🔓 **OPEN ACCESS**

Substrate recognition by a peptide-aminoacyl-tRNA ligase

Josseline Ramos-Figueroa[a,b,1] 🔵, Haoqian Liang[c,1] 🔵, and Wilfred A. van der Donk[a,b,c,2] 🔵

The continuing discovery of new peptide-aminoacyl-tRNA ligases (PEARLs) has unveiled a diverse array of enzymes with the unique potential to append amino acids to the C terminus of substrate peptides in an aminoacyl-tRNA-dependent manner. To date, PEARLs have been reported that can conjugate Cys, Ala, Trp, Gly, Leu, Asn, and Thr residues, but the basis of peptide substrate and aminoacyl-tRNA recognition is not known. Cell-free expression (CFE) has emerged as a powerful tool to rapidly assay activity of substrate variants, and we used the technique in this study to investigate the peptide substrate specificity of the PEARL $BhaB_C^{Trp}$. This enzyme that adds Trp was discovered previously during genome mining for ribosomally synthesized and post-translational modified peptides (RiPPs). The enzyme is remarkably tolerant of changes to the C-terminal amino acid of the peptide substrate, and truncation and replacement experiments suggest a minimal sequence requirement. An AlphaFold3 model provided insights into binding interactions of the substrate peptide BhaA-Ala to $BhaB_C^{Trp}$ and also generated predictions for tRNA, ATP, and Mg^{2+} binding modes that were tested by site-directed mutagenesis. The data suggest that several highly conserved residues in PEARLs recognize the 3′-CCA sequence present in all tRNAs. The minimal sequence required for Trp incorporation by $BhaB_C^{Trp}$ was employed as a protein tag for C-terminal labeling of eGFP, lysozyme, and MBP with Trp and 5-Br-Trp.

peptide bond | amino acyl-tRNA | conjugation

Techniques that enable the labeling of proteins have extensive applications in biomolecular research (1–7). The integration of noncanonical amino acids (ncAAs) provides proteins with distinctive bioorthogonal handles for conjugation with exogenous molecules, facilitating investigation of biological processes in vitro and in living organisms (8–10). To date, the introduction of ncAAs into protein is mostly achieved through orthogonal translation systems, in which orthogonal tRNA and aminoacyl-tRNA synthetase pairs are engineered or recoded to incorporate ncAAs at specific codons in the mRNA by the ribosome (11–17). Despite the power of this method, incompatibility of orthogonal components with the ribosome can lead to miscoding and decrease in incorporation efficiency (18). Nature offers potential alternative strategies using the diverse array of enzymes discovered in the biosynthetic pathways of peptide-derived natural products. For instance, a recently discovered class of enzymes enables amino acid appendage to peptides independent of the ribosome, making them promising candidates for ncAA incorporation through further bioengineering. These enzymes, termed peptide aminoacyl tRNA ligases (PEARLs), use amino acyl-tRNA as amino acid donor and catalyze the conjugation of the charged amino acid to the C terminus of a substrate precursor peptide in an ATP-dependent and template-independent manner (19–22). During the reaction, PEARLs utilize ATP to first phosphorylate the C-terminal carboxylate of their peptide substrates. Next, the activated carboxyl group is condensed with the amino group of aminoacyl-tRNA, forming a new amide bond (Fig. 1A) (19, 20). Hydrolysis of the tRNA then provides a peptide that has been extended by one amino acid. A collection of different PEARLs have been demonstrated to incorporate a wide range of amino acids into their substrates (19–21, 23, 24).

PEARLs are evolutionarily related to class I lanthipeptide dehydratases (LanBs) (25), enzymes that glutamylate the side chains of Ser and Thr residues in their substrate peptides in a glutamyl-tRNA-dependent mechanism. The glutamate is then eliminated by the LanB C-terminal elimination domain, resulting in the formation of dehydroamino acids (26). In some cases, including enzymes involved in the biosynthesis of thiopeptides and pyritides (27–29), the glutamylation and elimination domains are two separate proteins, resulting in the distinction between $LanB_A$ proteins (single enzymes) and $LanB_B$ proteins (separate glutamylation and elimination proteins) (21). PEARLs typically only contain the domain that resembles the LanB glutamylation domain (20, 26) and lack the

Significance

Peptide synthesis in nature is mostly accomplished by the ribosome or nonribosomal synthetases. Recently, a class of enzymes was discovered termed peptide aminoacyl-tRNA ligases (PEARLs) that use aminoacyl-tRNA to attach the amino acid to the C terminus of peptides. Given the importance of amide bond formation for the preparation of natural and non-natural peptides, this study investigated the substrate specificity of a representative PEARL. The enzyme is tolerant of a wide range of amino acids at the C terminus of the peptide substrate, providing avenues for engineering the enzyme to prepare designer molecules. Identification of a minimal substrate sequence and appendage of this sequence to various proteins allowed the conjugation of tryptophan and 5-bromo-tryptophan to their C termini.

Author affiliations: [a]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801; [b]HHMI, University of Illinois at Urbana-Champaign, Urbana, IL 61801; and [c]Department of Biochemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801

[1]J.R.-F. and H.L. contributed equally to this work.

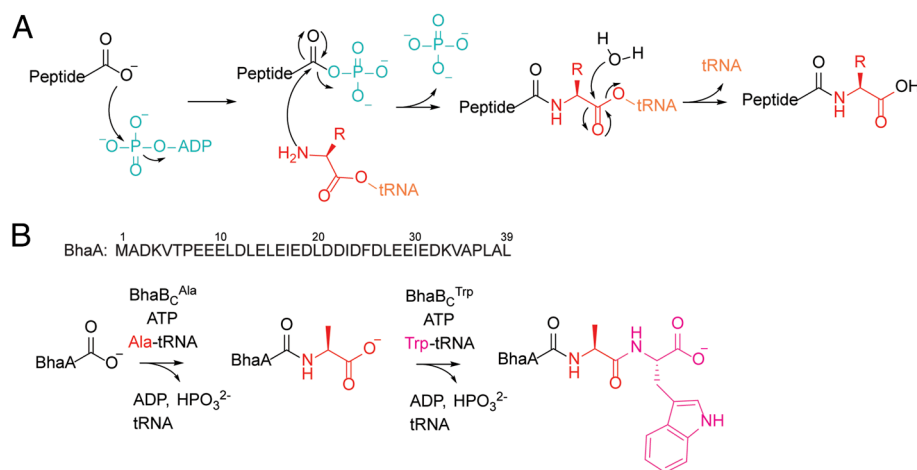[2]To whom correspondence may be addressed. Email: vddonk@illinois.edu.

Fig. 1. (*A*) Mechanism of nonribosomal peptide extension by PEARLs. (*B*) BhaA sequence and the cascade reactions catalyzed by $BhaB_C^{Ala}$ and $BhaB_C^{Trp}$ on BhaA.

elimination domain; these enzymes have been designated $LanB_C^{Xxx}$, with Xxx indicating which amino acid is appended to the substrate peptide from amino acyl-tRNA. A major difference between PEARLs and $LanB_{A/B}$ enzymes is that the former require ATP whereas the latter do not. A recent bioinformatic analysis (23) provided support for a model in which PEARLs evolved from ATP-GRASP proteins and that $LanB_{A/B}$ proteins further evolved and lost the ATP requirement. At present, no information is available for PEARLs or LanBs regarding the binding site of the aminoacyl-tRNA nor is it known where the active site for amide bond formation is located.

Despite the diverse bacterial origins of PEARL-encoding biosynthetic gene clusters (BGCs), thus far, all identified PEARLs from different phyla have exhibited activity with aminoacylated tRNA substrates in *Escherichia coli*, suggesting a relaxed specificity regarding the tRNA sequence. This flexibility also illustrates their potential compatibility with orthogonal translation components and positions them as promising candidates for introducing ncAAs into target peptides or proteins. Furthermore, the process of incorporating an amino acid at the end of a growing peptide and then exposing a new terminus for subsequent coupling to the next amino acid (Fig. 1*B*) has conceptual similarities to solid phase peptide synthesis (SPPS). Given the recent rise of peptides as a drug modality and the high chemical waste associated with SPPS (30), enzymatic means of synthesizing peptides in a nontemplated manner hold significant potential. Effective application of PEARLs for amino acid appendage to non-native peptides or proteins requires a comprehensive understanding of substrate recognition. Thus, in this study, we set out to investigate the mechanism of substrate recognition by a representative PEARL enzyme through site-directed mutagenesis of the substrate peptide.

