. Cells were plated on LB medium containing 50 µg/ml kanamycin sulfate and 100 µg/ml carbenicillin and incubated at 30 °C o/n. Transformants carrying leucine-zipper-reporter hybrids (pUT18-zip and pKT25-zip) were used as positive controls, whereas transformants containing pUT18C and pKT25 vector backbones served as negative controls. For quantification of interaction strength, cells were cultivated in 96-well plates, with each well containing 200 µl of LB medium supplemented with 0.5 mM IPTG as well as 50 µg/ml kanamycin sulfate and 100 µg/ml carbenicillin. Plates were incubated at 30 °C and under moderate agitation (550 rpm in Eppendorf ThermoMixer) for 16 hours. On the next morning, Costar 96Well White plates containing 200 µl of LB medium (0.5 mM IPTG, 50 µg/ml kanamycin sulfate, 100 µg/ml carbenicillin) in each well, were inoculated with 2 µl of o/n culture and luminescence output was monitored every 10 minutes for 42 hours in a Tecan Spark with 240 rpm at 30 °C. For detection of protein-protein interaction on culture plates, cells were plated on LB plates containing 40 µg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) and 0.5 mM IPTG as well as 50 µg/ml kanamycin sulfate and 100 µg/ml carbenicillin after transformation and incubated at 30 °C o/n. Liquid cultures containing 50 µg/ml kanamycin sulfate, 100 µg/ml carbenicillin and 0.5 mM IPTG were inoculated from single colonies and incubated at 30 °C for 8 hours. 2 µl of

liquid culture were spotted on LB plates containing 40 µg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) and 0.5 mM IPTG as well as 50 µg/ml kanamycin sulfate and 100 µg/ml carbenicillin. Pictures were taken after 32 hours of incubation at 30 °C.

Protein overproduction for isoelectric focusing

C-terminally His$_6$-tagged EF-P$_{Eco}$ was overproduced in *E. coli* BW25113 and *E. coli* BW25113 Δ*epmA* cells, harboring the pBAD33-*efp*-His$_6$ plasmid, and were grown in Miller modified LB at 37 °C. Furthermore, *E. coli* BW25113 was transformed with pBAD33-*efp*-His$_6$-*epmAB* to produce post-translationally modified EF-P. During exponential growth, 0.2% (w/v) L-arabinose was added to induce protein production and cells were grown o/n at 18 °C. On the next day, cells were harvested by centrifugation. The resulting pellet was resuspended on ice in HEPES buffer (50 mM HEPES, 100 mM NaCl, 50 mM KCl, 10 mM MgCl$_2$, 5% (w/v) glycerol, pH 7.0). Cells were lysed using a continuous-flow cabinet from Constant Systems Ltd. at 1.35 kb. The resulting lysates were clarified by centrifugation for 1.5 h at 4 °C at 234,998 x *g*. The His$_6$-tagged proteins were purified using Ni-NTA beads (Qiagen) according to the manufacturers instructions, using 20 mM imidazole for washing and 400 mM imidazole for elution. In a final step, the purified protein was dialyzed o/n in HEPES buffer to remove imidazole from the purification step.

Isoelectric focusing

For isoelectric focusing 0.5 µg of protein per lane was loaded on a native vertical isoelectric focusing gel with a pH gradient range of 4-7 (SERVAGel^{TM}) containing approximately 3% (w/v) SERVALYT^{TM}. Prior to loading, samples were mixed with 2x IEF sample buffer according to instructions and wells were rinsed with SERVA IEF Cathode buffer. Focusing was conducted for one hour at 50 V, one hour at 300 V and finally bands were sharpened for 30 min at 500 V. Western blotting was conducted as described above using 0.1 µg/ml Anti-EF-P$_{Eco}$ (Eurogentec).

NMR Experiments

All ^{15}N NMR relaxation experiments for EF-P and its variants were performed in 100 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$, pH 6.5 and 1 mM DTT. NMR data were recorded at 298 K for ~ 0.15-0.18 mM of EF-P$_{Eco}$ and its variants except for EF-P$_{Eco}$ P32S for which data was recorded at 0.09 mM due to low yields of expression. Pulse experiments were performed on an 800 MHz Bruker Avance III NMR spectrometer equipped with TXI cryogenic probehead. Amide ^{15}N relaxation data of R_1, R_2, and steady-state heteronuclear {^{1}H}–^{15}N-NOE experiments were performed as described (24, 25) T_1 data were measured with 11 different relaxation delays: 20, 50, 100, 150, 250, 400, 500, 650, 800, 1000, and 1300, where 150 ms was used as duplicate. T_2 data were determined by using eight different relaxation delays: 16, 32, 48, 64, 80. 96, 112, and 128 ms using 16 ms as duplicate. Duplicate time points were used for error estimation. The correlation time (τ_c) of the protein molecule was estimated using the ratio of

averaged T_2/T_1 values (24). Steady-state heteronuclear {^{1}H}–^{15}N-NOE experiments were recorded with and without 3 s of ^{1}H saturation. All relaxation experiments were acquired as pseudo-3D experiments. The spectra were processed with NMRPipe (26) and peak integration and relaxation parameter calculation was performed using PINT (27).

For the titration of EF-P$_{Eco}$ and its variants with 2x EarP$_{Ppu}$, both the proteins were dialyzed against 100 mM Na$_2$HPO$_4$/NaH$_2$PO$_4$, pH 7.5, 50 mM NaCl and 1 mM DTT. Experiments were recorded on an 800 MHz Bruker NMR spectrometer equipped with a TXI cryogenic probehead at 298 K. Protein backbone assignments for EF-P$_{Eco}$ and EF-P$_{Eco}$ loop$_{Ppu}$ were obtained from HNCACB, CBCA(CO)NH and HNCA experiments (28). Data analysis was performed in CcpNmr Analysis software (29). Resonance assignments of EF-P variants have been deposited at the BMRB with the following accession codes: XXX.

Bioinformatics software

Multiple sequences alignments (MSA) were built using MAFFT software (version 7.394, algorithm "l-ins-i") (30). HMMER web-server was used to predict domain architecture of the proteins (31). Phylogenetic trees were built using FastTree tool (32) with default settings and visualized using *ggtree* package in R (33). Sequence logos were built using *ggseqlogo* package in R (34).

Collection of the data for evolutionary analysis

EF-P/IF5A, EpmA, EpmB, EpmC and EarP sequences were collected by extracting sequences that belong to specific family from databases and additional filtration.

