

# An efficient pyrrolysyl-tRNA synthetase for economical production of MeHis-containing enzymes†

Amy E. Hutton,<sup>‡</sup> Jake Foster,<sup>‡</sup> James E. J. Sanders, Christopher J. Taylor, Stefan A. Hoffmann,<sup>‡</sup> Yizhi Cai, Sarah L. Lovelock<sup>\*</sup> and Anthony P. Green<sup>\*</sup>

Received 2nd February 2024, Accepted 16th February 2024

DOI: 10.1039/d4fd00019f

Genetic code expansion has emerged as a powerful tool in enzyme design and engineering, providing new insights into sophisticated catalytic mechanisms and enabling the development of enzymes with new catalytic functions. In this regard, the non-canonical histidine analogue  $N_\delta$ -methylhistidine (MeHis) has proven especially versatile due to its ability to serve as a metal coordinating ligand or a catalytic nucleophile with a similar mode of reactivity to small molecule catalysts such as 4-dimethylaminopyridine (DMAP). Here we report the development of a highly efficient aminoacyl tRNA synthetase (G1PyIRS<sup>MIFAF</sup>) for encoding MeHis into proteins, by transplanting five known active site mutations from *Methanomethylophilus alvus* (MaPyIRS) into the single domain PyIRS from *Methanogenic archaeon* ISO4-G1. In contrast to the high concentrations of MeHis (5–10 mM) needed with the Ma system, G1PyIRS<sup>MIFAF</sup> can operate efficiently using MeHis concentrations of ~0.1 mM, allowing more economical production of a range of MeHis-containing enzymes in high titres. Interestingly G1PyIRS<sup>MIFAF</sup> is also a 'polyspecific' aminoacyl tRNA synthetase (aaRS), enabling incorporation of five different non-canonical amino acids (ncAAs) including 3-pyridylalanine and 2-fluorophenylalanine. This study provides an important step towards scalable production of engineered enzymes that contain non-canonical amino acids such as MeHis as key catalytic elements.

## Introduction

In nature proteins perform a vast array of functions, including accelerating biochemical reactions, transporting molecules across membranes, providing structural support, and controlling signalling processes. Recent advances in high-

Manchester Institute of Biotechnology, School of Chemistry, The University of Manchester, Manchester, UK.  
E-mail: Anthony.green@manchester.ac.uk; Sarah.lovelock@manchester.ac.uk

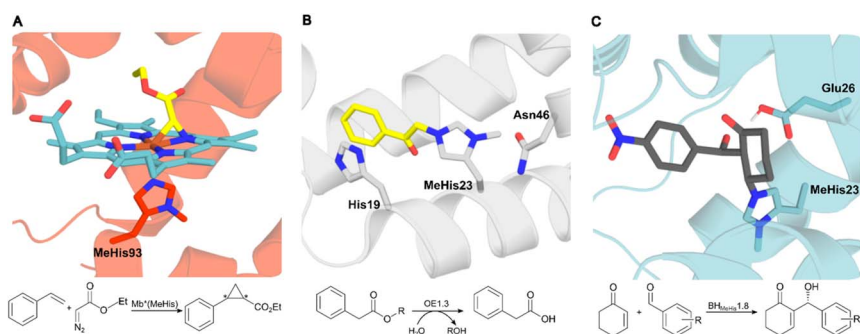
† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d4fd00019f>

‡ These authors contributed equally to this study.



throughput computation and experimentation have given us unprecedented control over protein sequence, structure and function, resulting in the development of a diverse array of engineered protein therapeutics, biocatalysts, and advanced biomaterials.<sup>1–6</sup> At present, our approaches to protein design, engineering and production typically only make use of nature's standard alphabet of twenty canonical amino acid building blocks. These standard amino acids are limited in their chemical diversity, which ultimately restricts our ability to develop proteins with new functions and desirable properties. To address this fundamental limitation, genetic code expansion (GCE) has emerged as a powerful and versatile technology to site selectively install new functional elements into proteins as non-canonical amino acids (ncAAs).<sup>7,8</sup> GCE employs orthogonal aminoacyl tRNA synthetase (aaRS)/tRNA pairs to direct the incorporation of ncAAs in response to a reassigned codon (most commonly the amber codon UAG) introduced into a gene of interest. To date, a variety of aaRS/tRNA pairs have been developed which display the required orthogonality across a range of host organisms. Pyrrolysyl aaRS/<sup>Pyl</sup>tRNA<sub>CUA</sub> pairs from methanogenic archaea have proven especially versatile, having been re-engineered to encode several hundred ncAAs in bacteria, yeast and mammalian cell lines.<sup>9</sup> These systems have allowed the development of new protein therapeutics, precision bioconjugates, responsive materials, protein-based vaccines and new biocontainment strategies.