Cell-free expression (CFE) is emerging as a powerful technique to rapidly probe the importance of specific residues in substrate recognition (31). In CFE, proteins are synthesized from input DNA templates using cell lysates or purified components that contain the necessary cellular machinery for transcription and translation (*SI Appendix,* Fig. S1) (31–33). This approach offers several advantages, including speed, absence of nucleases and proteases that potentially lead to low expression yield, and the ability to synthesize proteins (or RiPP products) that may be toxic or difficult to express in living cells (34–36). Indeed, CFE has recently been applied in several RiPP biosynthetic investigations (29, 37–41). For instance, CFE-based activity assessment of the

enzymes involved in the biosynthesis of thiopeptides, pyritides, and lasso peptides comprehensively investigated their substrate tolerance within a relatively short timeframe (29, 38, 39, 42–45), and discovery of novel lantibiotics was achieved through CFE-assisted development of libraries of nisin and lacticin 481 analogues (37, 41).

In this study, we utilized CFE methodology to investigate the substrate recognition mechanism of $BhaB_C^{Trp}$, a PEARL that adds Trp from tryptophanyl-tRNA to the C terminus of its substrate peptide (Fig. 1*B*) (21). Trp is the most rare amino acid in proteins and can function as a handle for site-selective modification (46, 47). $BhaB_C^{Trp}$ is encoded in a BGC from *Bacillus halodurans* C-125, *bha*, believed to produce a pyrroloquinoline-like natural product given its similarity to the ammosamide BGC (21). In the first biosynthetic step, the PEARL $BhaB_C^{Ala}$ adds Ala to the C terminus of the precursor peptide BhaA forming BhaA-Ala, then $BhaB_C^{Trp}$ installs a Trp to the C terminus of BhaA-Ala producing BhaA-Ala-Trp (Fig. 1*B*). The CFE experiments in this study demonstrate that $BhaB_C^{Trp}$ displays considerable tolerance toward variations in the BhaA-Ala sequence including substitution of the C-terminal residue, Ala40 (Fig. 1*B*), with polar, nonpolar, bulky, and small amino acids. Using AlphaFold3 (48), we generated a model of the quaternary complex of $BhaB_C^{Trp}$, $tRNA^{Trp}$, BhaA-Ala, and ATP providing a three-dimensional perspective of peptide substrate and tRNA engagement by the enzyme. We identified residues in $BhaB_C^{Trp}$ that were predicted to interact with the 3′ CCA sequence and the acceptor stem of tRNA and provide support for the model through mutagenesis studies. Furthermore, by evaluating the activity of $BhaB_C^{Trp}$ toward truncated BhaA substrates and the AlphaFold3 model, a minimal sequence was obtained for Trp incorporation that was used as a tag at the C termini of eGFP, lysozyme, and MBP and that also allowed incorporation of 5-Br-Trp at the C terminus of these proteins.

Results and Discussion

$BhaB_C^{Trp}$ Fully Modifies BhaA-Ala Generated by CFE.
In most previous studies that used CFE to investigate the substrate specificity of RiPP biosynthetic enzymes, variants of the substrate peptide were produced by CFE, and then, purified enzyme was added. In only a few cases, the enzymes themselves were also

produced by CFE (37, 39, 41). To establish proof-of-concept for the feasibility of using CFE to probe the substrate requirement of $BhaB_C^{Trp}$, expression of both BhaA-Ala and $BhaB_C^{Trp}$ was evaluated using the PURExpress® in vitro protein synthesis system (34). Pairwise CFE reactions were set up containing plasmid encoding hexahistidine (His_6) tagged BhaA-Ala or plasmids encoding both the peptide substrate and $BhaB_C^{Trp}$. The reactions were monitored by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (MS). Additionally, both reaction mixtures were analyzed by SDS-PAGE revealing a unique band corresponding to $BhaB_C^{Trp}$ (*SI Appendix,* Fig. S2). A peak corresponding to the mass of His_6-tagged BhaA-Ala was detected in the MALDI-TOF mass spectrum when substrate was expressed alone, and a shift in mass consistent with the condensation of a Trp was observed when enzyme and substrate were coexpressed, indicating $BhaB_C^{Trp}$ produced by CFE transformed the substrate BhaA-Ala into a product bearing a Trp appendage (Fig. 2*A*).

A C-Terminal 12-mer of BhaA-Ala Is Sufficient for Modification by $BhaB_C^{Trp}$. The activity of $BhaB_C^{Trp}$ in the CFE experiments prompted a detailed study into the features in the substrate peptide that are important for processing by the enzyme. To investigate the minimal requirement of the substrate length for $BhaB_C^{Trp}$, two shortened variants of BhaA-Ala were designed (Fig. 2 *B*, panel *i*). DNA encoding these peptide variants were acquired as double-stranded DNA fragments containing a T7 promoter, ribosome binding site (RBS), and each corresponding open reading frame (ORF) linked by spacers (*SI Appendix,* Fig. S3) (29). These DNA fragments were then combined with the plasmid encoding $BhaB_C^{Trp}$ for expression using PURExpress®. The assay mixtures were desalted and analyzed by MALDI-TOF MS. $BhaB_C^{Trp}$ catalyzed the extension of an 18-mer peptide truncant spanning Ile23 to Ala40. Conjugation of Trp to the C terminus of a C-terminal 12-mer peptide (from Glu29 to Ala40) was also observed albeit to a much lesser extent as judged by the low intensity signal of the +186 Da peak with respect to the mutant substrate peak (*SI Appendix,* Fig. S4). These results indicated that the residues from the N-terminus to Glu28 are not critical for the enzyme–substrate interaction, but that residues in the region from Ile23 to Glu28 are important for optimal processing by $BhaB_C^{Trp}$.

Ala Scanning to Investigate Substrate Specificity. To identify essential residues in BhaA-Ala for recognition by $BhaB_C^{Trp}$, blockwise Ala scanning of the C-terminal region of the peptide was performed (panel *ii*, Fig. 2*B*). Consistent with the previous sequence deletion data where residues Ile23 to Glu28 were not critical for activity, $BhaB_C^{Trp}$ added Trp to a variant where Phe25 to Glu28 (sequence FDLE) were replaced by Ala residues (Fig. 2*B* and *SI Appendix,* Fig. S5 and Table S2). $BhaB_C^{Trp}$ fully converted this mutant to the corresponding product, whereas the deletion mutant where these residues were completely removed was only partially processed. This result suggests that $BhaB_C^{Trp}$ is sensitive to the length of the peptide substrate, possibly for the peptide to attain secondary structure. In a similar manner, $BhaB_C^{Trp}$ also added Trp to variants in which a subset of four sequential amino acids from Leu27 to Val34 were substituted with Ala residues

(Fig. 2, panel *ii*, and *SI Appendix,* Fig. S5). However, when the C-terminal PLALA sequence was changed to PAAAA, $BhaB_C^{Trp}$ was not able to process the peptide variant. Previous bioinformatic analysis showed that the C-terminal PLALA sequence of BhaA-Ala is highly conserved among several PEARL substrate peptides from different phyla (19, 21). $BhaB_C^{Trp}$ was also unable to conjugate Trp to the end of a peptide in which the sequence preceding the LALA motif (Lys33 to Pro36) was substituted by Ala residues (Fig. 2*B* and *SI Appendix,* Fig. S5 and Table S2). Similarly, deletion of the EDKVA sequence abolished activity (Fig. 2*B*). Collectively, the deletion and Ala-block scanning experiments indicated that $BhaB_C^{Trp}$ is tolerant of changes up to roughly Pro36, but that substitutions beyond that point appear to disrupt activity.

