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*

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

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 Pseudomonas 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—β-lysylate a conserved lysine, K34EF-P, by EpmA. This EF-P-specific ligase...
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, R32EF-P, in the equivalent position (17, 24, 25). Rhamnosylation is mediated by the recently discovered inverting glycosyltransferase EarP, which utilizes dTDP-β-L-rhamnose (TDP-Rha) as donor substrate, resulting in β-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-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<sub>Ppu</sub>) 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<sub>Eco</sub>) (37) and O-GlcNAc transferase (OGT) from *Xanthomonas campestris* (OGT<sub>Xca</sub>) (38) as structural orthologues. Accordingly, EarP<sub>Ppu</sub> 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).
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 (β8 to β13) and seven α-helices (α8 to α14) (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 (β1 to β7), surrounded by the α-helices (α1 to α5 and α15) (see Fig. 4A and Fig. S3A). In the crystal structure, only β-strands β1, β2, and β3, as well as α-helices α1, α5, and α15, 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 EarPppu 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 carbohydrate-active enzymes (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 EarPppu crystal structure is located in a binding pocket that is composed of residues located in the C-domain (Fig. 3A). F191EarP, F252EarP, and F258EarP 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 Q255EarP. The diphosphate is recognized by hydrogen bonds formed with the side chain guanidine of R271EarP, the Y193EarP side chain hydroxyl, and backbone amides of E273EarP and D274EarP. The binding pocket is closed by the bulky side chain of Y193EarP, 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 EarPppu and EarPppu 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 EarPppu constructs with single-amino-acid substitutions of the individual residues forming the binding pocket and tested the activities of the EarPppu variants both in vivo and in vitro (Fig. 5). This included F191EarP, F252EarP, and F258EarP, which form the aromatic pocket, as well as Y193EarP, Q255EarP, R271EarP, and D274EarP, which are involved in hydrogen bond networking (Fig. SB).

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
the activation of the lysine-dependent acid stress response by the transcriptional activator CadC (11). Similarly, coproduction of wild-type EF-P<sub>Ppu</sub> and wild-type EarP<sub>Ppu</sub> (WT<sub>EarP</sub>) can restore β-galactosidase activity in an E. coli P<sub>cadBA</sub>Δ<sub>efp</sub>Δ<sub>lacZ</sub> strain (Fig. 5A and S1B). From the nine tested EarP<sub>Ppu</sub> substitution variants, we measured reduced β-galactosidase activities for the variants F191AEarP, Y193AEarP, R271AEarP, S275AEarP, and Y291AEarP. The variants R271AEarP and Y291AEarP failed to induce β-galactosidase expression at all (Fig. 5B and S1B).

In parallel, the enzymatic activity of EarP<sub>Ppu</sub> was investigated in vitro by employing an anti-Arg<sub>Rha</sub> antibody. The antibody was raised against a chemically synthesized glycopeptide antigen (SGR<sub>Rha</sub>NAAIVK) and specifically detects arginine rhamnosylation...
This in turn enabled the quantification of rhamnosylation rates of EF-P

\[ \text{Ppu} \] by Western blot analysis (Fig. 5C and D). In a first step, the \( K_m \) and \( k_{\text{cat}} \) of WT-\text{EarP} were determined to be 53 \( \mu \)M 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 \( \text{P. putida} \), \( \text{P. aeruginosa} \), and \( \text{E. coli} \), which were 3.5 mM, respectively.

(see Materials and Methods) (Fig. S1A). This in turn enabled the quantification of rhamnosylation rates of EF-P\(_{\text{Ppu}}\) by Western blot analysis (Fig. 5C and D). In a first step, the \( K_m \) and \( k_{\text{cat}} \) of WT-\text{EarP} were determined to be 53 \( \mu \)M 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 \( \text{P. putida} \), \( \text{P. aeruginosa} \), and \( \text{E. coli} \), which were 3.5 mM,
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 WTEarP 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 F252AEarP variant ($K_m = 3.4$ mM). Conversely, the $k_{cat}$ decreased up to 3,500 times when we measured the kinetics of the F191AEarP and Y193AEarP variants.