The availability of an expanded genetic code also opens exciting new opportunities in enzyme design and engineering. For example, GCE has been used to improve enzyme activity and stability,<sup>10,11</sup> to probe complex biological mechanisms,<sup>12–14</sup> and to develop enzymes with functions and modes of catalysis beyond those found in nature.<sup>15–20</sup> The non-canonical histidine analogue, *N*<sub>5</sub>-methylhistidine (MeHis), has proven to be an especially versatile tool in enzyme design and engineering research, leading to catalytically modified enzymes with augmented properties and entirely new functions, as well as new biocontainment



**Fig. 1** Crystal structures of MeHis-containing enzymes. (A) The crystal structure of the myoglobin variant Mb\*(MeHis) (PDB: 6G5B), in which substitution of His93 with MeHis affords an oxygen tolerant carbene transferase.<sup>23</sup> MeHis is shown as atom-coloured sticks, carbon in red, showing the Fe(III) carbenoid complex (atom-coloured stick, carbons in blue and yellow). (B) The designed esterase OE1.3 uses MeHis23 as a catalytic nucleophile.<sup>15</sup> The crystal structure (PDB: 6Q7R) shows MeHis23 (atom-coloured sticks, carbons in grey) alkylated with the mechanistic inhibitor bromoacetophenone (atom-coloured sticks, carbons in yellow). (C) Snapshot from an MD simulation of BH<sub>MeHis</sub>1.8\_Int2 complex. BH<sub>MeHis</sub>1.8 uses MeHis23 as a catalytic nucleophile (PDB: 8BPO, atom-coloured sticks, carbons in blue). Int2 is shown in atom-coloured sticks with carbons in grey.<sup>20</sup>



strategies (Fig. 1).<sup>15,20,21</sup> For example, MeHis has been used as a metal chelating ligand to probe the mechanisms of heme enzymes including ascorbate peroxidase and cytochrome *c* peroxidase.<sup>10,12</sup> MeHis ligands have also been used to augment metalloenzyme function, leading to improvements in both peroxidase and carbene transferase activities in engineered myoglobin variants.<sup>11,22,23</sup> MeHis has also been shown to act as a potent catalytic nucleophile, leading to the development of artificial hydrolases and proficient enzymes for valuable non-biological transformations such as the Morita–Baylis–Hillman reaction.<sup>15,20</sup>

To capitalize on these recent advances, it is important that we are able to produce MeHis-containing enzymes in an efficient and economical manner. At present, the engineered aaRS/tRNA pairs used to encode MeHis are relatively inefficient, typically requiring 5–10 mM concentrations of the expensive ncAA to be supplemented to the culture medium. For context, for a 30 kDa protein produced at 100 mg per litre of culture, this equates to only 0.03–0.06% of MeHis being incorporated into the target protein. Furthermore, even at these high MeHis concentrations, a substantial proportion of undesired truncated protein is typically observed. It is therefore evident that for any future large-scale applications of MeHis-containing enzymes, the efficiency of modified protein production will have to be significantly improved. Here we report an efficient system for producing MeHis-containing proteins in high titres using only low concentrations of ncAA, providing an important step towards scalable production of these modified biocatalysts.

## Results and discussion

MeHis is commonly incorporated into target proteins using engineered PylRS homologs from *Methanosarcina mazei* (*Mm*), *Methanosarcina barkeri* (*Mb*) or *Methanomethylophilus alvus* (*Ma*).<sup>24,25</sup> Two distinct sets of mutations have been reported to confer activity towards MeHis incorporation, either L121M, L125I, Y126F, M129A and V168F or L125I, Y126F, M129G, V168F and Y206F (based on

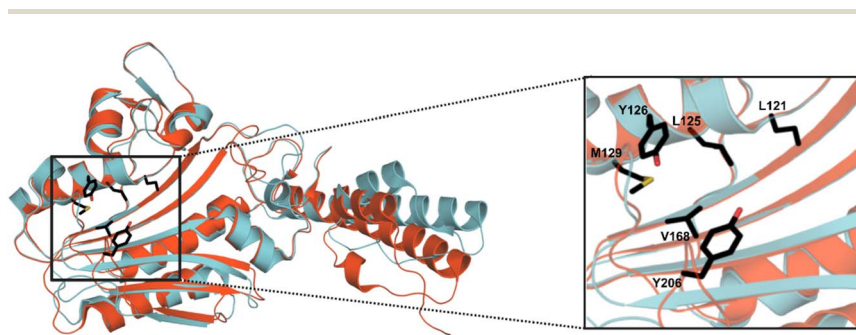


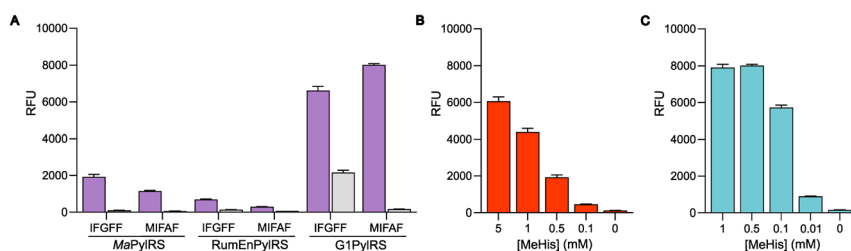
Fig. 2 Crystal structure overlay of *MaPylRS* and *G1PylRS*. An overlay of wild-type *MaPylRS* (red, PDB: 6JP2)<sup>26</sup> and wild-type *G1PylRS* (blue, PDB: 8IFJ).<sup>29</sup> Two distinct sets of mutations have been reported to confer activity towards MeHis incorporation in *MaPylRS*. The residues mutated in these two variants are shown as atom-coloured sticks with black carbons. The specific mutations present in each variant are detailed in the main body of text.