Proteins with EF-P_N Superfamily domain were extracted from Pfam database as the initial set of EF-P/IF5A (35). We have filtered out sequences that: (1) don't belong to the fully sequenced genomes [ftp://ftp.ncbi.nlm.nih.gov/genomes/GENOME_REPORTS/], (2) are marked in UniProtKB (36) as obsolete or fragments, (3) have "Uncertain" protein existence levels (UniProtKB notation), (4) have no gene name in their UniProtKB entry, (5) belong to the outliers of the lengths distribution in the protein family, (6) are missing the function GO:0003746 "translation elongation factor activity". In some species EF-P has nonfunctional paralogous protein - YeiP (Elongation factor P-like protein). All proteins that are encoded by 'yeiP' or 'elongation factor P-like protein' genes and sequences that share average identity with YeiP proteins more than 0.26 (excluding gaps) were removed. Functional EF-P is known to have 2 OB-Fold domain, while functional IF5A should have 1 OB-Fold domain (37). Number of OB-Fold domains in the sequences was predicted by hmmscan. All IF5A with no predicted OB-domains were removed. IF5A with 2 OB-domains (with exception for those that are localized in chloroplasts) also were removed. EF-P containing 1 or 0 OB-domains were also removed, except those EF-P that belong to the genomes with a second copy of EF-P with 2 OB-domains. Sequences *A0A0S8CMN6, Q3Z8J4, Q2Y9K0, A0A0C3RPA1, A0A0F7KHW6, A0A0S8CBC3, A0A0S8DH18, C7NBD9, U2STQ5, C9MW58, A0A0X8JVW4,*

A0A0E3ZAU2, D0GJB3, D1AKL3, D1AYT5, A0A0F9ZIK0, A0A0G0Q7K6 were excluded from the data set for weblogo building because of their untypical loop regions that bring gaps into the MSA of EF-P_N domains.

Sequences containing tRNA-synthetase II domain were downloaded from Pfam database, as this is the only domain that EpmA proteins are known to have. Sequences were filtered using (1) - (5) criteria. MSA and phylogenetic tree for the remaining sequences were built and the branch with sequences that are encoded by *epmA* genes (gene name '*epmA*', '*genX*', '*yjeA*' or '*poxA*' in UniProtKB) was taken as the set of EpmA proteins.

EpmB proteins are part of lysine aminomutases (LAMs) family that was downloaded from InterPro database (35). Initial filters (1) - (5) were applied to this dataset. Sequences without predicted Radical_SAM domain or with predicted full LAM_C domain were removed from the data set.

EpmC and EarP sequences were extracted from Pfam database by EpmC and DUF2331 domains, correspondingly, and additionally filtered by (1) - (3) criteria.

Final sets of EF-P/IF5A, EpmA, EpmB, EpmC and EarP count 4421, 858, 4894, 317 and 306 proteins, respectively (Table S1).

Evolutionary analysis

Phylogenetic tree of EF-P/IF5A proteins was built from MSA of the *EF-P_N Superfamily* domain sequences. The belonging of EF-P/IF5A and modification system proteins to the same genome was determined by the NCBI Taxonomy IDs corresponding to the sequences. EF-P/IF5A, EpmA, EpmB, EpmC and EarP proteins belonged to 4169, 845, 3976, 308 and 302 genomes, respectively. Among the genomes encoding EF-P/IF5A proteins 808, 773, 300 and 294 genomes also encoded EpmA, EpmB, EpmC and EarP, respectively (Table S1 and Figure S1B).

Evolutionary reconstruction

Using the MSA of the EF-P KOW-like N-domain I a rooted phylogenetic tree was built using the *phylogeny* tool from MAFFT online server (UPGMA algorithm with default settings) (30). The leafs of the tree were mapped to three states: 'EpmA', 'EarP' and '-' (no modification enzymes), representing the presence/absence of the modification proteins. The evolutionary state reconstruction of ancestral states was done using the maximum likelihood (ML) method from the phytools R package (38). The phylogenetic tree annotated with the likelihood of ancestral states was visualized using phytools. The area of the surface colored with the certain color represents the predicted probability of the ancestor to have the certain state. Using the same method, an evolutionary reconstruction of the K34R substitution was performed. The leafs of the tree were mapped to amino acids observed at the 34th position: A, G, H, K, M, N, Q, R and S. *E. coli* EF-P was used as reference sequence.

Supplemental References

1. Macinga **DR, Parojcic MM, Rather PN.** 1995. Identification and analysis of *aarP*, a transcriptional activator of the 2'-N-acetyltransferase in *Providencia stuartii*. J Bacteriol **177:**3407-3413.
2. **Guzman LM, Belin D, Carson MJ, Beckwith J.** 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P$_{BAD}$ promoter. J Bacteriol **177:**4121-4130.
3. **Studier FW, Moffatt BA.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol **189:**113-130.
4. **Karimova G, Dautin N, Ladant D.** 2005. Interaction network among *Escherichia coli* membrane proteins involved in cell division as revealed by bacterial two-hybrid analysis. J Bacteriol **187:**2233-2243.
5. **Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H.** 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol **2:**2006 0008.
6. **Fried L, Lassak J, Jung K.** 2012. A comprehensive toolbox for the rapid construction of *lacZ* fusion reporters. J Microbiol Methods **91:**537-543.
7. **Lassak J, Keilhauer EC, Fürst M, Wuichet K, Gödeke J, Starosta AL, Chen JM, Søgaard-Andersen L, Rohr J, Wilson DN, Häussler S, Mann M, Jung K.** 2015. Arginine-rhamnosylation as new strategy to activate translation elongation factor P. Nat Chem Biol **11:**266-270.
8. **Ude SCM.** 2013. The role of elongation factor EF-P in translation and in copy number control of the transcriptional regulator CadC in *Escherichia coli*. Dissertation. LMU München: Faculty of Biology, München.
9. **Datsenko KA, Wanner BL.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci USA **97:**6640-6645.
10. **Volkwein W, Maier C, Krafczyk R, Jung K, Lassak J.** 2017. A versatile toolbox for the control of protein levels using N^{ε}-acetyl-L-lysine dependent amber suppression. ACS Synth Biol **6:**1892-1902.
11. **Krafczyk R, Macosek J, Jagtap PKA, Gast D, Wunder S, Mitra P, Jha AK, Rohr J, Hoffmann-Roder A, Jung K, Hennig J, Lassak J.** 2017. Structural Basis for EarP-Mediated Arginine Glycosylation of Translation Elongation Factor EF-P. mBio **8:**e01412-01417.
12. **Peil L, Starosta AL, Lassak J, Atkinson GC, Virumae K, Spitzer M, Tenson T, Jung K, Remme J, Wilson DN.** 2013. Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. Proc Natl Acad Sci USA **110:**15265-15270.
13. **Lassak J, Henche AL, Binnenkade L, Thormann KM.** 2010. ArcS, the cognate sensor kinase in an atypical Arc system of *Shewanella oneidensis* MR-1. Appl Environ Microbiol **76:**3263-3274.
14. **Laemmli UK.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:**680-685.
15. **Ladner CL, Yang J, Turner RJ, Edwards RA.** 2004. Visible fluorescent detection of proteins in polyacrylamide gels without staining. Anal Biochem **326:**13-20.
16. **Bertani G.** 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. J Bacteriol **62:**293-300.
17. **Bertani G.** 2004. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. J Bacteriol **186:**595-600.
18. **Miller JH.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Epstein W, Kim BS.** 1971. Potassium transport loci in *Escherichia coli* K-12. J Bacteriol **108:**639-644.
20. **Tetsch L, Koller C, Haneburger I, Jung K.** 2008. 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.