To exclude the possibility that decreased enzyme activity was due to fold disruption, selected EarP$_{Ppu}$ variants (F191AEarP, Y193AEarP, F252AEarP, R271AEarP, D274AEarP, and Y291AEarP) 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 D274AEarP 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 D274EarP in the protein C-domain and an equally conserved arginine at position 23 (R23EarP) 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 D274AEarP 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 D274AEarP SEC profile is congruent with the one that we found when subjecting WTEarP to SEC. Accordingly, $K_m$ (TDP-Rha) and $k_{cat}$ values were determined from this protein fraction to be 206 $\mu$M 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 WTEarP as well as the above-mentioned nine substitution variants.
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<sub>Ppu</sub> and WTEarP do interact, then CyaA is reconstituted, which in turn allows induction of the lac promoter and results in lacZ expression. Accordingly, /H9252-galactosidase activity is a measure of the interaction strength. When coproducing EF-P<sub>Ppu</sub> with WTEarP, we determined ca. 250 MU, whereas combinations with solely T25 and T18 resulted in /H11021 MU, thus defining the threshold of the assay (Fig. S1C). Except for the R271AEarP and Y291AEarP 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 /H<sup>1</sup>H-/H<sup>15</sup>N HSQC results between unbound EF-P<sub>Ppu</sub> and EarP<sub>Ppu</sub>-bound EF-P<sub>Ppu</sub> (Fig. 6A). Triple-resonance experiments of EF-P<sub>Ppu</sub> 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<sub>Ppu</sub> (Fig. 6B and C). However, R32EF-P<sub>F</sub> and residues surrounding the rhamnosylation site (e.g., S30EF-P<sub>F</sub>, G31EF-P<sub>F</sub>, R32EF-P<sub>F</sub>,...
N33EF-\(\text{P}\) 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\(\text{Ppu}\) 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\(\text{Ppu}\) but might also be propagating effects. Therefore, we also investigated in vitro rhamnosylation of truncated EF-P\(\text{Ppu}\) variants comprising either amino acids 1 to 128 or amino acids 1 to 65 (Fig. 6D). Both truncations were readily rhamnosylated by EarP\(\text{Ppu}\), further corroborating that EF-P contact sites are predominantly located in the KOW-like N-domain.

In addition, we compared NMR interactions between EarP\(\text{Ppu}\) and unmodified EF-P\(\text{Ppu}\) or rhamnosylated EF-P\(\text{Ppu}\). 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-\(\beta\)-\(L\)-rhamnose to \(\alpha\)-rhamnosyl arginine (26, 27). Reportedly, inverting glycosyltransferases employ a direct-displacement \(S_N2\)-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\(\text{EarP}\) and D17\(\text{EarP}\) and glutamate E273\(\text{EarP}\)—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\(\text{EarP}\) and D17\(\text{EarP}\) variants as well as the E273\(\text{EarP}\) 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 D13\(\text{AEarP}\), D17\(\text{AEarP}\), and E273\(\text{AEarP}\) 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\(\text{Ppu}\) protein variants, \(^{15}\)N HSQCs were measured for D13\(\text{AEarP}\), D17\(\text{AEarP}\), and E273\(\text{AEarP}\). The spectra show no structural alterations from WT\(\text{EarP}\) (Fig. 7B, C, and D and see Fig. S7). Additionally, the variants D13\(\text{AEarP}\) and D17\(\text{AEarP}\) were titrated with TDP-Rha being indistinguishable from WT\(\text{EarP}\) perturbations. Interestingly, although D13\(\text{EarP}\) and D17\(\text{EarP}\) resonances could not be assigned, other residues in close proximity (G16\(\text{EarP}\) and G19\(\text{EarP}\)) exhibited strong perturbations not only in WT\(\text{EarP}\) but also in the D13\(\text{AEarP}\) and D17\(\text{AEarP}\) variants upon TDP-Rha binding, despite not forming direct ligand contacts (Fig. 7E). Similarly, we could measure TDP-Rha binding for E273\(\text{AEarP}\) variants using STD NMR (Fig. S5C). This confirms that these mutations do not affect donor substrate binding.