*Ma* numbering), affording *Ma*PylRS<sup>MIFAF</sup> or *Ma*PylRS<sup>IFGFF</sup> respectively (Fig. 2).<sup>26–28</sup> A recent study showed that *Ma*PylRS<sup>IFGFF</sup> gave modest improvements in protein yield compared with the analogous *Mm*PylRS<sup>IFGFF</sup> variant, although high concentrations of MeHis are required in both cases. With the aim of identifying more efficient systems, we elected to explore a wider range of PylRS homologs, namely *Methanogenic archaeon* ISO4-G1 (G1) PylRS and *Methanomassiliicoccales archaeon* RumEn M1 (RumEn) PylRS.<sup>25</sup> Similar to *Ma*PylRS, these homologs lack the N-terminal tRNA binding domain that is essential for activity in *Mm*PylRS and *Mb*PylRS. To develop engineered G1PylRS and RumEnPylRS for encoding MeHis, we mapped mutations from *Ma*PylRS<sup>MIFAF</sup> and *Ma*PylRS<sup>IFGFF</sup> into these homologs, which along with their respective cognate <sup>Pyl</sup>tRNA<sub>CUA</sub>, gave rise to four new aaRS/tRNA pairs for experimental characterization.

The resulting G1PylRS<sup>MIFAF/IFGFF</sup> and RumEnPylRS<sup>MIFAF/IFGFF</sup> variants were evaluated using an established GFP production assay and their activity compared to the analogous *Ma* systems. Of the two existing *Ma* variants, *Ma*PylRS<sup>IFGFF</sup> was shown to be slightly more effective in suppressing the UAG codon to produce full length GFP containing MeHis at position 150 (Fig. 3A). The UAG suppression efficiency of RumEnPylRS<sup>MIFAF/IFGFF</sup> was substantially reduced compared to the *Ma* variants. In contrast, both active-site transplanted G1PylRS variants showed substantially improved UAG suppression efficiency. G1PylRS<sup>MIFAF</sup> displays both high activity and specificity for MeHis (purple bars) over incorporation of canonical amino acids (grey bars), whereas the G1PylRS<sup>IFGFF</sup> variant suffers from a high background of phenylalanine incorporation in cultures grown in the absence of MeHis. This newly engineered G1PylRS<sup>MIFAF</sup> variant produces approximately 4-fold more full length GFP than *Ma*PylRS<sup>IFGFF</sup> when 0.5 mM MeHis is supplied to the culture medium.

To further compare the G1PylRS<sup>MIFAF</sup> and *Ma*PylRS<sup>IFGFF</sup> systems, we next evaluated GFP production across a range of MeHis concentrations (Fig. 3B, C and ESI Fig. S1<sup>†</sup>). Remarkably, G1PylRS<sup>MIFAF</sup> can operate efficiently using a MeHis concentration of 0.1 mM, with detectable levels of GFP production even observed



**Fig. 3** Production of MeHis-containing GFP using engineered PylRS homologs. (A) Bar chart showing production of GFP containing MeHis at position 150, using either *Ma*PylRS<sup>MIFAF/IFGFF</sup>/<sup>Pyl</sup>tRNA<sub>CUA</sub>, RumEnPylRS<sup>MIFAF/IFGFF</sup>/<sup>Pyl</sup>tRNA<sub>CUA</sub> or G1PylRS<sup>MIFAF/IFGFF</sup>/<sup>Pyl</sup>tRNA<sub>CUA</sub> pairs. Cultures grown in the presence of 0.5 mM MeHis (purple bars) or with no MeHis supplemented (grey bars). Error bars represent the standard deviation of measurements made in triplicate. (B) Bar chart showing GFP production in cultures containing varying MeHis concentrations (0–5 mM) using *Ma*PylRS<sup>IFGFF</sup>/<sup>Pyl</sup>tRNA<sub>CUA</sub>. Error bars represent the standard deviation of measurements made in triplicate. (C) Bar chart showing GFP production in cultures containing varying MeHis concentrations (0–1 mM) using G1PylRS<sup>MIFAF</sup>/<sup>Pyl</sup>tRNA<sub>CUA</sub>. Error bars represent the standard deviation of measurements made in triplicate.



at 0.01 mM. Increasing the MeHis concentration to 0.5 mM or 1 mM led to only modest improvements in MeHis incorporation, suggesting that G1PylRS<sup>MIFAF</sup> is saturated at a concentration between 0.1 and 0.5 mM. For comparison, G1PylRS<sup>MIFAF</sup> was a more effective aaRS at 0.1 mM MeHis than MaPylRS<sup>IFGFF</sup> at 1 mM. Even using 50 times more MeHis (5 mM), MaPylRS<sup>IFGFF</sup> is only marginally more active. To illustrate the efficacy of G1PylRS<sup>MIFAF</sup>, we further increased the stringency of GFP production assays by introducing an additional UAG codon at position 40 (Fig. 4). The G1PylRS<sup>MIFAF</sup> system is able to efficiently read through two UAG codons to produce GFP containing MeHis at positions 40 and 150, with only minor reductions in protein yield (Fig. 4B). In contrast, with the less efficient MaPylRS<sup>IFGFF</sup> system yields of doubly modified GFP are extremely low. It is notable that many enzymes contain multiple catalytically important histidine residues.<sup>30–33</sup> The ability to efficiently produce proteins containing multiple MeHis residues will open up new opportunities to study and/or tune the functions of these enzymes.