21. **Miller JH.** 1992. A short course in bacterial genetics: a laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory, Cold Spring Harbor N.Y.

22. **Starosta AL, Lassak J, Peil L, Atkinson GC, Woolstenhulme CJ, Virumae K, Buskirk A, Tenson T, Remme J, Jung K, Wilson DN.** 2014. A conserved proline triplet in Val-tRNA Synthetase and the origin of elongation factor P. Cell Rep **9:**476-483.

23. **Schneider CA, Rasband WS, Eliceiri KW.** 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods **9:**671-675.

24. **Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE.** 1994. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. Biochemistry **33:**5984-6003.

25. **Korzhnev DM, Skrynnikov NR, Millet O, Torchia DA, Kay LE.** 2002. An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. J Am Chem Soc **124:**10743-10753.

26. **Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A.** 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR **6:**277-293.

27. **Niklasson M, Otten R, Ahlner A, Andresen C, Schlagnitweit J, Petzold K, Lundstrom P.** 2017. Comprehensive analysis of NMR data using advanced line shape fitting. J Biomol NMR **69:**93-99.

28. **Sattler M, Schleucher J, Griesinger C.** 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog Nucl Mag Res Sp **34:**93-158.

29. **Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED.** 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins **59:**687-696.

30. **Katoh K, Standley DM.** 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol **30:**772-780.

31. **Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR.** 2015. HMMER web server: 2015 update. Nucleic Acids Res **43:**W30-38.

32. **Price MN, Dehal PS, Arkin AP.** 2010. FastTree 2-approximately maximum-likelihood trees for large alignments. PLoS ONE **5:**e9490.

33. **Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y.** 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol **8:**28-36.

34. **Wagih O.** 2017. ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics **33:**3645-3647.

35. **Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A.** 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res **44:**D279-285.

36. **UniProt Consortium T.** 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res **46:**2699.

37. **Rossi D, Kuroshu R, Zanelli CF, Valentini SR.** 2014. eIF5A and EF-P: two unique translation factors are now traveling the same road. Wiley Interdiscip Rev RNA **5:**209-222.

38. **Revell LJ.** 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol **3:**217-223.

Supplemental information – Chapter 5

Supporting Information

A Versatile Toolbox for the Control of Protein Levels using N^ϵ-acetyl-L-lysine Dependent Amber Suppression

Wolfram Volkwein[1], Christopher Maier[1], Ralph Krafczyk[1], Kirsten Jung[1]* & Jürgen Lassak[1]*

[1]Center for integrated Protein Science Munich (CiPSM) at the Department of Biology I, Microbiology,

Ludwig-Maximilians-Universität München, Großhaderner Strasse 2-4, 82152 Martinsried, Germany.

*To whom correspondence should be addressed:

E-mail: jung@lmu.de

E-mail: juergen.lassak@lmu.de

Supplementary Tables

Supplementary Table S1. Strains used in this study.

Stain	Relevant genotype or description	Reference or source
E. coli DH5αλpir	F⁻ 80*lacZ*ΦM15 (*lacZYA-argF*)U196 *recA1 hsdR17 deoR thi-1 supE44 gyrA96 relA1*/*pir*	1
E. coli BW25113	F⁻ λ⁻ Δ(*araD-araB*)567 Δ*lacZ4787*(::rrnB-3) *rph-1* Δ(*rhaD-rhaB*)568 *hsdR514*	2
E. coli MG1655	wild-type; F- lambda- *ilvG rfb50 rph-1*	3
Salmonella enterica	Serovar Typhimurium LT2	ATCC 19585
Vibrio cholera C6706	O1 El Tor isolate from Peru	Centers for Disease Control and Prevention (CDC)
E. coli LF1	F⁻ λ⁻ *ilvG rfb50 rph-1 rpsL150* P$_{lac}$::*rpsL-neo-kan::lacZ*$^{\Delta 1-100bp}$, *Kan^r, Strp^s*	4
E. coli LF1-LacZ-K	Derivative from *E. coli* LF1, contains an insertion of a lysine codon (AAA) after position 9 of the *lacZ* gene, *Kan^s Str^r*	this study
E. coli LF1-LacZ(Am)	Derivative from *E. coli* LF-1, contains an insertion of an amber codon (TAG) after position 9 of the *lacZ* gene, *Kan^s, Str^r*	this study
E. coli LF1-LacZ-PPK	Derivative from *E. coli* LF1, contains an insertion of a proline-proline-lysine motif (CCG CCG AAA) after position 9 of the *lacZ* gene, *Kan^s, Str^r*	this study
E. coli LF1-LacZ-PP(Am)	Derivative from *E. coli* LF1, contains an insertion of a proline-proline-amber motif (CCG CCG TAG) after position 9 of the *lacZ* gene, *Kan^s, Str^r*	this study
E. coli LF1-AcKRST	Derivative from *E. coli* LF1, contains the acetyl lysyl-tRNA-synthetase (AcKRS), a mutated version of the PylRS from *Methanosarcina mazei* towards N^ε-acetyl lysine specificity, according to Umehara *et al.*[5] and the cognate amber suppressor tRNA$_{CUA}$ (*pylT*), *Kan^s, Str^r*	this study
E. coli LF1-AcKRST/*dapA:cam^r*	Derivative from *E. coli* LF1-AcKRST were the *dapA* gene was replaced by a chloramphenicol resistance cassette using λ-RED recombinase, *Cam^r*	this study

Supplementary Table S2. Plasmids used in this study.