We next attempted to determine whether specific residues in the C-terminal five residues are critical for activity. We first focused on the two Leu residues (Leu37 and Leu39, panel *iii*, Fig. 2*B*). Starting from the inactive PAAAA variant, reintroduction of either Leu37 or Leu 39 was sufficient to restore Trp conjugation activity (*SI Appendix,* Fig. S6). Activity was not specific to Leu residues as $BhaB_C^{Trp}$ was also active when Phe substitutions were introduced at positions 37 and 39 (Fig. 2*B*). Furthermore, Ala38 was not a critical residue as substitution with Phe still resulted in a substrate that was processed by $BhaB_C^{Trp}$. We also explored the importance of Pro36 for activity and found that replacement with Ala resulted in a substrate for $BhaB_C^{Trp}$. These findings show that the enzyme is quite forgiving in terms of individual substitutions and suggest that it may be a secondary structural feature that is recognized. A discussion based on AlphaFold3 models is provided in a later section of this study.

$BhaB_C^{Trp}$ Is Tolerant of Changes in the C-Terminal Residue of BhaA-Ala. The C-terminal Ala of the LALA motif was investigated next with a series of single-site variants (panel *iv*, Fig. 2*B*). $BhaB_C^{Trp}$ conjugated Trp to all investigated variants in which Ala40 was substituted with a wide array of amino acids, including polar or charged residues (Ser, Lys, and Glu) and hydrophobic and aromatic residues (Leu, Phe, Trp, and Tyr) (Fig. 2*B* and *SI Appendix,* Fig. S7 and Table S2). $BhaB_C^{Trp}$ also added Trp to the carboxyl group of Pro40 and Gly40, amino acids with distinct conformational preferences. Of these variants, A40P and A40W were only partially processed by the enzyme. The relatively low specificity for the C-terminal residue suggests that the size, polarity, or charge of the side chain of the C-terminal residue is not critical for recognition and again suggests that secondary structure and/or substrate length may be important.

Last, we made extended substrates in which Ala blocks were inserted before Ala35 (BhaA-Ala extend-1), Lys33 (BhaA-Ala extend-2), and Glu31 (BhaA-Ala extend-3, panel *v*, Fig. 2*B*). When these variants were used in the CFE assays, $BhaB_C^{Trp}$ added Trp to all peptides as shown by MALDI-TOF MS analysis (Fig. 2*B* and *SI Appendix,* Fig. S8 and Table S2).

AlphaFold3 Predictions of the BhaA-Ala:$BhaB_C^{Trp}$:$tRNA^{Trp}$:ATP Quaternary Complex. To rationalize the results of the CFE studies and obtain insights into the potential mechanism of tRNA recognition, we employed the recently launched AlphaFold3 program with its added feature of ligand modeling including RNA (48, 49). In addition to BhaA-Ala, we included ATP, Mg^{2+} ions, and $tRNA^{Trp}$ from *E. coli,* but without posttranscriptional
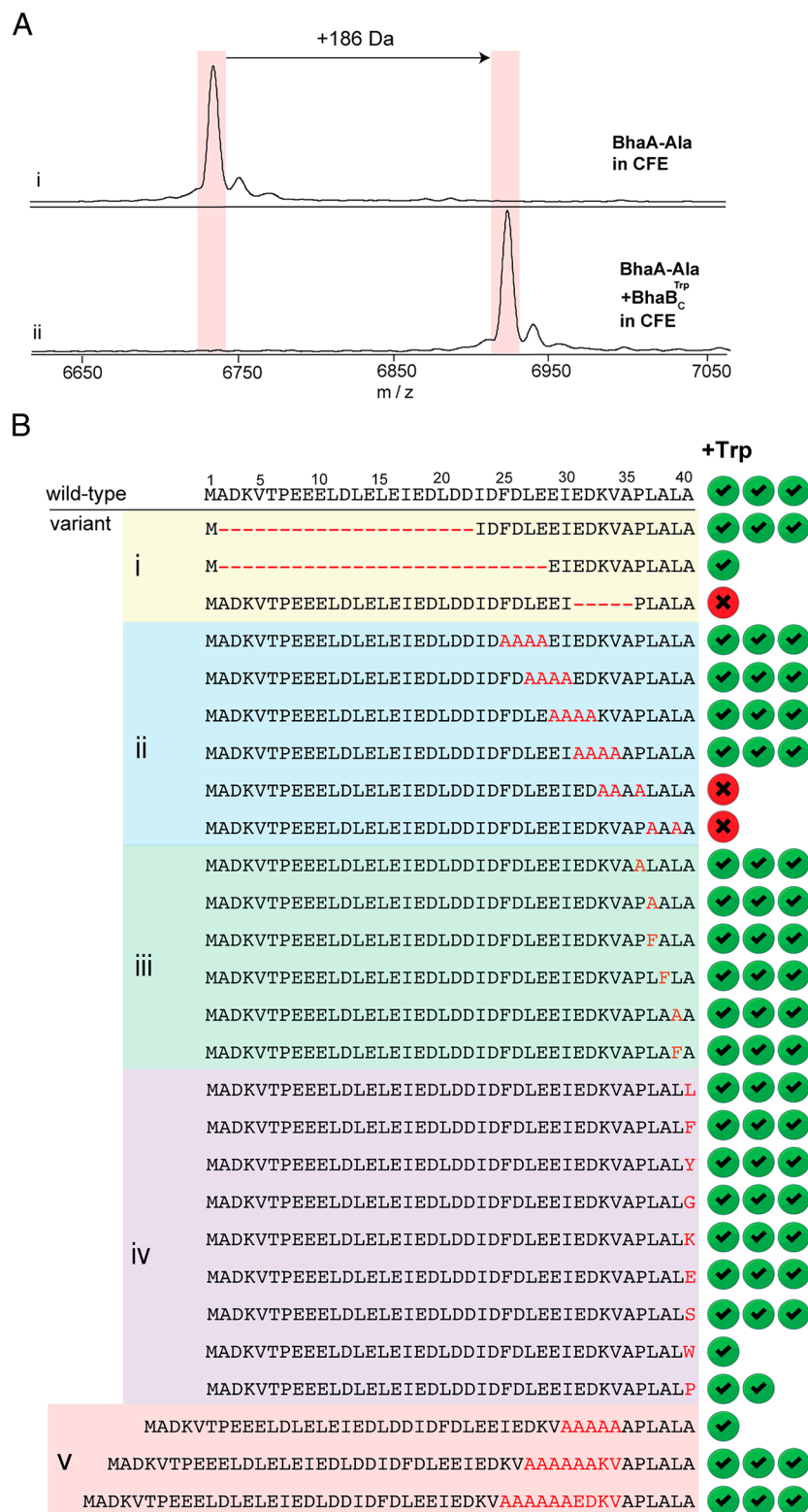
Fig. 2. Activity and substrate specificity assessment of $\text{BhaB}_\text{C}^\text{Trp}$ made by CFE. MALDI-TOF MS analysis of (*A*) CFE of (*i*) His$_6$–BhaA-Ala. [fMet + M + H]$^+$ calcd. *m/z* = 6733.2, obsd. *m/z* = 6734.7. (*ii*) CFE of substrate along with $\text{BhaB}_\text{C}^\text{Trp}$. (fMet + M + H + Trp − H$_2$O)$^+$ calcd. *m/z* = 6919.3, obsd. *m/z* = 6921.5. (*B*) Summary of $\text{BhaB}_\text{C}^\text{Trp}$ activity toward Bha-Ala variants including (*i*) truncated substrates, (*ii*) block-wise Ala replacements, (*iii* and *iv*) single-site variants, and (*v*) sequence extensions to evaluate changes in register of the C-terminal PLALA sequence. Three check marks indicate >90% conversion, two check marks ~50% conversion, and one check mark indicates product is formed but more than 50% of the starting peptide remained in at least two independent duplicate experiments under identical conditions (*SI Appendix*, Figs. S3–S8). A cross indicates no conversion.

modifications since these are not currently supported. AlphaFold-based algorithms predict the structure of wild-type (WT) proteins based on covariance of amino acid sequences. Assessing the

confidence in AlphaFold3-based predicted complexes involves analysis of two parameters, the predicted local distance difference test (pLDDT) and the predicted aligned error (PAE). Compared

to earlier versions that determined a pLDDT value as a per-residue measure of local confidence, AlphaFold3 predicts these values for each atom in the predicted structure. Regions with pLDDT values larger than 90 are regarded as high-confidence predictions that can be used for binding analysis including the side chains of residues. pLDDT values higher than 70 but lower than 90 represent good confidence suggesting correct backbone prediction. pLDDT values lower than 70 are not reliable (50). In addition, PAE values represent the level of confidence on the relative position of two amino acid residues within the predicted structure. When these residues are present in two different domains or objects of a complex, high PAE values indicate that the relative position between these two amino acids cannot be predicted confidently. In our evaluation of the predicted quaternary structure, we used both pLDDT and PAE values. We will show only the top model in the discussion below, but the top five structures were highly similar (*SI Appendix,* Fig. S9).