Having established an efficient aaRS/tRNA pair, our attention now turned to the production of engineered enzymes that use MeHis as an important catalytic element. To this end we selected the engineered peroxidase APX2\_MeHis, where MeHis serves as an axial ligand and leads to dramatically improved turnover numbers, and designed enzymes OE1.4 (stereoselective hydrolase) and BH\_MeHis1.8 (Morita–Baylis–Hillmanase) that both employ MeHis as a catalytic nucleophile.<sup>10,15,20</sup> Using only 0.1 mM MeHis, these engineered enzymes are all produced in >100 mg L<sup>-1</sup> in standard laboratory *Escherichia coli* strains and culture conditions, corresponding to an impressive 4–6% of the total MeHis supplemented being incorporated into protein (Table 1). In all cases, protein yields achieved with G1PylRS<sup>MIFAF</sup> are substantially higher than those produced with MaPylRS<sup>IFGFF</sup> using 10 times higher MeHis concentrations (1 mM).

Given the high efficiency of G1PylRS<sup>MIFAF</sup>, we wondered whether this aaRS is highly specific for MeHis or whether it could also be used to encode other ncAAs. We therefore tested G1PylRS<sup>MIFAF</sup> activity towards a small panel of ncAAs (10 mM) using the aforementioned GFP production assay. In addition to MeHis, G1PylRS<sup>MIFAF</sup> is able to encode five of the other ncAAs tested (Fig. 5, structures 4, 6, 7, 8, and 9). These substrates include the hydrophobic ncAAs 3-(2-naphthyl)

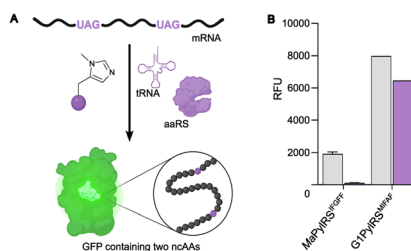


Fig. 4 Production of GFP containing two MeHis residues. (A) Schematic representation of modified GFP expression, introducing two MeHis residues in response to UAG codons at positions 40 and 150. (B) Bar chart showing GFP production containing MeHis at either position 150 (grey bars) or at positions 40 and 150 (purple bars) using either MaPylRS<sup>IFGFF</sup>/PyltRNA<sub>CUA</sub> or G1PylRS<sup>MIFAF</sup>/PyltRNA<sub>CUA</sub>. Cultures were supplemented with 0.5 mM MeHis. Error bars represent the standard deviation of measurements made in triplicate.



Table 1 Protein titres of engineered enzymes containing MeHis

Synthetase	[MeHis] (mM)	Protein yield (mg L <sup>-1</sup> )		
		APX2	BH <sub>MeHis</sub> 1.8	OE1.4
<i>Ma</i> PyIRS <sup>IFGFF</sup>	1	107	58	71
G1PyIRS <sup>MIFAF</sup>	0.1	188	127	106

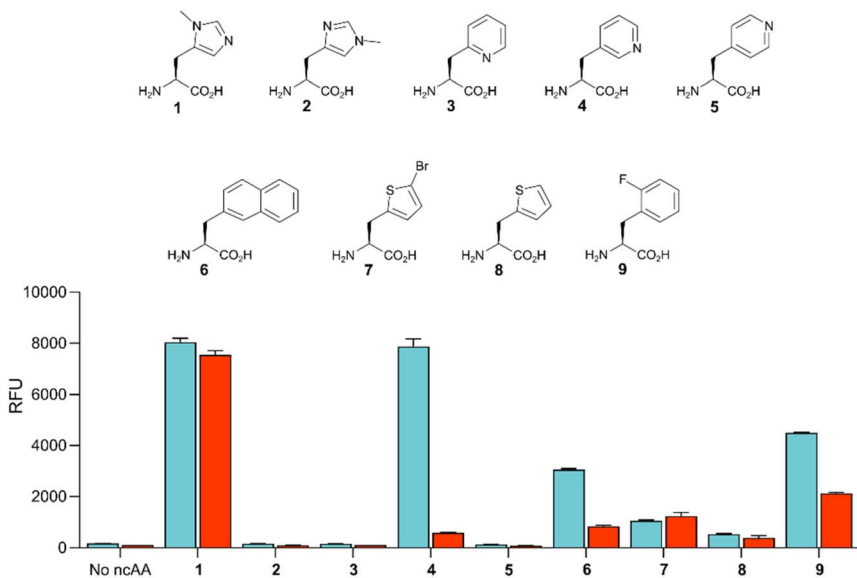


Fig. 5 Comparison of the polyspecificity of G1PyIRS<sup>MIFAF</sup> and *Ma*PyIRS<sup>IFGFF</sup> systems. Bar chart showing the activities of G1PyIRS<sup>MIFAF</sup> (blue) and *Ma*PyIRS<sup>IFGFF</sup> (red) towards a small panel of nCAAs (10 mM, 1–9). PyIRS activity was determined by monitoring production of modified GFP containing the appropriate nCAA at position 150. Error bars represent the standard deviation of measurements made in triplicate.

alanine and 2-fluorophenylalanine, as well as 3-pyridylalanine, a potentially valuable histidine analogue that is a poor substrate for *Ma*PyIRS<sup>IFGFF</sup> and *Ma*PyIRS<sup>MIFAF</sup> (ESI Fig. S2†). The ability of G1PyIRS<sup>MIFAF</sup> to efficiently discriminate between phenylalanine and 2-fluorophenylalanine is particularly notable.