Plasmid	Features	References
pBBR1MCS-2	Broad-host-range cloning vector, 5.1 kb, ND[a]; *Kan*[r]	6
pBBR1MCS-3	Broad-host-range cloning vector, 5.2 kb, ND[a]; *Tet*[r]	6
pBBR1MCS-4	Broad-host-range cloning vector, 5.0 kb, ND[a]; *Amp*[r]	6
pBBR1MCS-5	Broad-host-range cloning vector, 4.9 kb, ND[a]; *Gm*[r]	6
pBBR1MCS-2 AcKRST	Broad-host-range cloning vector, Contains the acetyl lysyl-tRNA synthetase (*pylS**) under control of P$_{glnS}$ and the suppressor tRNA$_{CUA}$ (*pylT*) under the control of P$_{proK}$, inserted into the *Nsi*I restriction site, 6.8 kb, ND[a]; *Kan*[r]	this study
pBBR1MCS-3 AcKRST	Broad-host-range cloning vector. Contains the acetyl lysyl-tRNA synthetase (*pylS**) under control of P$_{glnS}$ and the suppressor tRNA$_{CUA}$ (*pylT*) under the control of P$_{proK}$, inserted into the *Age*I restriction site, 6.9 kb, ND[a]; *Tet*[r]	this study
pBBR1MCS-4 AcKRST	Broad-host-range cloning vector. Contains the acetyl lysyl-tRNA synthetase (*pylS**) under control of P$_{glnS}$ and the suppressor tRNA$_{CUA}$ (*pylT*) under the control of P$_{proK}$, inserted into the *Nsi*I restriction site, 6.6 kb, ND[a]; *Amp*[r]	this study
pBBR1MCS-5 AcKRST	Broad-host-range cloning vector. Contains the acetyl lysyl-tRNA synthetase (*pylS**) under control of P$_{glnS}$ and the suppressor tRNA$_{CUA}$ (*pylT*) under the control of P$_{proK}$, inserted into the *Nsi*I restriction site, 6.5 kb, ND[a]; *Gm*[r]	this study
pBBR1MCS-5 TT-RBS-lux	PCR template for the amplification of the *lux* operon *luxCDABE* from *Photorhabdus luminescens*, ND[a]; *Gm*[r]	7
pBAD/HisA	pBR322-derived expression vector, contains the promotor P$_{BAD}$ of the arabinose operon *araBAD* from *E. coli* and its regulatory gene *araC*, 4.1 kb, *Amp*[r]	Invitrogen
pBAD/HisA-Lux	Contains the *lux* operon inserted into the *Nco*I / *Kpn*I restriction sites, 9.8 kb, *Amp*[r]	this study
pBAD/HisA-Lux(Am)	Contains the *lux* operon with an amber codon (TAG) in *luxC* at position 3, inserted into the *Nco*I / *Bsr*GI restriction sites of pBAD/HisA-Lux, 9.8 kb, *Amp*[r]	this study
pBAD/HisA-Lux(2Am)	Contains the *lux* operon with two amber codons (TAG) in *luxC* at position 3, inserted into the *Nco*I / *Bsr*GI restriction sites of pBAD/HisA-Lux, 9.8 kb, *Amp*[r]	this study
pBAD/HisA-Lux-RA(Am)	Contains the *lux* operon with an arginine alanine amber motif (CGG GCT TAG) in *luxC* at position 3, inserted into the *Nco*I / *Bsr*GI restriction sites of pBAD/HisA-Lux, 9.8 kb, *Amp*[r]	this study
pBAD/HisA-Lux-DA(Am)	Contains the *lux* operon with an aspartic-acid alanine amber motif (GAT GCT TAG) in *luxC* at position 3, inserted into the *Nco*I / *Bsr*GI restriction sites of pBAD/HisA-Lux, 9.8 kb, *Amp*[r]	this study
pBAD/HisA-Lux-HH(Am)	Contains the *lux* operon with an histidine histidine amber motif (CAC CAC TAG) in *luxC* at position 3, inserted into the *Nco*I / *Bsr*GI restriction sites of pBAD/HisA-Lux, 9.8 kb, *Amp*[r]	this study
pBAD/HisA-Lux-KDP(Am)	Contains the *lux* operon with an lysine aspartic-acid proline amber motif (AAA GAT CCG TAG) in *luxC* at position 3, inserted into the *Nco*I / *Bsr*GI restriction sites of pBAD/HisA-Lux, 9.8 kb, *Amp*[r]	this study
pBAD/HisA-Lux-PP(Am)	Contains the *lux* operon with an proline amber motif (CCG TAG) in *luxC* at position 3, inserted into the *Nco*I / *Bsr*GI restriction sites of pBAD/HisA-Lux, 9.8 kb, *Amp*[r]	this study
pBAD/HisA-Lux-KDPP(Am)	Contains the *lux* operon with an lysine aspartic acid proline proline amber motif (AAA GAT CCG CCG TAG) in *luxC* at position 3, inserted into the *Nco*I / *Bsr*GI restriction sites of pBAD/HisA-Lux, 9.8 kb, *Amp*[r]	this study
pNL1.1	PCR template for amplification of NanoLuc® (*Nluc*), 3.1 kb, *Amp*[r]	Promega
pBBR1MCS-2 AcKRST Nluc	Derived from pBBR1MCS-2 AcKRST, contains *Nluc* under control of P$_{BAD}$ inserted into the *Xba*I / *Kpn*I restriction sites, 8.7 kb, *Kan*[r]	this study
pBBR1MCS-2 AcKRST Nluc(Am)	Derived from pBBR1MCS-2 AcKRST, contains *Nluc* under control of P$_{BAD}$ inserted into the *Xba*I / *Kpn*I restriction sites, insertion of an amber codon in *Nluc* at position 3, 8.7 kb, *Kan*[r]	this study
pBBR1MCS-2 AcKRST Nluc PP(Am)	Derived from pBBR1MCS-2 AcKRST, contains *Nluc* under control of P$_{BAD}$ inserted into the *Xba*I / *Kpn*I restriction sites, insertion of an proline proline amber motif in *Nluc* at position 3, 8.7 kb, *Kan*[r]	this study
pBAD/HisA-Kan	pBAD/HisA (Invitrogen) derived expression vector, the ampicillin resistance cassette (*bla*) was replaced with a	this study