The binding of tRNA to LanB and PEARL enzymes has remained elusive as no structures have been reported. It was therefore striking that AlphaFold3 places the 3′-end of the tRNA^{Trp} molecule right into the putative active site of BhaB$_C^{Trp}$ based on previous mutagenesis studies on the Cys-adding PEARL TglB (Fig. 3*A*) (20). A recent bioinformatic analysis of PEARLs concluded that they contain an ATP-GRASP-like ATP binding site that is involved in activation of the C-terminal carboxylate group of its peptide substrate by phosphorylation (23). In line with this prediction, AlphaFold3 positioned the ATP together with two Mg^{2+} ions in the postulated ATP-GRASP binding site making several interactions with BhaB$_C^{Trp}$ residues (Glu813, Arg815, and Asp558) that have already been shown experimentally to be important for the phosphorylation step catalyzed by TglB (Glu801, Arg803, and Glu542) (20, 23). Finally, the model also placed the C-terminal carboxylate of the peptide substrate right next to the γ-phosphate of ATP (Fig. 3*B*). Hence, in terms of the chemistry that BhaB$_C^{Trp}$ catalyzes, the predicted quaternary complex is primed for catalytic activity.

The confidence in the predicted complex structure was assessed using the pLDDT (*SI Appendix,* Fig. S10) and PAE values (*SI Appendix,* Fig. S11). The quaternary complex is shown in Fig. 3*C* colored by pLDDT, illustrating good to high confidence

in the model in key areas of interaction of the enzyme with both the peptide substrate and certain parts of the tRNA substrate. Individual pLDDT plots for enzyme, tRNA, and peptide substrate in the top five models are shown in *SI Appendix,* Fig. S10. For all five models, most of the tertiary structure of BhaB$_C^{Trp}$ is predicted with good to high confidence (pLDDT > 80), with low confidence regions (50 < pLDDT < 70) for the first N-terminal 23 amino acids, and for residues at positions 78 to 87 and 105 to 116. PAE plots are shown in *SI Appendix,* Fig. S11; as mentioned, the PAE values measure the relative confidence between any two residues at any part of the complex. ChimeraX (51) was used to visually evaluate the confidence of interactions at the interface between the individual components of the quaternary complex involving BhaB$_C^{Trp}$, BhaA-Ala, ATP, and tRNA^{Trp} (*SI Appendix,* Fig. S11*C*). Low PAE values (<5 Å) were observed for interactions of the enzyme with the C-terminal helix of BhaA-Ala and with ATP, with slightly larger PAE values for the interactions between BhaB$_C^{Trp}$ and the 3′ CCA sequence of tRNA^{Trp}. Significantly larger PAE values were observed for putative interactions with the anti-codon region. Below we will discuss in turn the predicted interactions with ATP, tRNA, and the peptide substrate.

AlphaFold3 Predictions of the Interaction of BhaB$_C^{Trp}$ with ATP. All enzyme residues embedded in the ATP-GRASP-like fold were predicted with high confidence (pLDDT values > 90), and as such, the side chain residue interactions with ATP within this fold are further discussed. In the model, Glu813 interacts through hydrogen bonds with the 2′ and 3′ hydroxyl groups of the ribose moiety, and with one of the two Mg^{2+} ions (Fig. 3*B*) that coordinates to the β-phosphate oxygens. In addition, Asp558 coordinates to the same Mg^{2+} ion and interacts with Arg815, which in turn engages the γ phosphate oxygens of ATP and the C-terminal carboxylate of BhaA-Ala. A second Mg^{2+} ion is predicted to be coordinated by Glu575 and possibly the α, β, and γ phosphates of ATP. As noted above, several of the corresponding residues of the PEARL TglB were previously shown to be important for phosphorylation of its peptide substrate (20), providing further support for the predicted binding model.

The side chains of Arg715 and Phe755 are in close proximity to the adenine group of ATP with Arg715 likely interacting with
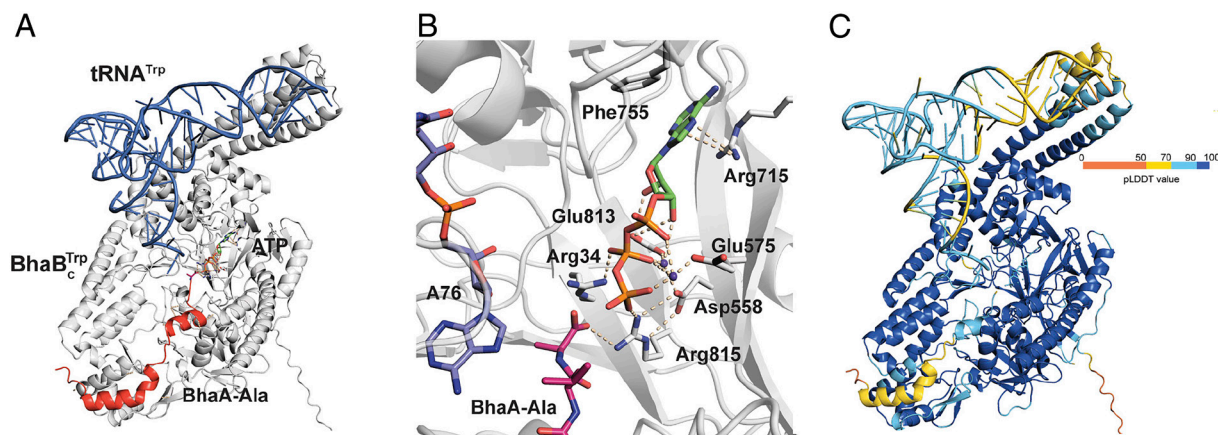

Fig. 3. (*A*) General view of the quaternary complex comprising BhaB$_C^{Trp}$:tRNA^{Trp}:BhaA-Ala:ATP predicted by AlphaFold3. The enzyme is in gray, tRNA in blue, and substrate peptide in red. (*B*) Predicted ATP binding to the postulated ATP-GRASP domain in BhaB$_C^{Trp}$. Key interactions are shown with wheat-colored dashes including predicted polar interactions between the BhaA-Ala C-terminal carboxylate and Arg815 and Arg34 of BhaB$_C^{Trp}$. BhaA-Ala residues are shown as pink sticks and the carbon atoms of the 3′ adenosine (A76) are shown in marine blue sticks. (*C*) Quaternary complex shown in a pLDDT-based coloring scheme. Dark blue corresponds to pLDDT values higher than 90, light blue to pLDDT scores higher than 70 but less than 90, yellow to pLDDT values less than 70 but higher than 50, and low pLDDT scores (<50) are shown in orange.

the adenine moiety through a cation–pi interaction. This interaction is reminiscent of a previously observed interaction in the cocrystal structure of TbtB, a glutamyl-tRNA-dependent $LanB_B$ enzyme, coordinated to an AMP analog that carried a glutamate on the 2′-amino group (52). TbtB catalyzes a different reaction from that catalyzed by PEARLs and does not utilize ATP as substrate. However, the X-ray structure displays a very similar interaction of the adenine moiety of the glutamyl-tRNA analog with an Arg and Phe as the interaction observed in the AlphaFold3 model of $BhaB_C^{Trp}$ involving Arg715 and Phe755 and the adenine of ATP.