## Conclusion

In summary, we have developed a highly efficient aaRS for encoding MeHis by introducing five known active site mutations into a single domain PyIRS from *Methanogenic archaeon* ISO4-G1. The successful development of G1PyIRS<sup>MIFAF</sup> serves to highlight the importance of exploring a wider range of PyIRS homologs when developing orthogonal translation components. This G1PyIRS<sup>MIFAF</sup> has allowed the efficient and economical production of a range of MeHis-containing enzymes using only 0.1 mM MeHis supplemented to the culture medium. Moving forward, there are several avenues for investigation to further enhance the



production of valuable proteins containing MeHis to underpin any future commercial applications. Firstly, it is likely that even more efficient G1PylRS<sup>MIFAF</sup> descendants can be developed through directed evolution using established high-throughput assays. Switching to high-density fermentation technologies will also likely boost protein yields. Secondly, we can take advantage of engineered or synthetic *E. coli* strains that have been specifically tailored for more efficient nAAs incorporation.<sup>8,34–36</sup> Alternatively, we can explore engineered yeast strains or mammalian cell lines that can be advantageous for selected protein applications.<sup>37–39</sup> Finally, considering MeHis is a naturally occurring amino acid, we can envision the development of engineered production hosts that contain the necessary biosynthetic machinery to produce MeHis and direct its selected incorporation into target proteins. For these reasons, we are optimistic that the work presented in this paper will provide an important step towards commercially viable production of MeHis-containing enzymes.

## Methods

### Materials

All chemicals and biological materials were obtained from commercial suppliers. *Escherichia coli* DH10 $\beta$  cells were purchased from Thermo Fisher. *Escherichia coli* BL21(DE3), Q5 polymerase, T4 ligase and restriction enzymes from New England BioLabs (NEB). PylRS and tRNA gene sequences were synthesized by Twist Bioscience. Primers were synthesized by IDT. Kanamycin, chloramphenicol and 5-aminolevulinic acid were purchased from Sigma-Aldrich. (*S*)-2-Amino-3-(5-bromothiophen-2-yl) propanoic acid, 2-fluoro-L-phenylalanine, 3-(3-pyridyl)-L-alanine and 3-(2-naphthyl)-L-alanine were purchased from Fluorochem. 3-(2-Thienyl)-L-alanine and 3-(2-pyridyl)-L-alanine were purchased from Fisher Scientific UK. 3-(4-Pyridyl)-L-alanine was purchased from Alfa-Aesar. H-His(1-Me)-OH was from abcr. LB agar, LB media, 2xYT media, isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG) and arabinose were from Formedium. H-His(3-Me)-OH (MeHis) was from Bachem.

Defined autoinducing medium (500 mL) had the following composition: 25 mL aspartate (5%, pH = 7.5), 25 mL glycerol (10% w/v), 20 mL 18-amino-acid mix (5 g L<sup>-1</sup> glutamic acid, 5 g L<sup>-1</sup> aspartic acid, 5 g L<sup>-1</sup> lysine·HCl, 5 g L<sup>-1</sup> arginine·HCl, 5 g L<sup>-1</sup> alanine, 5 g L<sup>-1</sup> proline, 5 g L<sup>-1</sup> glycine, 5 g L<sup>-1</sup> threonine, 5 g L<sup>-1</sup> serine, 5 g L<sup>-1</sup> glutamine, 5 g L<sup>-1</sup> asparagine·H<sub>2</sub>O, 5 g L<sup>-1</sup> valine, 5 g L<sup>-1</sup> leucine, 5 g L<sup>-1</sup> isoleucine, 5 g L<sup>-1</sup> phenylalanine, 5 g L<sup>-1</sup> tryptophan, 5 g L<sup>-1</sup> methionine, histidine-HCl·H<sub>2</sub>O 5 g L<sup>-1</sup>, sterile filtered), 1.25 mL arabinose (20% w/v), 20 mL 25  $\times$  M salts (0.625 M NaH<sub>2</sub>PO<sub>4</sub>, 0.625 M KH<sub>2</sub>PO<sub>4</sub>, 1.25 M NH<sub>4</sub>Cl, 0.125 M Na<sub>2</sub>SO<sub>4</sub>), 1 mL MgSO<sub>4</sub> (1 M), 0.625 mL glucose (40% w/v), 100  $\mu$ L 5000  $\times$  trace metals solution (20 mM CaCl<sub>2</sub>·2H<sub>2</sub>O, 10 mM MnCl<sub>2</sub>·H<sub>2</sub>O, 10 mM ZnSO<sub>4</sub>·7H<sub>2</sub>O, 2 mM CoCl<sub>2</sub>·6H<sub>2</sub>O, 2 mM CuCl<sub>2</sub>, 2 mM NiCl<sub>2</sub>, 2 mM Na<sub>2</sub>MoO<sub>4</sub>·2H<sub>2</sub>O, 2 mM NaSeO<sub>3</sub>, 2 mM H<sub>3</sub>BO<sub>3</sub>, 50 mM FeCl<sub>3</sub>), 1 mL MgSO<sub>4</sub> (1 M), 500  $\mu$ L IPTG (0.1 M). The solution was made up with sterile water to 500 mL.