	kanamycin resistance cassette from pBBR1MCS-2, 4.2 kb, *Kan^r*	
pBAD/HisA-Kan Lux	Derived from pBAD/HisA (Kan), contains the *lux* operon inserted restriction free, 9.8 kb, *Kan^r*	this study
pBAD/HisA-Kan-Lux(Am)	Derived from pBAD/HisA (Kan), contains the *lux* operon inserted restriction free, insertion of an amber codon (TAG) in *luxC* at position 3, 9.8 kb, *Kan^r*	this study
pBAD/HisA-dapA	Contains *dapA* inserted between the *Xho*I / *Eco*RI restriction sites, 5.0 kb, *Amp^r*	this study
pBAD/HisA-dapA(Am)	Derived from pBAD/HisA-dapA, contains *dapA* with an amber codon on position 3 of the *dapA* open reading frame, inserted between the *Xho*I / *Eco*RI restriction sites, 5.0 kb, *Amp^r*	this study
pACYCDuet™-1	Standard expression vector, ORI P15A, 4 kb, *Cam^r*	Novagen
pACYCDuet-PylRST	Contains the pyrrolysyl-tRNA-synthetase (PylRS) / cognate amber suppressor tRNA$_{CUA}$ (*pylT*) pair from *Methanosarcina mazei*, 5.5 kb, *Cam^r*	[8]
pACYCDuet-AcKRST	Contains the acetyl lysyl-tRNA-synthetase (AcKRS), a mutated version of the PylRS from *Methanosarcina mazei* towards N^ϵ-acetyl lysine specificity, according to Umehara *et al.*[5] and the cognate amber suppressor tRNA$_{CUA}$ (*pylT*), 5.5 kb, *Cam^r*	this study
pRed/ET	λ-RED recombinase in pBAD24; *Amp^r*	Gene Bridges

[a] ND, the incompatibility group of pBBR1 MCS plasmids has not been defined [9]; compatible with IncP, IncQ and IncW group plasmids, as well as ColE1- and P15a-based replicons [6]. *Amp^r*, *Cam^r*, *Kan^r*, *Tet^r*, *Gm^r* and *Str^r* are ampicillin, chloramphenicol, kanamycin, tetracycline, gentamycin, and streptomycin resistance, respectively.

Supplementary Table S3. Primers used in this study.

Primer name	Sequence	Restriction site	Reference
Sequencing and control primers			
AcKRS chk fw	AAT AAG TTC CTC ACA AAG GCA AAC GAA GAC		this study
pBBR1MCS245 NsiI chk rev	CAA GGC GAC AAG GTG CTG ATG		this study
pBBR1MCS3 AgeI chk rev	TGC GAT GAG TGG CAG GGC GGG GC		this study
Seq33 fw-100	GGC GTC ACA CTT TGC TAT GC		[10]
AraC chk fw	CCT GAC CGC GAA TGG TGA GAT TGA GA		this study
Ara Prom chk fw	CTG ACG CTT TTT ATC GCA ACT CTC TAC TG		this study
dapA chk fw	CAT GAA GCT CCG CAA GCG GT		this study
dapA chk rev	CAC CAG ATA ATG TTG CGA TGA CAGT		this study
Primers to generate a mutated version of the PylRS from *Methanosarcina mazei* towards Nε-acetyl lysine specificity (AcKRS). Described primers a directly derived from the library primers described by Umehara *et al.*[5]			
AcKRS L301M Y306L L309A OL fw	AGA ACT TCT GCC TGA GAC CCA TGA TGG CTC CAA ACC TTC TGA ACT ACG CGC GCA AGC TTG ACA GGG CCC		this study
AcKRS L301M Y306L L309A OL rev	GGG CCC TGT CAA GCT TGC GCG CGT AGT TCA GAA GGT TTG GAG CCA TCA TGG GTC TCA GGC AGA AGT TCT		this study
AcKRS C348F OL fw	CAT GCT GAA CTT CTT TCA GAT GGG ATC G		this study
AcKRS C348F OL rev	CGA TCC CAT CTG AAA GAA GTT CAG CAT G		this study
MPYSf	TTT CCC TGA ATT CCG GCA AGC		[5]
AcKRS rev	ATC GGC GAG AAA GTC AGC AGG CCG CGC GGC CGC TTA CAG GTT GGT AGA AAT CCC GTT ATA GTA AGA CTC		this study
Primers for Lux reporter constructs			
KpnI LuxE rev	GCT TCG AAT TCC CAT ATG GTA CCT TAT CAA ACG CTT CGG TTA AGC TCA A	*Kpn*I	this study
NcoI LuxC-wt fw	CTA ACA GGA GGA ATT AAC CAT GGG CAC TAA AAA AAT TTC ATT CAT TAT TAA CGG CCA G	*Nco*I	this study
NcoI LuxC:Am fw	CTA ACA GGA GGA ATT AAC CAT GGG CTA GAC TAA AAA AAT TTC ATT CAT TAT TAA CGG CCA G	*Nco*I	this study
NcoI LuxC:2Am fw	CTA ACA GGA GGA ATT AAC CAT GGG CTA GTA GAC TAA AAA AAT TTC ATT CAT TAT TAA CGG CCA G	*Nco*I	this study
NcoI LuxC:RA Am fw	CTA ACA GGA GGA ATT AAC CAT GGG CCG GGC TTA GAC TAA AAA AAT TTC ATT CAT TAT TAA CGG CCA G	*Nco*I	this study
NcoI LuxC:DA Am fw	CTA ACA GGA GGA ATT AAC CAT GGG CGA TGC TTA GAC TAA AAA AAT TTC ATT CAT TAT TAA CGG CCA G	*Nco*I	this study
NcoI LuxC:HH Am fw	CTA ACA GGA GGA ATT AAC CAT GGG CCA CCA CTA GAC TAA AAA AAT TTC ATT CAT TAT TAA CGG CCA G	*Nco*I	this study
NcoI LuxC:KDP Am fw	CTA ACA GGA GGA ATT AAC CAT GGG CAA AGA TCC GTA GAC TAA AAA AAT TTC ATT CAT TAT TAA CGG CCA G	*Nco*I	this study
NcoI LuxC:PP Am fw	CTA ACA GGA GGA ATT AAC CAT GGG CCC GCC GTA GAC TAA AAA AAT TTC ATT CAT TAT TAA CGG CCA G	*Nco*I	this study
NcoI LuxC:KDPP Am fw	CTA ACA GGA GGA ATT AAC CAT GGG CAA AGA TCC GCC GTA GAC TAA AAA AAT TTC ATT CAT TAT TAA CGG CCA G	*Nco*I	this study
LuxC rev	TCA AAA TCT TTT TTG GCA TTC GGT		this study
Primers for *E. coli* LF1 LacZ XXX strain construction and verification			
lacI 583bp sense	GTC TGC GTC TGG CTG GCT GGC ATA		[4]
lacZ 500bp anti	CGA CTG TCC TGG CCG TAA CCG ACC		[4]
lacZ 220bp anti	AGC TTT CCG GCA CCG CTT CTG		[4]
LacZ-K-OL-fw	TCA CTG GCC AAA GTC GTT TTA C		this study
LacZ-K-OL-rev	GTA AAA CGA CTT TGG CCA GTG A		this study
LacZ-PPK-OL-fw	TCA CTG GCC CCG CCG AAA GTC GTT TTA C		this study
LacZ-PPK-OL-rev	GTA AAA CGA CTT TCG GCG GGG CCA GTG A		this study
LacZ-Am-OL-fw	TCA CTG GCC TAG GTC GTT TTA C		this study
LacZ-Am-OL-rev	GTA AAA CGA CCT AGG CCA GTG A		this study
LacZ-PPAm-OL-fw	TCA CTG GCC CCG CCG TAG GTC GTT TTA C		this study
LacZ-PPAm-OL-rev	GTA AAA CGA CCT ACG CGC GGG CCA GTG A		this study
Primers for *E. coli* LF1-AcKRST strain construction			
Plac-AcKRST-OL-fw	GGA AAC AGC TAT GGA TAA AAA ACC ACT AAA		this study