AlphaFold3 Predictions of the Interaction of $BhaB_C^{Trp}$ with $tRNA^{Trp}$. The $tRNA^{Trp}$ molecule in all models is overall of lower confidence but still mostly with pLDDT values >60, with higher pLDDT scores (80 to 90) for the T- and D-arm regions (base positions 43 to 65) and the 3′-CCA end (base positions 73 to 76) (Fig. 3C and *SI Appendix,* Fig. S10C). Lower pLDDT values were consistently observed for bases in the anticodon arm and loop (bases 25 to 42). The predicted structure revealed potential interactions between $BhaB_C^{Trp}$ and the $tRNA^{Trp}$ structure, with these interactions localized at positively charged patches of the enzyme including the highly conserved Arg208, Tyr205, Gln201, and Arg197 residues (pLDDT > 90) residing in a helix comprising residues 193–211, as well as Arg429 in a loop made up of residues 415–432 (70 < pLDDT < 90). Arg208 is predicted to interact with the phosphate ester bond between nucleotides C75 and A76 of the terminal 3′-CCA sequence (70 < pLDDT < 90) in addition to the 2′ hydroxyl groups of both C75 and A76 (Fig. 4A). In support of an important role of Arg208 for tRNA binding, mutation of the corresponding Arg residue in a previous study of TglB resulted in complete abolishment of PEARL activity without impairing ATP-dependent activation of the C-terminal carboxylate of its substrate TglA (20). Since all tRNAs end in CCA at the 3′ end, the model explains the complete conservation of Arg208 in all experimentally characterized PEARLs even though they add a wide variety of amino acids. In the model, Tyr205 interacts with the pyranose oxygen of the ribose ring of C75, as well as a phosphate ester oxygen and the 2′ hydroxyl of C74. Gln201 is predicted to hydrogen bond with N3 of C74, and Arg197 to interact with N2 of the discriminator base G73 and with O4 of

U72 (Fig. 4A). Finally, Arg429 is located in a proximal loop and potentially interacts with oxygens of two phosphate ester bonds of the 3′ CCA motif; however, because of the lower pLDDT values of this region, the side chain positioning is of lower confidence. Intriguingly, the anticodon loop of the bound $tRNA^{Trp}$ appears to engage through polar interactions with a helix bundle of $BhaB_C^{Trp}$ that was previously noted as a unique part of $LanB_B$ and PEARL enzymes (52) (*SI Appendix,* Fig. S12). This type of predicted interaction was also observed for AlphaFold3 models of $BhaB_C^{Ala}$ and TglB with their cognate tRNAs (*SI Appendix,* Fig. S12) suggesting it may be a general feature of PEARLs. However, the pLDDT and PAE values for enzyme and tRNA in this region are of low confidence (*SI Appendix,* Fig. S11C) and this potential interaction will require investigation using other tools such as the Flexizyme technology that can change the tRNA sequence without affecting charging with a desired amino acid (53).

We probed the predictions of the AlphaFold3 model by disruption of putative interactions between the enzyme and the 3′-CCA sequence of tRNA. We made R208A, Y205A, and R429A mutants of $BhaB_C^{Trp}$ and assessed their activities in the CFE assay. All three mutants showed severe decreases in the activity of $BhaB_C^{Trp}$ (*SI Appendix,* Fig. S13) indicating that despite less confident pLDDT values, the predicted interactions mediated by Arg208, Tyr205, and Arg429 appear important for catalysis. The model leaves unresolved how the amino acid attached to the tRNA may be recognized by the enzyme. A comparison of the AlphaFold3 models for the PEARLs TglB (adding Cys) and $BhaB_C^{Trp}$ shows that the area where the amino acid is attached is surrounded by flexible loops that are likely to engage with the amino acid (*SI Appendix,* Fig. S14).

AlphaFold3 Predictions of the Interaction of $BhaB_C^{Trp}$ with Bha-Ala. Finally, we discuss the binding of BhaA-Ala in the quaternary complex obtained with AlphaFold3. The BhaA-Ala secondary structure was predicted with good confidence (70 < pLDDT < 90) for the C-terminal residues from positions 27 to 39 with significantly lower scores toward the N-terminus of the peptide (Fig. 3C and *SI Appendix,* Fig. S10C). As mentioned, AlphaFold3 placed the C terminus of the BhaA-Ala peptide right into the same area as the 3′-end of the $tRNA^{Trp}$ and the
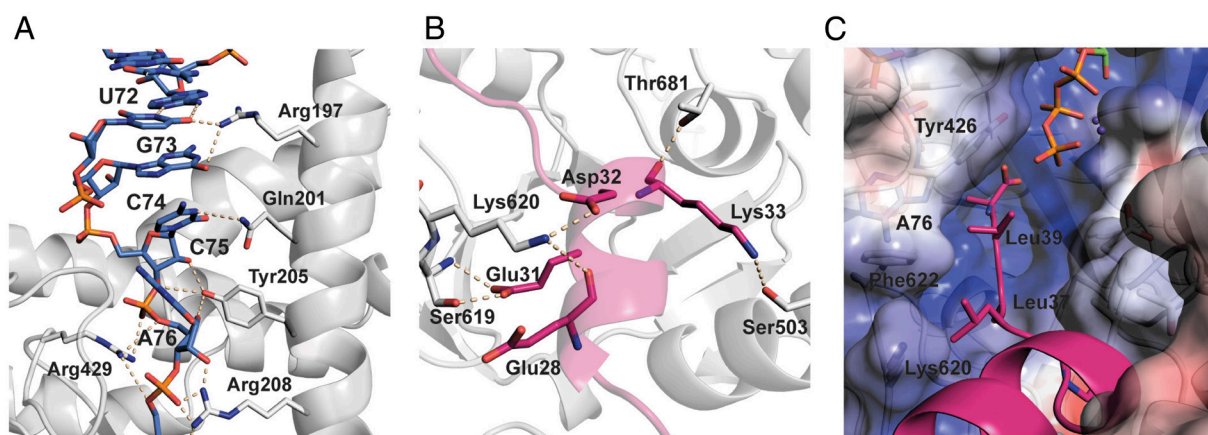

Fig. 4. Predicted interactions between $BhaB_C^{Trp}$ and its peptide and tRNA substrates. (A) Predicted binding interactions of the 3′ CCA end of the $tRNA^{Trp}$ with residues in $BhaB_C^{Trp}$. tRNA bases are shown in blue sticks, and enzyme residues with white gray sticks. Polar interactions are indicated with wheat-colored dashes. (B) Predicted binding mode of BhaA-Ala (pink) and interactions with $BhaB_C^{Trp}$ residues (white gray). (C) Predicted binding mode of the C-terminal PLALA motif of BhaA-Ala (in pink) bound to a groove in $BhaB_C^{Trp}$ that leads to the ATP binding site. Electrostatic potential surface of the enzyme is shown with white patches indicating hydrophobic regions, blue patches positively charged regions, and red patches negatively charged zones.

γ-phosphate of ATP (Fig. 3B). Although this region had lower pLDDT values for residues on $BhaB_C^{Trp}$ (<70) (*SI Appendix,* Fig. S10B), it contains several completely conserved amino acid residues. In the model, Arg815 of $BhaB_C^{Trp}$ makes an electrostatic interaction with the C-terminal carboxylate of BhaA-Ala to likely position it for phosphoryl transfer. In agreement with the model, the corresponding residue in TglB (Arg803) was shown to be important for the phosphorylation of its substrate peptide. Arg34 is also located in the vicinity of the C-terminal carboxylate of BhaA-Ala and the ATP β phosphate, with one of the side chain guanidinium nitrogens positioned 3.2 Å from the C-terminal carboxylate while the adjacent nitrogen is 2.9 Å from the β phosphate oxygen (Fig. 3B). The corresponding residue in TglB (Arg10) was previously identified as a residue important for Cys appendage to its substrate TglA, but as nonessential for the phosphorylation step (20). Hence, the residue may be more important for activation of the amino acid attached to tRNA or the hydrolysis of tRNA. AlphaFold3 currently cannot use amino acyl-tRNA and thus determination of the precise role of Arg34/Arg10 will require structural biology approaches.

In addition to the interactions in the active site, several other interactions are predicted between BhaA-Ala and $BhaB_C^{Trp}$ (Fig. 4B). The predicted conformation of BhaA-Ala features two helical segments, from Glu8 to Asp21 and from Leu27 to Val34. The pLDDT values show a much higher confidence level for the C-terminal helix than for the N-terminal helix (pLDDT < 70, Fig. 3C and *SI Appendix*, Fig. S10B). Furthermore, the deletion mutants in the CFE studies showed that the segment of Bha-Ala that contains the N-terminal helix is dispensable for activity. Therefore, we focus here solely on the predicted interactions of $BhaB_C^{Trp}$ with the C terminus of BhaA-Ala, with a discussion of its N-terminus in the *SI Appendix*. We do note that the predicted conformation of BhaA-Ala does not result in any interactions with the RiPP recognition element (RRE) (54) that is present in $BhaB_C^{Trp}$ (see discussion in *SI Appendix*). Consistent with the possibility that the RRE may not be important for substrate engagement, the segment of the substrate peptide that would have been expected to bind to the RRE (residues 25–28) could be deleted or replaced by four Ala residues without loss of activity, an observation that was also made for another PEARL (TglB) (20).