### DNA constructs

pEVOL\_MaPylRS<sup>IFGFF</sup>/Ma<sup>Pyl</sup>tRNA<sub>CUA</sub> was available from a previous study.<sup>26</sup> PylRS genes (G1PylRS<sup>MIFAF</sup> and RumEnPylRS<sup>MIFAF</sup>), optimized for *E. coli* expression, and



the tRNAs ( $G1^{Pyl}tRNA_{CUA}$  and  $RumEn^{Pyl}tRNA_{CUA}$ ) were synthesized by Twist Bioscience. Two copies of each PylRS gene and their corresponding tRNA were cloned into their respective pEVOL vectors using *NdeI/PstI* and *BglIII/SalI* restriction sites for PylRS genes and *ApaLI/XhoI* for the tRNA. To make  $MaPylRS^{MIFAF}$ ,  $G1PylRS^{IFGFF}$  and  $RumEnPylRS^{IFGFF}$  primers to introduce the required mutations were used to make gene fragments which were combined using overlap extension PCR. Two copies of each gene were cloned into their respective pEVOL vectors using *NdeI/PstI* and *BglIII/SalI* restriction sites.

### GFP expression assays

Chemically competent *E. coli* BL21(DE3) cells containing the appropriate pEVOL vector were transformed with either pET28\_GFP\_Asn150TAG or pET28\_GFP\_Asn40TAG\_Asn150TAG plasmid. Single colonies of freshly transformed cells were cultured in 5 mL of LB media containing  $50 \mu\text{g mL}^{-1}$  kanamycin and  $25 \mu\text{g mL}^{-1}$  chloramphenicol for 18 h at 30 °C. Expression cultures were grown in 96-deepwell blocks sealed with a breathable membrane. 20  $\mu\text{L}$  of the starter culture was used to inoculate 480  $\mu\text{L}$  of defined auto-induction medium containing  $50 \mu\text{g mL}^{-1}$  kanamycin and  $25 \mu\text{g mL}^{-1}$  chloramphenicol (for the cultures with Asn40TAG\_Asn150TAG plasmid, IPTG was removed from the auto-induction medium and added when the cultures reached  $OD_{600} = 0.6$ ). Expression cultures were grown in the presence of the appropriate nCAA (0–10 mM) and incubated at 30 °C with shaking at 850 rpm for 20 h.  $OD_{600}$  and GFP fluorescence ( $\lambda_{\text{excitation}}$ : 395 nm,  $\lambda_{\text{emission}}$ : 509 nm) measurements were recorded using a BMG LabTech CLARIOstar spectrophotometer.

### Protein production and purification of MeHis-containing proteins

For the expression of APX2\_MeHis, chemically competent *E. coli* BL21(DE3) containing either pEVOL\_ $MaPylRS^{IFGFF}/Ma^{Pyl}tRNA_{CUA}$  or pEVOL\_ $G1PylRS^{MIFAF}/G1^{Pyl}tRNA_{CUA}$  were transformed with pET29b\_APX\_MeHis. A single colony of freshly transformed cells were cultured in 5 mL of LB media containing  $50 \mu\text{g mL}^{-1}$  kanamycin and  $25 \mu\text{g mL}^{-1}$  chloramphenicol for 18 h at 30 °C. 300  $\mu\text{L}$  of the starter cultures was used to inoculate 30 mL 2xYT medium supplemented with  $50 \mu\text{g mL}^{-1}$  kanamycin,  $25 \mu\text{g mL}^{-1}$  chloramphenicol, 5-aminolevulinic acid (1 mM final) and MeHis (1–0.1 mM final) and cultures were grown at 37 °C, 200 rpm to an  $OD_{600}$  of 0.6. Protein expression was induced with the addition of IPTG (0.1 mM final) and arabinose (5 mM final) and the cultures grown for a further 20 h at 20 °C.

For the expression of  $BH_{MeHis}1.8$  and OE1.4, chemically competent *E. coli* DH10 $\beta$  containing either pEVOL\_ $MaPylRS^{IFGFF}/Ma^{Pyl}tRNA_{CUA}$  or pEVOL\_ $G1PylRS^{MIFAF}/G1^{Pyl}tRNA_{CUA}$  were transformed with either pBbe8K\_  $BH_{MeHis}1.8$  or pBbe8K\_OE1.4. A single colony of freshly transformed cells were cultured in 5 mL of LB media containing  $50 \mu\text{g mL}^{-1}$  kanamycin and  $25 \mu\text{g mL}^{-1}$  chloramphenicol for 18 h at 30 °C. 300  $\mu\text{L}$  of the starter cultures was used to inoculate 30 mL 2xYT medium supplemented with  $50 \mu\text{g mL}^{-1}$  kanamycin,  $25 \mu\text{g mL}^{-1}$  chloramphenicol, and MeHis (1–0.1 mM final) and cultures were grown at 37 °C, 200 rpm to an  $OD_{600}$  of 0.6. Protein expression was induced with the arabinose (10 mM final) and the cultures grown for a further 20 h at 20 °C.

The cells were harvested and purified as stated above for PylRS purification. Purified proteins were desalted using 10DG desalting columns (Bio-Rad) with PBS





pH 7.4 and analysed by SDS-PAGE and protein MS. Protein concentrations were determined by measuring the absorbance at 280 nm using calculated extinction coefficients (ExPASy ProtParam).