To investigate interactions between EF-P\(\text{Ppu}\) and the D13\(\text{AEarP}\), D17\(\text{AEarP}\), and E273\(\text{AEarP}\) 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-\(\beta\)-D-galactopyranoside)-containing LB plates (Fig. 7F, S1C).

To further corroborate our findings on the in vitro-inactive D13\(\text{AEarP}\), D17\(\text{AEarP}\), and E273\(\text{AEarP}\) variants, they were subjected to an in vivo experiment in which we investigated their ability to activate EF-P\(\text{Ppu}\) (Fig. 5A). Additional substitutions—D13N\(\text{EearP}\), D17N\(\text{EearP}\), and E273Q\(\text{DearP}\)—were also included in the study. Expectedly, coproduction of the D13\(\text{AEarP}\), D17\(\text{AEarP}\), and E273\(\text{AEarP}\) variants with EF-P\(\text{Ppu}\) phenocopies \(\Delta\text{efp}\)
with respect to \( P_{\text{cadBA}} \) activation and in vivo rhamnosylation (Fig. 7G; Fig. S1B). Similar results were obtained with the D17N/E EarP and E273QEarP variants, whereas the D13EEarP and E273DEarP 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 WTEarP, 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-lysyl-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-acetylglucosaminylates (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 D13EarP, D17EarP, and E273EarP in the EarP glycosyltransferase family might play a role in activating the R32 guanidino group of EF-P. Especially D17EarP and E273EarP—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 D13EarP, D17EarP, and E273EarP, 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 POFTU1 (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 F252EarP and F258EarP 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 R271EarP in the equivalent position and a substitution to alanine (R271AEarP) 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 R32EF-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 elf-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, R271EarP might stabilize the nucleotide product, thereby facilitating leaving group departure. Upon successful inverting glycosyl transfer from TDP-Rha to R32EF-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. For 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 *Ppa* 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 HiYield 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 Russell (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-*ep*<sub>pu</sub> 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 pUT18C-earP and pKT25-earP 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<sub>6</sub>-tagged EarP<sub>pu</sub> variants (pBAD33-earP<sub>pu</sub>) were overproduced in *E. coli* MG194 by addition of 0.2% arabinose to exponentially growing cells and subsequent cultivation at 18°C o/n. N-terminally His<sub>6</sub>-tagged EarP (pACYC-DUET-earP<sub>pu</sub>) and His<sub>6</sub>-SUMO-tagged EF-P<sub>pu</sub> (pET-SUMO-earP<sub>pu</sub>) 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<sub>pu</sub> (EF-P<sup>RM</sup>) was produced by cooverproduction with His<sub>6</sub>-tagged EarP<sub>pu</sub>. Cells were lysed by sonication, and His<sub>6</sub>-tagged proteins were purified using Ni-nitrilotriacetic acid (Ni-NTA; Qiagen) according to the manufacturer's instructions. The His<sub>6</sub>-SUMO tag was removed by incubation with 1 μg/ml His<sub>6</sub>-Ulp1 (71) overnight. Subsequently, tag-free EF-P<sub>pu</sub> 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, 15N-labeled nitrogen (15NH<sub>4</sub>Cl) and 13C-labeled glucose were used. For NMR backbone assignment of EarP<sub>pu</sub> additionally 99.8%-pure heavy water D<sub>2</sub>O (Sigma-Aldrich) was used instead of H<sub>2</sub>O in growth medium to allow partial deuterium of the protein in order to reduce cross-relaxation effects and increase the signal-to-noise ratio. Size exclusion chromatography of EarP<sub>pu</sub> and the D274AEarP variant was performed in 100 mM NaPi (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<sub>pu</sub> *E. coli* BL21(DE3) cells expressing N-terminally His<sub>6</sub>-tagged EarP<sub>pu</sub> were cultivated in 1 liter M9 minimal medium at 37°C to an optical density at 600