Several interactions are predicted between $BhaB_C^{Trp}$ and the C-terminal helical segment of BhaA-Ala (Leu27 to Val34), with the Glu31 side chain on BhaA-Ala interacting through a hydrogen bond with Ser619 of the enzyme, the Asp32 side chain and Glu28 main chain carbonyl of the peptide interacting with Lys620 on the enzyme, and the Lys33 side chain of the substrate forming a hydrogen bond with Ser503 of the protein and its main chain carbonyl interacting with the nearby Thr681 side chain of $BhaB_C^{Trp}$ (Fig. 4B). However, based on the CFE experiments, none of these interactions, if predicted correctly, are critical as shown by the Ala-block variants and the Ala-block insertions (Fig. 2B). We also confirmed the nonessentiality of these interactions with a complementary set of mutations of the enzyme. The S619D variant of $BhaB_C^{Trp}$ fully processed the WT substrate in CFE experiments as shown by MALDI-TOF MS and the S503R mutant also did not impair the activity of $BhaB_C^{Trp}$ (*SI Appendix,* Fig. S13).

Based on the AlphaFold3 model, the residues that comprise the critical PLALA C-terminal motif bend away from the preceding helical axis by about 40 degrees and turn toward the active site pocket (Fig. 4C). The isobutyl groups of Leu37 and Leu39 are

pointing out of a groove that guides the terminus of the peptide to the active site (Fig. 4C). This predicted conformation provides an explanation why substitution of these Leu residues with Phe was tolerated by the enzyme, but it does not explain why experimentally the C-terminal PAAAA mutant was not a substrate for $BhaB_C^{Trp}$ (Fig. 2). AlphaFold relies on covariance in natural sequences to predict intra- and intermolecular interactions. As such, it has been shown to not be a good tool to interrogate non-natural mutations (55–58), such as those introduced in our CFE experiments. We speculate that the secondary structure of the PAAAA mutant is perturbed, possibly by extending the helix and making the 40° turn less favorable or by introducing alternative binding interactions of the terminal residues with the enzyme that similarly prevents the C terminus from turning into the active site.

$BhaB_C^{Ala}$ adds an Ala residue to BhaA, a peptide that differs by only one amino acid from the substrate for $BhaB_C^{Trp}$ (Fig. 1B). How the two enzymes (27% sequence identity) differentiate between these two very similar peptides is not understood. We made an AlphaFold3 complex consisting of $BhaB_C^{Ala}$, $tRNA^{Ala}$, BhaA, and ATP to investigate whether these models have enough information to explain the observed selectivity (*SI Appendix,* Fig. S15). The overall structure of the two complexes is very similar with respect to tRNA and ATP binding, but the conformations of the peptide substrates are quite different and pLDDT values are of low confidence for most of the peptide (*SI Appendix,* Fig. S15 *C* and *D*). The models provide no obvious interactions to explain the observed selectivity, possibly because the amino acids attached to the tRNA may change the observed binding poses. Further biochemical studies will be required to better understand this aspect of PEARL catalysis.

The AlphaFold3 model of BhaA-Ala binding to $BhaB_C^{Trp}$ provides a potential explanation for the observed tolerance regarding the C-terminal residue of the peptide substrate. The model suggests that there may be a pocket that is available between a partially disordered helix from Ile580 to His590 and a beta sheet from Asp30 to Ser35 that may accommodate the various side chains of the variants in Fig. 2 *B*, panel *iv* (*SI Appendix,* Fig. S16).

LEEIEDKVAPLALA as a Tag for Trp Appendage to the C Termini of Proteins. The CFE evaluation of substrate specificity showed that the C-terminal 12-mer of BhaA-Ala was a substrate for $BhaB_C^{Trp}$ but with poor conversion (Fig. 2B). The AlphaFold3 model suggested that the 14-mer sequence of BhaA-Ala from Leu27 to Ala40 forms a helical structure (Fig. 4 *B* and *C*). Thus, we tested this 14-mer sequence (LEEIEDKVAPLALA) as a tag to the C terminus of Enhanced Green Fluorescent Protein (eGFP), lysozyme, and maltose-binding protein (MBP) in proof-of-concept experiments to assess whether Trp could be enzymatically added to these proteins. An N-terminal His_6-tag was added and a C-terminal TEV-cleavage site, followed by a SGSSGGSSG linker and the 14-mer BhaA-Ala sequence (e.g., Fig. 5). Each fusion protein was expressed in *E. coli* along with $BhaB_C^{Trp}$. The proteins were purified by metal affinity chromatography, treated with TEV protease, and the C-terminal peptide was analyzed by MALDI-TOF MS. A similar procedure was performed with protein substrates expressed without $BhaB_C^{Trp}$ as a control. For all three tested proteins incorporation of a C-terminal Trp was observed when coexpressed with $BhaB_C^{Trp}$ with some unmodified protein also observed (*SI Appendix,* Figs. S17–S19).
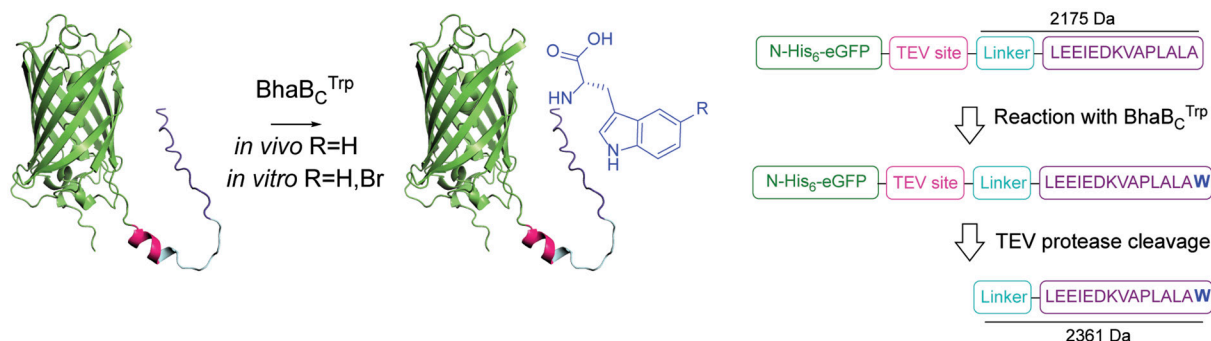
Fig. 5. Representative example of the strategy used to append Trp or 5-Br-Trp to the C terminus of proteins, shown here for eGFP. The TEV site was added for ease of analysis by MALDI-TOF MS in *SI Appendix,* Figs. S17–S19.

We next used purified fusion proteins for in vitro assays with purified $BhaB_C^{Trp}$, *E. coli* TrpRS, and in vitro transcribed tRNA^{Trp} and observed full incorporation of Trp (*SI Appendix,* Figs. S17–S19). Addition of proteinogenic amino acids to the C terminus of proteins has limited utility as the same amino acids could have been encoded on the encoding DNA. PEARLs or future engineered variants have the potential, however, to add noncanonical amino acids, either using substrate-tolerant natural amino acyl-tRNA synthetases or their evolved variants. As an example, we selected L-5-bromo-Trp as a noncanonical amino acid as it has previously been shown to be tolerated by *E. coli* tryptophanyl-tRNA synthetase (TrpRS) (59) and has been used for site-selective conjugation chemistry (47, 60). Indeed, 5-Br-Trp could be incorporated, albeit only to a certain extent as we also observed incorporation of nonbrominated Trp (*SI Appendix,* Fig. S20). This result was puzzling as Trp was not added to the assay and NMR characterization of 5-Br-Trp did not show contamination. Previous crystallographic studies have alluded to the possibility that Trp-AMP can copurify with TrpRS (61), thus we presumed that the observed contaminant might originate from our purified TrpRS. To solve this issue, we next used a procedure where TrpRS was incubated with excess L-5-Br-Trp and ATP to wash out any bound Trp-AMP (*SI Appendix*). Using this TrpRS for in vitro experiments allowed improved incorporation of 5-Br-Trp into the C-terminal modified protein substrates (*SI Appendix,* Figs. S17–S19, *Bottom* panels).