### MS analysis

Purified protein samples were diluted to a final concentration of 0.5 mg mL<sup>-1</sup> with 0.1% acetic acid. MS analysis was performed using a 1200 series Agilent LC, 5 µL injection into 5% acetonitrile (with 0.1% formic acid) and desalted inline for 1 min. Protein was eluted over 1 min using 95% acetonitrile and 5% water. The resulting multiply charged spectrum was analyzed using an Agilent QTOF 6510 and deconvoluted using Agilent MassHunter software.

## Author contributions

A. E. H. and J. F. carried out all laboratory work. J. E. J. S. identified homologs to test and created plasmids for these. All authors discussed the results and participated in writing the manuscript. A. P. G. initiated and directed the research.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

We acknowledge a UKRI Engineering Biology Mission Award (Grant no. BB/Y00812X/1), the European Research Council (ERC Starting Grant no. 757991 to A. P. G.), UK Research and Innovation (UKRI Frontier Research Guarantee to A. P. G. EP/Y023722/1), the Biotechnology and Biological Sciences Research Council (BB/W014483/1 and BB/X000974/1, to A. P. G.), the Human Frontier Science Program research grant (RGP0004/2022). A. E. H. was supported by a BBSRC Industrial CASE PhD studentship (BB/S507040/1) supported by GSK. J. F. was supported by an integrated catalysis Doctoral Training Program (EP/023755/1). J. S. was supported by an EPSRC DTG studentship. C. T. was supported by a Centre of Excellence for Biocatalysis, Biotransformations and Biocatalytic Manufacture Studentship. We are grateful to the Manchester SYNBIOCHEM Centre (BB/M017702/1), the Future Bio-manufacturing Hub (EP/S01778X/1) and the Henry Royce Institute for Advanced Materials (financed through EPSRC grant no. EP/R00661X/1, EP/S019367/1, EP/P025021/1 and EP/P025498/1). We thank R. Spiess and M. Trelore (Manchester Institute of Biotechnology) for acquiring protein mass spectra.

## References

- 1 S. L. Lovelock, R. Crawshaw, S. Basler, C. Levy, D. Baker, D. Hilvert and A. P. Green, *Nature*, 2022, **606**, 49–58.
- 2 R. L. Koder and P. L. Dutton, *Dalton Trans.*, 2006, 3045.
- 3 B. A. Smith and M. H. Hecht, *Curr. Opin. Chem. Biol.*, 2011, **15**, 421–426.
- 4 B. Leader, Q. J. Baca and D. E. Golan, *Nat. Rev. Drug Discovery*, 2008, **7**, 21–39.



- 5 M. A. Huffman, A. Fryszkowska, O. Alvizo, M. Borra-Garske, K. R. Campos, K. A. Canada, P. N. Devine, D. Duan, J. H. Forstater, S. T. Grosser, H. M. Halsey, G. J. Hughes, J. Jo, L. A. Joyce, J. N. Kolev, J. Liang, K. M. Maloney, B. F. Mann, N. M. Marshall, M. McLaughlin, J. C. Moore, G. S. Murphy, C. C. Nawrat, J. Nazor, S. Novick, N. R. Patel, A. Rodriguez-Granillo, S. A. Robaire, E. C. Sherer, M. D. Truppo, A. M. Whitaker, D. Verma, L. Xiao, Y. Xu and H. Yang, *Science*, 2019, **366**, 1255–1259.
- 6 M. Schober, C. MacDermaid, A. A. Ollis, S. Chang, D. Khan, J. Hosford, J. Latham, L. A. F. Ihnken, M. J. B. Brown, D. Fuerst, M. J. Sanganeer and G.-D. Roiban, *Nat. Catal.*, 2019, **2**, 909–915.
- 7 C. C. Liu and P. G. Schultz, *Annu. Rev. Biochem.*, 2010, **79**, 413–444.
- 8 J. W. Chin, *Nature*, 2017, **550**, 53–60.
- 9 W. Wan, J. M. Tharp and W. R. Liu, *Biochim. Biophys. Acta, Proteins Proteomics*, 2014, **1844**, 1059–1070.
- 10 A. P. Green, T. Hayashi, P. R. E. Mittl and D. Hilvert, *J. Am. Chem. Soc.*, 2016, **138**, 11344–11352.
- 11 M. Pott, T. Hayashi, T. Mori, P. R. E. Mittl, A. P. Green and D. Hilvert, *J. Am. Chem. Soc.*, 2018, **140**, 1535–1543.
- 12 M. Ortmayer, K. Fisher, J. Basran, E. M. Wolde-Michael, D. J. Heyes, C. Levy, S. L. Lovelock, J. L. R. Anderson, E. L. Raven, S. Hay, S. E. J. Rigby and A. P. Green, *ACS Catal.*, 2020, **10**, 2735–2746.
- 13 M. R. Seyedsayamdost, C. S. Yee, S. Y. Reece, D. G. Nocera and J. Stubbe, *J. Am. Chem. Soc.*, 2006, **128**, 1562–1568.
- 14 Y. Wu and S. G. Boxer, *J. Am. Chem. Soc.*, 2016, **138**, 11890–11895.
- 15 A. J. Burke, S. L. Lovelock, A. Frese, R. Crawshaw, M. Ortmayer, M. Dunstan, C. Levy and A. P. Green, *Nature*, 2019, **570**, 219–223.
- 16 I. Drienovská, C. Mayer, C. Dulson and G. Roelfes, *Nat. Chem.*, 2018, **10**, 946–952.
- 17 C. Mayer, C. Dulson, E. Reddem, A. W. H. Thunnissen and G. Roelfes, *Angew. Chem., Int. Ed.*, 2019, **58**, 2083–2087.
- 18 N. Sun, J. Huang, J. Qian, T.-P. Zhou, J. Guo, L. Tang, W. Zhang, Y. Deng, W. Zhao, G. Wu, R.-Z. Liao, X. Chen, F. Zhong and Y. Wu, *Nature*, 2022, **611**, 715–720.
- 19 J. S. Trimble, R. Crawshaw, F. J. Hardy, C. W. Levy, M. J. B. Brown, D. E. Fuerst, D. J. Heyes, R. Obexer and A. P. Green, *Nature*, 2022, **611**, 709–714.
- 20 A. E. Hutton, J. Foster, R. Crawshaw, F. J. Hardy, L. O. Johannissen, E. F. Gerard, Z. Birch-Price, R. Obexer, S. Hay and A. P. Green, *Nat. Commun.*, 2024, **15**, 1956.
- 21 F. Gan, R. Liu, F. Wang and P. G. Schultz, *J. Am. Chem. Soc.*, 2018, **140**, 3829–3832.
- 22 D. M. Carminati and R. Fasan, *ACS Catal.*, 2019, **9**, 9683–9697.
- 23 T. Hayashi, M. Tinzl, T. Mori, U. Kregel, J. Proppe, J. Soetbeer, D. Klose, G. Jeschke, M. Reiher and D. Hilvert, *Nat. Catal.*, 2018, **1**, 578–584.
- 24 W. Wan, J. M. Tharp and W. R. Liu, *Biochim. Biophys. Acta, Proteins Proteomics*, 2014, **1844**, 1059–1070.
- 25 J. C. W. Willis and J. W. Chin, *Nat. Chem.*, 2018, **10**, 831–837.
- 26 C. J. Taylor, F. J. Hardy, A. J. Burke, R. M. Bednar, R. A. Mehl, A. P. Green and S. L. Lovelock, *Protein Sci.*, 2023, **32**, e4640.
- 27 H. Xiao, F. B. Peters, P.-Y. Yang, S. Reed, J. R. Chittuluru and P. G. Schultz, *ACS Chem. Biol.*, 2014, **9**, 1092–1096.