September/October 2017 Volume 8 Issue 5 e01412-17 mbio.asm.org 13
nm (OD_{600}) of 0.6. One hundred microliters of threonine, 100 μg lysine, and 50 μg isoleucine were added to feedback inhibit methionine biosynthesis (72). Additionally, 50 μg L-(-)-serine 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-pro and EF-P_{Ppu} were dialyzed against 100 mM NaPi, pH 7.6, 5 mM dithiothreitol (DTT), whereas a buffer composed of 100 mM NaPi, 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. CH2Cl2 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 F254 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-% C18 column (250 by 21.2 mm). Analytical RP-HPLC was measured on a JASCO system with a Phenomenex Aeres Peptide 5-% C18 column (250 by 4.6 mm). In all cases, mixtures of water (eluents A and acetoneitrile (eluents 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 C18 column (2.6 μm, 100 Å) at 40°C (Fig. 8).

Glycopeptide SGR™-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 (Oxyme Pure), 0.5 eq of N,N'-diisopropylcarbodiimide (DIC), and 5 eq of N,N'-diisopropylcarbodiimide. All coupling reactions were conducted at 75°C and 28 W for 5 min. Removal of the allyloxy-carbonyl-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™-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 C44H82N14O16 [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™-NAAIVK. To test the specificity of the purified polyclonal antibodies toward EF-P_{Ppu}, 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-ArgRha™ 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% (w/v) glycogen and 1.7% (w/v) agarose and transferred onto a nitrocellulose membrane by vertical Western blotting. Antigens were detected using 0.1 μg/ml anti-His6 tag (Abcam, Inc.), 0.2 μg/ml anti-EF-P, or 0.25 μg/ml of anti-ArgRha. 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 8** Synthesis of glycopeptide SGR™-NAAIVK. a, SiPhH3 (phenylsilane), Pd(PPh3)4, CH2Cl2; b, 1-((tert-butoxycarbonyl)-3-(2,4-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).
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<sub>Ppu</sub> (0.1 μM) and unmodified EF-P<sub>Ppu</sub> (2.5 μM) constant. A mixture of EarP<sub>Ppu</sub> and unmodified EF-P<sub>Ppu</sub> was equilibrated to 30°C in 100 mM NaP<sub>P</sub>, (pH 7.6). The reaction was started by the addition of TDP-Rha and 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<sub>Ppu</sub> 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<sub>Ppu</sub>. Kin<sub>cat</sub> and K<sub>m</sub> values were determined by fitting reaction rates (in nanomoles per milligram per second) to the Michaelis-Menten equation using SigmaPlot.

Fold recognition. Fold recognition models were generated using the online user interface of Phyre2 (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<sub>600</sub> of 0.5 (5 × 10<sup>8</sup> cells/ml), harvested by centrifugation, and resuspended in 25 ml 100 mM NaP<sub>P</sub>, (pH 7.6) (2 × 10<sup>10</sup> 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<sub>Ppu</sub> (0.1 μM) and unmodified EF-P<sub>Ppu</sub> (2.5 μM) was equilibrated to 30°C in 10 μl 100 mM NaP<sub>P</sub>, (pH 7.6). The reaction was started by addition of 10 μl lysate from ~2 × 10<sup>10</sup> or ~2 × 10<sup>9</sup> 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<sub>Ppu</sub> 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 intensitites 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<sup>3</sup> for E. coli (84) and 2.1 μm<sup>3</sup> 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 <sup>15</sup>N<sup>-</sup> and <sup>13</sup>C-labeled EF-P (700 μM) in 100 mM NaP<sub>P</sub>, 50 mM NaCl, and 5 mM DTT (pH 7.6), 0.02% NaN<sub>3</sub> was added to the sample. Sequential resonance assignment was obtained from two-dimensional (2D) 1H-15N HSQC and three-dimensional (3D) HNCACB, CBCACONH, and HNCACB backbone experiments, using a constant time during <sup>13</sup>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 <sup>13</sup>C<sub>α</sub> and <sup>13</sup>C<sub>β</sub> chemical shifts and random coil chemical shifts of the same nuclei to assign a secondary structure to the EF-P sequence (Fig. 5E) 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 1H, 15N- and 13C-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 15N-labeled protein sample and monitoring the progress of the titration by recording 1H-15N HSQC. First, 15N-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. 15N-labeled 41 μM rhamnosylated EF-P was titrated with unlabeled EarP to a 1:2 EF-P/EarP molar ratio without any intermediate steps. 15N-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. 15N-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 CSP = Δδ/H0.15, 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, 1H-15N HSQC spectra of 15N-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 WTEarP 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 7.6, 150 mM NaCl, 1 mM DTT, and 10% D2O. 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 WTEarP sample with a 100-fold excess of ligand, and after a total saturation time of 5 s, with 4,096 scans performed for the WTEarP 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. CRYOSOL (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 His6-tagged EarP, expressed as a selenomethionine 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 an 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], reffmac [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.