Conclusion and Outlook

By utilizing a cell-free expression system, the peptide substrate requirements of the PEARL $BhaB_C^{Trp}$ were examined in a faster manner than traditional mutagenesis using a heterologous expression host for which the success rates of functional expression are still low (62). The enzyme showed unexpected tolerance to changes to the C-terminal position of the substrate BhaA-Ala and the part of the substrate that was anticipated to interact with the RRE was dispensable for enzyme activity. RREs are ubiquitous in RiPP systems, and in many cases, their interactions with the LP are well documented biochemically or crystallographically (26, 54, 63–67). However, such interactions have not yet been shown for PEARLs, and examples of vestigial RREs have been reported previously (68). In support of the general nonessentiality of the substrate segment that was expected to interact with the RRE, investigation of another PEARL, the Cys-adding TglB from *Pseudomonas syringae*, also showed that the enzyme remained active when only the C-terminal twelve residues of its substrate TglA were present (20). Instead of the expected RRE-interacting segment, the CFE data and the AlphaFold3

prediction of substrate engagement points to the importance of a C-terminal helix and the PLALA sequence that veers the C terminus into the active site. Using the information gleaned from the CFE experiments and AlphaFold3 model, a tag was developed allowing the conjugation of a Trp residue at the C terminus of three model proteins, eGFP, MBP, and lysozyme. These experiments suggest that PEARLs may have utility for protein labeling.

The enhanced abilities of AlphaFold3 to predict binding of cofactors, metal ions, and nucleotides provided a remarkable picture of a potential quaternary complex that appears to be primed for catalysis. Although the tRNA in the model is not charged with Trp and does not contain any posttranscriptional modifications, the C terminus of the peptide is right next to the γ-phosphate of ATP and the 3′-end of the tRNA is in close proximity to the C terminus of the peptide. The model explains previous site-directed mutagenesis results with TglB and was used to design mutants of $BhaB_C^{Trp}$ that were predicted to disrupt the interactions with the 3′-CCA sequence of the tRNA. In agreement with the model, the designed mutants showed strongly diminished activity. The model also suggests intriguing additional interactions with the anticodon loop that can be tested with methods that are beyond the scope of the current study. Importantly, although the model is suggestive of two sequential reactions from a quaternary complex, in the absence of kinetic studies we cannot rule out a ping-pong type mechanism in which ADP leaves the active site before the aminoacyl-tRNA enters (23).

This investigation provides a platform for future studies to test the possibility of engineering PEARLs for non-natural substrates. On the one hand, the substrate tolerance demonstrated in this work is promising. On the other hand, several questions will need to be resolved to facilitate targeted directed evolution approaches toward new activities. These include understanding the details of the sites of recognition of the tRNA sequence, and the still puzzling observation that these enzymes never add more than one residue. While this feature is beneficial in the context of peptide synthesis, it is somewhat surprising given the variety of peptides that were accepted as substrates in this study.

Materials and Methods

Plasmids and PCR constructs were made with the primers listed in *SI Appendix*. CFE assays consisted of 5 μL reactions using the PURExpress® in vitro protein synthesis kit. Each reaction contained 2.5 μL of solution A, 1 μL of solution B, 0.2 μL murine RNase inhibitor, and 1.0 μL of substrate DNA solution (100 ng of template), or 1 μL of 300 to 400 ng/μL enzyme plasmid solution and 0.8 μL of 100 to 200 ng/μL substrate plasmid solution. The reactions were conducted for 5 h at 37 °C in a heating block, desalted, and analyzed by MALDI-TOF MS. Detailed

procedures for CFE, protein purification, and MS are provided in *SI Appendix*. For the protein tagging experiments, pACYCDuet-1 plasmids encoding tagged eGFP, lysozyme, and MBP were cotransformed with a pET28a plasmid encoding BhaB$_C^{Trp}$. After Ni-NTA purification, the eluted His-tagged protein substrates were digested with 1 μg tobacco etch virus (TEV) in 50 mM Tris-HCl pH 8 and 1 mM DTT overnight at 4 °C followed by MS analysis. All primary data for all figures have been deposited at Mendeley data (https://data.mendeley.com/datasets/88tzntsmhs/1) (69).