	CAC TCT GAT A		
Plac-AcKRST-OL-rev	GGT TTT TTA TCC ATA GCT GTT TCC TGT GTG AAA TTG TTA TCC		this study
Plac-OL-sRBS-lacZ-fw	ACA AAT ACA ATA AAT AAA GTA AGG AGG TAC ATT ATG ACC ATG ATT ACG GAT TCA CTG GCC G		this study
Plac-AcKRST-OL-sRBS-lacZ-rev	AAT GTA CCT CCT TAC TTT ATT TAT TGT ATT TGT GCA AAA AAG CCT GCT CGT TGA GC		this study
Primer for *E. coli* LF1-AcKRST/*dapA:cam*[r] strain construction			
FRT dapA fw	CCA GGC GCG ACT TTT GAA CAG AGT AAG CCA TCA AAT CTC CCT AAA CTT TAA ATT AAC CCT CAC TAA AGG GCG		this study
FRT dapA rev	CAT ACC AAA CGT ACC ATT GAG ACA CTT GTT TGC ACA GAG GAT GGC CCA TGT AAT ACG ACT CAC TAT AGG GCT C		this study
Primers for *dapA* constructs			
dapA EcoRI fw	CCG GAA TTC TTA CAG CAA ACC GGC ATG CTT AAG	*Eco*RI	this study
dapA XhoI rev	CCG CTC GAG ATG TTC ACG GGA AGT ATT GTC GCG	*Xho*I	this study
Amber dapA XhoI rev	CCG CTC GAG ATG TTC TAG GGA AGT ATT GTC GCG	*Xho*I	this study
Primers for broad-host-range cloning vectors			
AcKRST NsiI fw	TGC ATG CAT TCA TCA ATC ATC CCC ATA ATC CTTG	*Nsi*I	this study
AcKRST NsiI rev	TGC ATG CAT GCA AAA AAG CCT GCT CGT TGA GCA G	*Nsi*I	this study
AcKRST BspEI fw	GGC ATT CTC CGG ATC ATC AAT CAT CCC CAT AAT CCT TG	*Bsp*EI	this study
AcKRST BspEI rev	GGC ATT CTC CGG AGC AAA AAA GCC TGC TCG TTG AGC AG	*Bsp*EI	this study
Primers for resistance cassette exchange in pBAD/HisA			
pBAD/HisA amp kan ex fw	TAT ATG AGT AAA CTT GGT CTG ACA GTC AGA AGA ACT CGT CAA GAA GGC GA		this study
pBAD/HisA amp kan ex rev	CTT TTT GTT TAT TTT TCT AAA TAC ATA GCT TGC AGT GGG CTT ACA TGG CG		this study
Primers for NanoLuc® reporter constructs			
Ara XbaI fw	TGC TCT AGA TTA TGA CAA CTT GAC GGC TAC	*Xba*I	this study
Ara Nluc OL rev	CAT GGT TAA TTC CTC CTG TTA GC		this study
Ara Nluc OL fw	GCT AAC AGG AGG AAT TAA CCA TGG TCT TCA CAC TC		this study
Ara Nluc am OL fw	GCT AAC AGG AGG AAT TAA CCA TGG GCT AGG TCT TCA CAC TCG AAG AT		this study
Ara Nluc am PP OL fw	GCT AAC AGG AGG AAT TAA CCA TGG GCC CGC CGT AGG TCT TCA CAC TCG AAG AT		this study
Nluc KpnI rev	CGG GGT ACC TTA CGC CAG AAT GCG TTC	*Kpn*I	this study

Supplementary Table S4. Amber suppression in *Salmonella enterica* LT2 (A) and *Vibrio cholerae* El Tor C6706 (B). Cells were grown aerobically overnight at 37 °C in LB and on the next day, luminescence production in response to $N\varepsilon$-acetyl-L-lysine (AcK) and L-arabinose (Ara) was measured for the reporter pBBR1MCS-2 AcKRST Nluc, pBBR1MCS-2 AcKRST Nluc^{Am} and pBBR1MCS-2 AcKRST Nluc$^{PP(Am)}$, using the Nano-Glo® luciferase assay system.