ACKNOWLEDGMENTS

We thank Ingrid Weitl for excellent technical assistance. We thank Wolfram Volkwein for fruitful discussions. The SAXS and X-ray diffraction experiments were performed on beamlines BM29 and ID29/ID23-1, respectively, at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. We are grateful to local contacts at the ESRF for providing assistance in using beamlines BM29, ID29, and ID23-1. We are also grateful to Christoph Müller, Irmgard Sinning, and Klemens Wild for fruitful discussions regarding the crystal structure of EarP. Lastly, we thank Bernard Henrissat and the glycogenomics group at AFMB in Marseille, France, for building EarP into the CAZy database as the new family GT104.

J.L., K.J., and A.H.R. gratefully acknowledge financial support from the DFG Research Training Group GRK2062 (Molecular Principles of Synthetic Biology). Moreover, J.L. is grateful for DFG grant LA 3658/1-1. J.H. acknowledges support from the European Molecular Biology Laboratory (EMBL). P.K.A.J. acknowledges EMBL and the EU Marie Curie Actions CoFund for an EIPOD fellowship. K.J. and A.H.R. additionally thank the Center for integrated Protein Science Munich (Cluster of Excellence grant Exc114/2). The work of J.R., P.M., and A.K.J. was supported by U.S. National Institutes of Health grant GM 105977.

A.H.R., S.W., and D.G. performed the organic synthesis and NMR analysis of small molecules and wrote the corresponding section of Materials and Methods. R.K. performed the confirmation of antibody specificity raised against the rhamnosyl-arginine-comprising peptide. Additionally, R.K. constructed the EarP<sub>Ppu</sub>- and EF-P<sub>Ppu</sub>-encoding plasmids and purified all proteins used for biochemical analyses, NMR studies, and X-ray crystallography. R.K. also performed the biochemical in vivo/in vitro characterization of EarP<sub>Ppu</sub> and determined concentrations of TDP-β-L-rhamnose in E. coli, P. putida, and P. aeruginosa. TDP-β-L-rhamnose was synthesized by J.R., P.M., and A.K.J., J.H., J.M., and P.K.A.J. performed and analyzed all protein NMR experiments. The crystallization screen was set up by J.M. J.H., and P.K.A.J. solved the crystal structure of EarP<sub>Ppu</sub>. J.L., J.H., and K.J. designed the study. The manuscript was written by R.K., J.M., P.K.A.J., K.J., J.H., and J.L.

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 106:30–34. https://doi.org/10.1073/pnas.0809211106.

3. Tanner DR, Cariello DA, Woolstenholme 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. Woolstenholme CJ, Parajuli S, Healey DW, Valverde DP, Petersen EN, Starosta AL, Gudosh 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, Woolstenholme CJ, Kim JR, Saini P, Buskirk AR, Dever TE. 2013. elf5A 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. elf5A 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. elf5A func-
ions globally in translation elongation and termination. Mol Cell 66: 194–205.e5. https://doi.org/10.1016/j.molcel.2017.03.003.

8. Doerfler LK, Wohlgemuth I, Kothe C, Peske F, Urlaub H, Rodnina MV. 2013. EF-P residue is essential for rapid synthesis of proteins containing consecutive proline residues. Science 339:85–88. https://doi.org/10.1126/science.1229917.

9. Hersch SJ, Wang M, Zou SB, Moon KM, Foster LJ, Ibbas M, Naravare 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, 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 110:15262–15270. https://doi.org/10.1073/pnas.1310642110.