- 28 V. Beránek, J. C. W. Willis and J. W. Chin, *Biochemistry*, 2019, **58**, 387–390.
- 29 T. Yanagisawa, E. Seki, H. Tanabe, Y. Fujii, K. Sakamoto and S. Yokoyama, *Int. J. Mol. Sci.*, 2023, **24**, 6256.
- 30 C. Krebs, D. Galonić Fujimori, C. T. Walsh and J. M. Bollinger, *Acc. Chem. Res.*, 2007, **40**, 484–492.
- 31 J. C. Lewis, P. S. Coelho and F. H. Arnold, *Chem. Soc. Rev.*, 2011, **40**, 2003–2021.
- 32 F. H. Vaillancourt, E. Yeh, D. A. Vosburg, S. Garneau-Tsodikova and C. T. Walsh, *Chem. Rev.*, 2006, **106**, 3364–3378.
- 33 P. H. Walton and G. J. Davies, *Curr. Opin. Chem. Biol.*, 2016, **31**, 195–207.
- 34 N. Ostrov, M. Landon, M. Guell, G. Kuznetsov, J. Teramoto, N. Cervantes, M. Zhou, K. Singh, M. G. Napolitano, M. Moosburner, E. Shrock, B. W. Pruitt, N. Conway, D. B. Goodman, C. L. Gardner, G. Tyree, A. Gonzales, B. L. Wanner, J. E. Norville, M. J. Lajoie and G. M. Church, *Science*, 2016, **353**, 819–822.
- 35 J. Fredens, K. Wang, D. de la Torre, L. F. H. Funke, W. E. Robertson, Y. Christova, T. Chia, W. H. Schmied, D. L. Dunkelmann, V. Beránek, C. Uttamapinant, A. G. Llamazares, T. S. Elliott and J. W. Chin, *Nature*, 2019, **569**, 514–518.
- 36 T. M. Wannier, A. M. Kunjapur, D. P. Rice and G. M. Church, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 3090–3095.
- 37 S. M. Richardson, L. A. Mitchell, G. Stracquadanio, K. Yang, J. S. Dymond, J. E. DiCarlo, D. Lee, C. L. V. Huang, S. Chandrasegaran, Y. Cai, J. D. Boeke and J. S. Bader, *Science*, 2017, **355**, 1040–1044.
- 38 Y. Zhao, C. Coelho, A. L. Hughes, L. Lazar-Stefanita, S. Yang, A. N. Brooks, R. S. K. Walker, W. Zhang, S. Lauer, C. Hernandez, J. Cai, L. A. Mitchell, N. Agmon, Y. Shen, J. Sall, V. Fanfani, A. Jalan, J. Rivera, F.-X. Liang, J. S. Bader, G. Stracquadanio, L. M. Steinmetz, Y. Cai and J. D. Boeke, *Cell*, 2023, **186**, 5220–5236.
- 39 J. Sanders, S. A. Hoffmann, A. P. Green and Y. Cai, *Curr. Opin. Biotechnol.*, 2022, **75**, 102691.