A

Salmonella enterica LT2	- AcK - Ara	+AcK (1mM)	+ Ara (0.2%)	+AcK (1mM) + Ara (0.2%)
■ nLuc	23,925	38,640	1,935,681	1,559,454
▨ Amber nLuc	19,516	56,584	215,665	692,714
▨ PP Amber nLuc	2,332	8,646	52,880	124,702

B

Vibrio cholerae El Tor C6706	- AcK - Ara	+AcK (1mM)	+ Ara (0.2%)	+AcK (1mM) + Ara (0.2%)
■ nLuc	24,780	41,781	1,760,650	1,834,598
▨ Amber nLuc	686	2,163	256,137	672,902
▨ PP Amber nLuc	33	203	56,303	177,643

Supplementary Figures
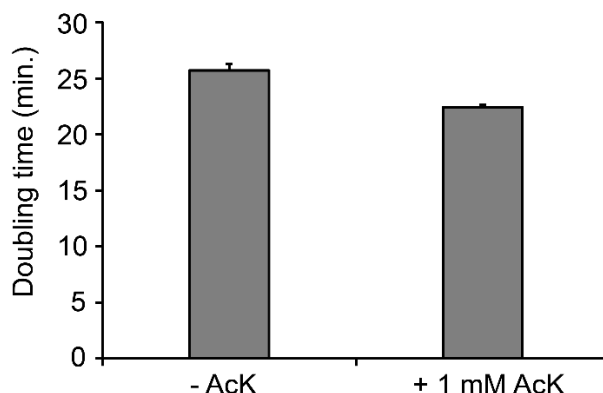
Supplementary Figure S1. Influence of $N\varepsilon$-acetyl-L-lysine (AcK) on the doubling time of *E. coli* BW25113 during exponential growth. Cells were grown in LB in the absence (- AcK) and in the presence of 1 mM AcK (+ 1 mM AcK). Error bars represent the standard deviation of data from three different experiments.
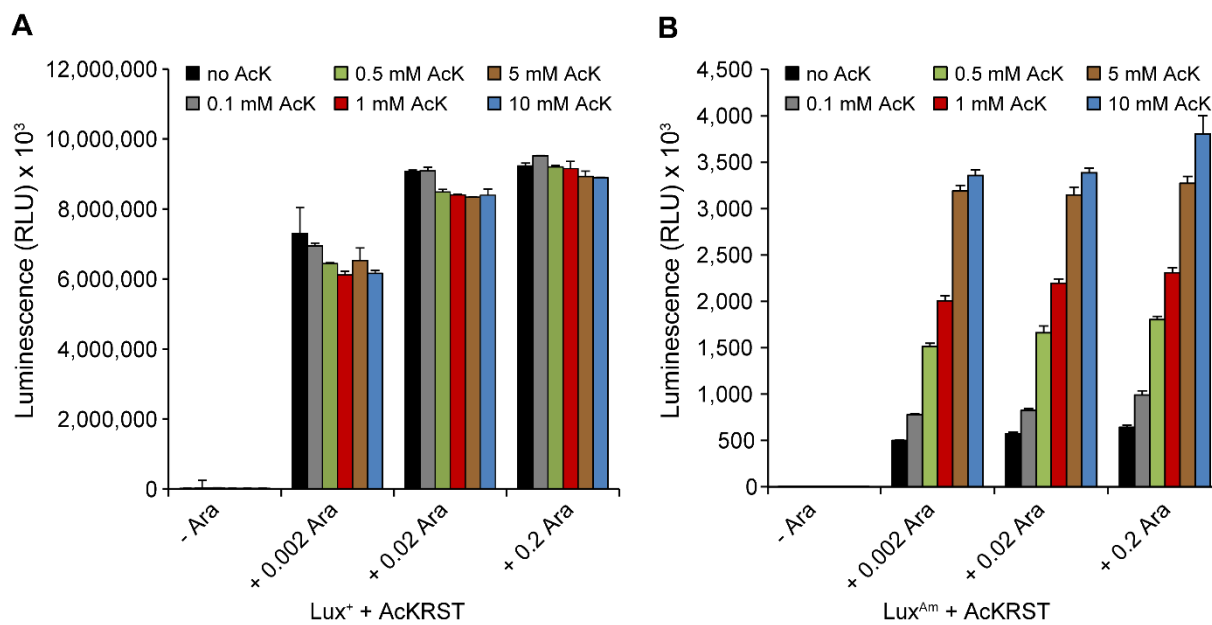
Supplementary Figure S2. Luminescence production of Lux$^+$ (A) and Lux^{Am} (B) in *E. coli* BW25113 in response to L-arabinose (Ara) and of $N\varepsilon$-acetyl-L-lysine (AcK). The maximal luminescence normalized to the OD_{600} from a 16 h time course experiment is depicted in response to Ara and AcK for Lux$^+$ (A) and Lux^{Am} (B). Error bars represent the standard deviation of data from three different experiments.
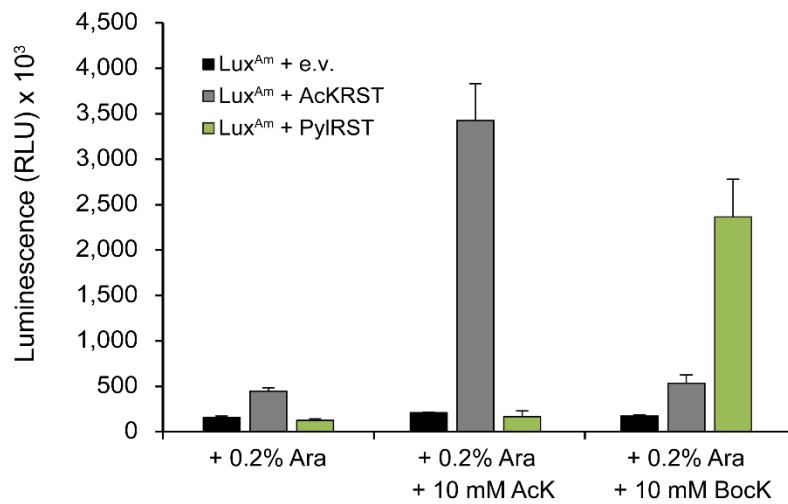
Supplementary Figure S3. Comparison of PylRS and AcKRS based amber suppression. *E. coli* BW25113 cells were transformed with the Lux-amber reporter (Lux^{Am}) in combination with either pACYCDuet-AcKRST (encoding the acetyl lysyl-tRNA synthetase/tRNA$_{CUA}$ pair, AcKRST), pACYCDuet-PylRST (encoding the pyrrolysyl-tRNA synthase/tRNA$_{CUA}$ pair, PylRST) or the empty vector (e.v.) as negative control. These cells were grown aerobically in LB supplemented with 0.2% (w/v) L-arabinose (Ara) at 37°C in a microtiter plate within a Tecan Infinite® F500 system (TECAN). The maximal luminescence production in response to $N\varepsilon$-acetyl-L-lysine (AcK) and $N\varepsilon$-*tert*-butoxycarbonyl-L-lysine (BocK) normalized to the OD$_{600}$ from a 16 h time course experiment is depicted. Error bars represent the standard deviation of data from three different experiments.
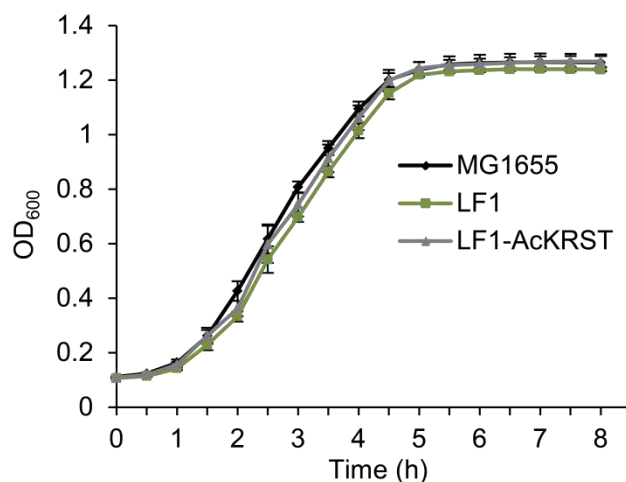
Supplementary Figure S4. Influence of the chromosomal integration of the acetyl lysyl-tRNA synthetase/tRNA$_{CUA}$ pair (AcKRST) on cell growth. *E. coli* MG1655, LF1 and LF1-AcKRST were grown aerobically at 37°C in a microtiter plate within a Tecan Infinite® F500 system (TECAN) in LB. Error bars represent the standard deviation of data from three different experiments.
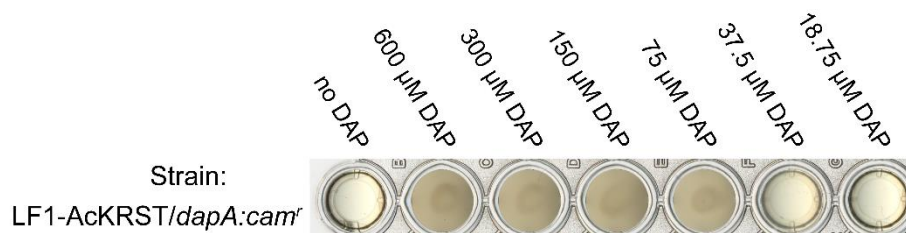
Supplementary Figure S5. Determination of the minimal diaminopimelic acid (DAP) concentration required for the growth of the *dapA* deficient LF1-AcKRST/*dapA*:*cam^r* strain. Depicted are cells grown for 16h at 37 °C aerobically in a microtiter plate in LB, supplemented with varying DAP concentrations ranging from 600 µM to 0 µM.