11. Ude S, Lassak J, Starosta AL, Kraxenberger T, Wilson DN, Jung K. 2013. 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.

12. Woolstenhulme CJ, Gydosh NR, Green R, Buskirk AR. 2015. High-precision analysis of translational pausing by ribosome profiling in bacteria lacking EF-P. Cell Rep 11:113–21. https://doi.org/10.1016/j.celrep.2015.03.014.

13. 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 101:9995–9960. http://www.pnas.org/10.1073/pnas.0309661101.

14. Blaha G, Stanley RE, Steitz TA. 2009. Formation of the first peptide consecutive proline residues. Science 339:85–88. https://doi.org/10.1126/science.1136194.

15. 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. https://doi.org/10.1093/nar/gku768.

16. Breton C, Fourmel-Guex S, Palic MM. 2012. Recent structures, evolution and mechanisms of glycosylation transferases. Curr Opin Struct Biol 22:540–549. https://doi.org/10.1016/j.sbi.2012.06.007.

17. Singh DG, Lokamo J, Lokamo 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.

18. Peterson JS, Giogga C, O’Reilly LA, Bankovacki A, Nachbur U, Infusini G, Webb AI, 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.

19. Guex N, Peitsch MC. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18: 2714–2723. https://doi.org/10.1002/(SICI)10.1002/elps.115081805.

20. 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.2009.22.

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

22. 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.

23. 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.

24. Saibil HR, Walker D, Shi Y, Breij T, Breij J, Coates 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/10.1111/j.1158-1083.1998.tb01054.x.

25. Martinez-Fleites C, Macauley MS, He Y, Shen DL, Vodac PJ, 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/nsmb1443.

26. 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.

27. Li X, Krafczyk R, Macošek J, Li YL, Zou Y, Simon B, Parker BL, Payne RJ, 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 antibiotic. Chem Sci 7:6995–7001. https://doi.org/10.1039/c6sc02889f.

28. Wang S, Corliss L, Sharp PP, Rajkovic A, Ibbas 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/C6CS03847F.

29. Coutinho PM, Deleury E, Davies GJ, Henrisatt 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.

30. Karimova G, Pidoux J, Ullmann A, Ladant D. 1998. A bacterial two-component system linking of arginine-rhamnosylated translation elongation factor P trig-
EarP Structure and Biochemistry

65. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. 1989. Site-directed

55. Lira-Navarrete E, Valero-González J, Villanueva R, Martínez-Júlvez M,

50. Spiro RG. 2002. Protein glycosylation: nature, distribution, enzymatic

46. Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases:

52. Wong Fok Lung T, Giogha C, Creuzburg K, Ong SY, Pollock GL, Zhang

51. Li S, Zhang L, Yao Q, Li L, Dong N, Rong J, Gao W, Ding X, Sun L, Chen

58. Qasba PK, Ramakrishnan B, Boeggeman E. 2005. Substrate-induced

59. Ni L, Sun M, Yu H, Chokhawala H, Chen X, Fisher AJ. 2006. Cytidine

40. Chai S, Choe J. 2011. Crystal structure of elongation factor P from

45. Lai L, Henrissat B, Davies GJ, Witters SG. 2008. Glycosyltransferases:

41. Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases:

42. Kasz P, Wemmer ED, Johnson FE, Ferrin TE. 2004. UCSF Chimera—a visualization system for explor-
Krafczyk et al.

87. Schwarzinger S, Kroon GJ, Foss TR, Chung J, Wright PE, Dyson HJ. 2001. Dependence 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 anomalous data. Acta Crystallogr D Biol Crystallogr 59:3047–3052. https://doi.org/10.1107/S090744491301639X.

95. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, 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.

96. 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.

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

98. Adams PD, Brünger AT, Connolly MDL, deLange WM, Grzesiek S, Hopfinger E, Karplus K, Kossiakoff AA, McCammon JA, Murzin AG, Nunn JN, Pfeiffer J, Pannu NS, Poon LL, 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. Brünger 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, Gos 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, Comolli M, 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/btl272.

106. Vranken WF, Boucher W, Stevens TJ, Foqh 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.