1. G. T. Hermanson, *Bioconjugate Techniques* (Academic Press, San Diego, CA, 1996).
2. T. Gronemeyer, G. Godin, K. Johnsson, Adding value to fusion proteins through covalent labelling. *Curr. Opin. Biotechnol.* **16**, 453–458 (2005).
3. M. J. Hinner, K. Johnsson, How to obtain labeled proteins and what to do with them. *Curr. Opin. Biotechnol.* **21**, 766–776 (2010).
4. N. Stephanopoulos, M. B. Francis, Choosing an effective protein bioconjugation strategy. *Nat. Chem. Biol.* **7**, 876–884 (2011).
5. K. Lang, J. W. Chin, Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.* **114**, 4764–4806 (2014).
6. A. M. Saleh *et al.*, Non-canonical amino acid labeling in proteomics and biotechnology. *J. Biol. Eng.* **13**, 43 (2019).
7. G. Vandemoortele, S. Eyckerman, K. Gevaert, Pick a tag and explore the functions of your pet protein. *Trends Biotechnol.* **37**, 1078–1090 (2019).
8. J. T. Ngo, D. A. Tirrell, Noncanonical amino acids in the interrogation of cellular protein synthesis. *Acc. Chem. Res.* **44**, 677–685 (2011).
9. R. E. Bird *et al.*, Bioorthogonal chemistry and its applications. *Bioconjug. Chem.* **32**, 2457–2479 (2021).
10. Z. Zou, Y. Ji, U. Schwaneberg, Empowering site-specific bioconjugations in vitro and in vivo: Advances in sortase engineering and sortase-mediated ligation. *Angew. Chem. Int. Ed.* **63**, e202310910 (2024).
11. L. Wang, A. Brock, B. Herberich, P. G. Schultz, Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498–500 (2001).
12. A. J. Link, M. L. Mock, D. A. Tirrell, Non-canonical amino acids in protein engineering. *Curr. Opin. Biotechnol.* **14**, 603–609 (2003).
13. C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182–188 (1989).
14. C. C. Liu, P. G. Schultz, Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
15. D. Cervettini *et al.*, Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase-tRNA pairs. *Nat. Biotechnol.* **38**, 989–999 (2020).
16. H. Hirose, C. Tsiamantas, T. Katoh, H. Suga, In vitro expression of genetically encoded non-standard peptides consisting of exotic amino acid building blocks. *Curr. Opin. Biotechnol.* **58**, 28–36 (2019).
17. C. Soni *et al.*, A translation-independent directed evolution strategy to engineer aminoacyl-tRNA synthetases. *ACS Cent. Sci.* **10**, 1211–1220 (2024).
18. M. Smeenk, J. Agramunt, K. M. Bonger, Recent developments in bioorthogonal chemistry and the orthogonality within. *Curr. Opin. Chem. Biol.* **60**, 79–88 (2021).
19. C. P. Ting *et al.*, Use of a scaffold peptide in the biosynthesis of amino acid-derived natural products. *Science* **365**, 280–284 (2019).
20. Z. Zhang, W. A. van der Donk, Nonribosomal peptide extension by a peptide amino-acyl tRNA ligase. *J. Am. Chem. Soc.* **141**, 19625–19633 (2019).
21. P. N. Daniels *et al.*, A biosynthetic pathway to aromatic amines that uses glycyl-tRNA as nitrogen donor. *Nat. Chem.* **14**, 71–77 (2022).
22. P. N. Daniels, W. A. van der Donk, Substrate specificity of the flavoenzyme BhaC(1) that converts a C-Terminal Trp to a hydroxyquinone. *Biochemistry* **62**, 378–387 (2023).
23. J. Ramos Figueroa, L. Zhu, W. A. van der Donk, Unexpected transformations during pyrroloiminoquinone biosynthesis. *J. Am. Chem. Soc.* **146**, 14235–14245 (2024).
24. Y. Yu, W. A. van der Donk, Biosynthesis of 3-thia-α-amino acids on a carrier peptide. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2205285119 (2022).
25. L. M. Repka, J. R. Chekan, S. K. Nair, W. A. van der Donk, Mechanistic understanding of lanthipeptide biosynthetic enzymes. *Chem. Rev.* **117**, 5457–5520 (2017).
26. M. A. Ortega *et al.*, Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature* **517**, 509–512 (2015).
27. G. A. Hudson, Z. Zhang, J. I. Tietz, D. A. Mitchell, W. A. van der Donk, In vitro biosynthesis of the core scaffold of the thiopeptide thiomuracin. *J. Am. Chem. Soc.* **137**, 16012–16015 (2015).
28. T. Ozaki *et al.*, Insights into the biosynthesis of dehydroalanines in goadsporin. *ChemBioChem* **17**, 218–223 (2016).
29. D. T. Nguyen *et al.*, Accessing diverse pyridine-based macrocyclic peptides by a two-site recognition pathway. *J. Am. Chem. Soc.* **144**, 11263–11269 (2022).
30. A. Isidro-Llobet *et al.*, Sustainability challenges in peptide synthesis and purification: From R&D to production. *J. Org. Chem.* **84**, 4615–4628 (2019).
31. A. D. Silverman, A. S. Karim, M. C. Jewett, Cell-free gene expression: An expanded repertoire of applications. *Nat. Rev. Genet.* **21**, 151–170 (2020).
32. N. E. Gregorio, M. Z. Levine, J. P. Oza, A user's guide to cell-free protein synthesis. *Methods Protoc.* **2**, 24 (2019).
33. D. Garenne *et al.*, Cell-free gene expression. *Nat. Rev. Methods Primers.* **1**, 49 (2021).
34. Y. Shimizu *et al.*, Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–755 (2001).
35. J. P. Chen *et al.*, Improving the soluble expression of difficult-to-express proteins in prokaryotic expression system via protein engineering and synthetic biology strategies. *Metab. Eng.* **78**, 99–114 (2023).
36. K. Khambhati *et al.*, Exploring the potential of cell-free protein synthesis for extending the abilities of biological systems. *Front. Bioeng. Biotechnol.* **7**, 248 (2019).
37. R. Liu *et al.*, A cell-free platform based on nisin biosynthesis for discovering novel lanthipeptides and guiding their overproduction in vivo. *Adv. Sci.* **7**, 2001616 (2020).
38. A. A. Vinogradov *et al.*, Minimal lactazole scaffold for in vitro thiopeptide bioengineering. *Nat. Commun.* **11**, 2272 (2020).
39. Y. Si, A. M. Kretsch, L. M. Daigh, M. J. Burk, D. A. Mitchell, Cell-free biosynthesis to evaluate lasso peptide formation and enzyme-substrate tolerance. *J. Am. Chem. Soc.* **143**, 5917–5927 (2021).
40. M. I. McLaughlin, Y. Yu, W. A. van der Donk, Substrate recognition by the peptidyl-(S)-2-mercaptoglycine synthase TglHI during 3-thiaglutamate biosynthesis. *ACS Chem. Biol.* **17**, 930–940 (2022).
41. W. Q. Liu *et al.*, Cell-free biosynthesis and engineering of ribosomally synthesized lanthipeptides. *Nat. Commun.* **15**, 4336 (2024).
42. A. A. Vinogradov *et al.*, Promiscuous enzymes cooperate at the substrate level en route to lactazole A. *J. Am. Chem. Soc.* **142**, 13886–13897 (2020).
43. J. S. Chang, A. A. Vinogradov, Y. Zhang, Y. Goto, H. Suga, Deep learning-driven library design for the de novo discovery of bioactive thiopeptides. *ACS Cent. Sci.* **9**, 2150–2160 (2023).
44. A. A. Vinogradov, J. S. Chang, H. Onaka, Y. Goto, H. Suga, Accurate models of substrate preferences of post-translational modification enzymes from a combination of mRNA display and deep learning. *ACS Cent. Sci.* **8**, 814–824 (2022).
45. S. E. Barrett *et al.*, Substrate interactions guide cyclase engineering and lasso peptide diversification. *Nat. Chem. Biol.* **21**, 412–419 (2024).
46. X. Xie *et al.*, Oxidative cyclization reagents reveal tryptophan cation-π interactions. *Nature* **627**, 680–687 (2024).
47. N. Saha, F. N. U. Vidya, Y. Luo, W. A. van der Donk, V. Agarwal, Transformation-guided genome mining provides access to brominated lanthipeptides. *Org. Lett.* **27**, 984–988 (2025).
48. J. Abramson *et al.*, Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
49. R. Roy, H. M. Al-Hashimi, AlphaFold3 takes a step toward decoding molecular behavior and biological computation. *Nat. Struct. Mol. Biol.* **31**, 997–1000 (2024).
50. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
51. E. C. Meng *et al.*, UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).
52. I. R. Bothwell *et al.*, Characterization of glutamyl-tRNA-dependent dehydratases using nonreactive substrate mimics. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 17245–17250 (2019).
53. Y. Goto, T. Katoh, H. Suga, Flexizymes for genetic code reprogramming. *Nat. Protoc.* **6**, 779–790 (2011).
54. B. J. Burkhart, G. A. Hudson, K. L. Dunbar, D. A. Mitchell, A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nat. Chem. Biol.* **11**, 564–570 (2015).
55. O. Carugo, pLDDT values in AlphaFold2 protein models are unrelated to globular protein local flexibility. *Crystals* **13**, 1560 (2023).
56. M. A. Pak *et al.*, Using AlphaFold to predict the impact of single mutations on protein stability and function. *PLoS ONE* **18**, e0282689 (2023).
57. G. R. Buel, K. J. Walters, Can AlphaFold2 predict the impact of missense mutations on structure? *Nat. Struct. Mol. Biol.* **29**, 1–2 (2022).
58. S. Chaaban, G. Ratkevičiūtė, C. Lau, AI told you so: Navigating protein structure prediction in the era of machine learning. *Biochemist* **46**, 7–12 (2024).
59. A. Chatterjee, H. Xiao, P. Y. Yang, G. Soundararajan, P. G. Schultz, A tryptophanyl-tRNA synthetase/ tRNA pair for unnatural amino acid mutagenesis in *E. coli. Angew. Chem. Int. Ed. Engl.* **52**, 5106–5109 (2013).
60. N. Montua, P. Thye, P. Hartwig, M. Kühle, N. Sewald, Enzymatic peptide and protein bromination: The BromoTrp tag. *Angew. Chem. Int. Ed.* **63**, e202314961 (2024).
61. M. Xiang *et al.*, An asymmetric structure of bacterial TrpRS supports the half-of-the-sites catalytic mechanism and facilitates antimicrobial screening. *Nucleic Acids Res.* **51**, 4637–4649 (2023).
62. A. E. Kadjo, A. S. Eustáquio, Bacterial natural product discovery by heterologous expression. *J. Ind. Microbiol. Biotechnol.* **50**, kuad044 (2023).
63. J. Koehnke *et al.*, Structural analysis of leader peptide binding enables leader-free cyanobactin processing. *Nat. Chem. Biol.* **11**, 558–563 (2015).
64. T. L. Grove *et al.*, Structural insights into thioether bond formation in the biosynthesis of sactipeptides. *J. Am. Chem. Soc.* **139**, 11734–11744 (2017).
65. J. R. Chekan, C. Ongpipattanakul, S. K. Nair, Steric complementarity directs sequence promiscuous leader binding in RiPP biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 24049–24055 (2019).
66. A. M. Kloosterman, K. E. Shelton, G. P. van Wezel, M. H. Medema, D. A. Mitchell, RRE-finder: A genome-mining tool for class-independent RiPP discovery. *mSystems* **5**, e00267-20 (2020).
67. M. Montalbán-López *et al.*, New developments in RiPP discovery, enzymology and engineering. *Nat. Prod. Rep.* **38**, 130–239 (2021).
68. Z. Zhang *et al.*, Biosynthetic timing and substrate specificity for the thiopeptide thiomuracin. *J. Am. Chem. Soc.* **138**, 15511–15514 (2016).
69. J. Ramos-Figueroa, H. Liang, W. A. van der Donk, Data associated with "Substrate recognition by a peptide-aminoacyl-tRNA ligase". Mendeley Data. https://data.mendeley.com/datasets/88tzntsmhs/1. Deposited 4 March 2025.