REFERENCES

(1) Macinga, D. R., Parojcic, M. M. and Rather, P. N. (1995) Identification and analysis of *aarP*, a transcriptional activator of the 2'-N-acetyltransferase in *Providencia stuartii*. *J. Bacteriol. 177*, 3407-3413.

(2) Datsenko, K. A. and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA 97*, 6640-6645.

(3) Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. and Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science 277*, 1453-1462.

(4) Fried, L., Lassak, J. and Jung, K. (2012) A comprehensive toolbox for the rapid construction of *lacZ* fusion reporters. *J. Microbiol. Methods 91*, 537-543.

(5) Umehara, T., Kim, J., Lee, S., Guo, L. T., Soll, D. and Park, H. S. (2012) *N*-acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess *N*-acetyl lysine specificity in vitro and in vivo. *FEBS Lett. 586*, 729-733.

(6) Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M., 2nd and Peterson, K. M. (1995) Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene 166*, 175-176.

(7) Gödeke, J., Heun, M., Bubendorfer, S., Paul, K. and Thormann, K. M. (2011) Roles of two *Shewanella oneidensis* MR-1 extracellular endonucleases. *Appl. Environ. Microbiol. 77*, 5342-5351.

(8) Gattner, M. J., Vrabel, M. and Carell, T. (2013) Synthesis of ε-*N*-propionyl-, ε-*N*-butyryl-, and ε-*N*-crotonyl-lysine containing histone H3 using the pyrrolysine system. *Chem. Commun. 49*, 379-381.

(9) Antoine, R. and Locht, C. (1992) Isolation and molecular characterization of a novel broad-host-range plasmid from *Bordetella bronchiseptica* with sequence similarities to plasmids from gram-positive organisms. *Mol. Microbiol. 6*, 1785-1799.

(10) Ude, S., Lassak, J., Starosta, A. L., Kraxenberger, T., Wilson, D. N. and Jung, K. (2013) Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science 339*, 82-85.

Danksagung

Curriculum Vitae

PERSONAL INFORMATION

Ralph Fabian Krafczyk

Munich, Germany

Ralph.Krafczyk@gmail.com

EDUCATION

10/2013 - 10/2015   **M.Sc. in Biology (Grade 1.13)**, Department of Biology,
                       Ludwig-Maximilians-Universität Munich, Germany
10/2010 - 07/2013   **B.Sc. in Biology (Grade 2.05)**, Department of Biology,
                       Ludwig-Maximilians-Universität Munich, Germany

EXPERIENCE

08/2015 - 12/2018   Doctoral student - LMU Munich, Microbiology, Kirsten Jung Lab

01/2014 - 06/2014   Research assistant - LMU Munich

09/2013 - 11/2013   Student assistant - LMU Munich

02/2013 - 03/2013   Student assistant - LMU Munich

UNDERGRADUATE RESEARCH PROJECTS

02/2015 - 08/2015 LMU Munich, Germany; Project for Master thesis with Prof. Dr. Kirsten Jung: Initial characterization of the EF-P arginine rhamnosyltransferase EarP

04/2013 - 06/2013 LMU Munich, Germany; Project for Bachelor thesis with Prof. Dr. Heinrich Jung: Generation and characterization of a functional, cystein-free variant of the L-carnitine/γ-butyrobetaine-antiporter CaiT from *Escherichia coli*

Languages

German - first language

English - very good knowledge

French - basic knowledge

Memberships

Member of the the Association for General and Applied Microbiology (VAAM)

Member of the Federation of European Microbiological Societies (FEMS)

Member of the Research Training Group „Molecular Principles of Synthetic Biology" GRK2062 of the German Research Foundation (DFG)