Published in final edited form as:

Biochim Biophys Acta. 2017 November ; 1861(11 Pt B): 3009–3015. doi:10.1016/j.bbagen.2017.03.003.

A genomically modified *Escherichia coli* strain carrying an orthogonal *E. coli* histidyl-tRNA synthetase tRNA^{His} pair

Markus Englert^{a,1}, Oscar Vargas-Rodriguez^{a,1}, Noah M. Reynolds^a, Yane-Shih Wang^{a,†}, Dieter Söll^{a,b,*}, and Takuya Umehara^{a,c,*}

^aDepartment of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA

^bDepartment of Chemistry, Yale University, New Haven, CT 06520, USA

^cDepartment of Biological Science and Technology, Tokyo University of Science, 6-3-1 Niijuku, Katsushika-ku, Tokyo 125-8585, Japan

Abstract

Background—Development of new aminoacyl-tRNA synthetase (aaRS)•tRNA pairs is central for incorporation of novel non-canonical amino acids (ncAAs) into proteins via genetic code expansion (GCE). The *Escherichia coli* and *Caulobacter crescentus* histidyl-tRNA synthetases (HisRS) evolved divergent mechanisms of tRNA^{His} recognition that prevent their cross-reactivity. Although the *E. coli* HisRS•tRNA^{His} pair is a good candidate for GCE, its use in *C. crescentus* is limited by the lack of established genetic selection methods and by the low transformation efficiency of *C. crescentus*.

Methods—*E. coli* was genetically engineered to use a *C. crescentus* HisRS•tRNA^{His} pair. Superfolder green fluorescent protein (sfGFP) and chloramphenicol acetyltransferase (CAT) were used as reporters for read-through assays. A library of 313 ncAAs coupled with the sfGFP reporter system was employed to investigate the specificity of *E. coli* HisRS *in vivo*.

Results—A genomically modified *E. coli* strain (named MEOV1) was created. MEOV1 requires an active *C. crescentus* HisRS•tRNA^{His} pair for growth, and displays a similar doubling time as the parental *E. coli* strain. sfGFP- and CAT-based assays showed that the *E. coli* HisRS•tRNA^{His} pair is orthogonal in MEOV1 cells. A mutation in the anticodon loop of *E. coli* tRNA^{His}_{CUA} elevated its suppression efficiency by 2-fold.

*Corresponding authors: tumehara@rs.noda.tus.ac.jp (T. Umehara) and dieter.soll@yale.edu (D. Söll).

¹These authors have contributed equally to this work.

[†]Present address: Institute of Biological Chemistry, Academia Sinica, 128, Academia Road Sec. 2, Nankang, Taipei 115, Taiwan.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conclusions—The *C. crescentus* HisRS•tRNA^{His} pair functionally complements an *E. coli* *hisS* strain. The *E. coli* HisRS•tRNA^{His} is orthogonal in MEOV1 cells. *E. coli* tRNA^{His}_{CUA} is an efficient amber suppressor in MEOV1.

General significance—We developed a platform that allows protein engineering of *E. coli* HisRS that should facilitate GCE in *E. coli*.

Keywords

Genetic code expansion; Orthogonal pair; Non-canonical amino acids; Aminoacyl-tRNA synthetase; tRNA; Synthetic biology

1. Introduction

For over a decade, genetic code expansion (GCE) strategies have been used to incorporate a wide range of non-canonical amino acids (ncAAs) into proteins in living cells, making this technology a powerful tool for investigation of protein structure and function [1].

Incorporation of ncAAs via GCE relies on the reassignment of a codon (a stop codon is typically used) mediated by an aaRS•tRNA pair that functions to provide the ncAA-tRNA for ribosomal protein synthesis. The aaRS•tRNA pairs used in GCE must be orthogonal to the host organism; that is, the aaRS should exclusively interact with its cognate tRNA, and the tRNA should not be a substrate for the host's aaRSs. To fulfil this requisite for GCE, most aaRS•tRNA pairs used in model organisms (*e.g.*, *E. coli*, mouse, or human) are derived from organisms belonging to a different domain of life. Additionally, aaRSs with weak or no tRNA anticodon specificity are most attractive for GCE applications [2].

Tyrosyl- and pyrrolysyl-tRNA synthetase (TyrRS and PylRS) from *Methanocaldococcus jannaschii* and *Methanosarcina barkeri/mazei*, respectively, are the most commonly used aaRS•tRNA pairs in GCE as they and their engineered variants account for incorporation of more than 150 ncAAs [1, 3, 4]. However, despite their profound utility, the discovery and development of new aaRS•tRNA pairs remains imperative for GCE for two reasons [3]. First, new aaRS•tRNA pairs should expand the diversity of ncAAs in GCE. The scope of ncAAs incorporated by TyrRS and PylRS is biased towards aromatic, bulky, and large amino acids, in line with their large amino acid binding pockets, limiting the repertoire of ncAAs they can be used to incorporate. Second, a set of diverse aaRS•tRNA pairs will be essential for the development of platforms to incorporate multiple ncAAs into a single polypeptide [5, 6].

Recently, histidyl-tRNA synthetases (HisRS) from *Caulobacter crescentus* and *Escherichia coli* were reported to be orthogonal to each other *in vitro* [7] and *in vivo* [8]. In *E. coli*, an unusual G₋₁ and the C73 base are identity elements for acylation by HisRS [9], whereas the *C. crescentus* tRNA^{His} lacks G₋₁ and contains A73, a base not recognized by *C. crescentus* HisRS [7]. Instead, *C. crescentus* HisRS recognizes an extended D-loop, U72, and the anticodon bases (GUG) as main identity elements for histidylation [7]. Thus, the introduction of *E. coli* HisRS and *E. coli* amber suppressor tRNA^{His}_{CUA} enabled amber suppression in *C. crescentus* [8]. Despite this achievement it was clear that further development of the orthogonal *E. coli* pair was limited in *Caulobacter*, as this organism's

molecular genetic versatility is not as well understood as that of *E. coli*. Thus, we explored an alternative strategy to exploit *E. coli* HisRS•tRNA^{His} for GCE. We planned to switch *E. coli*'s housekeeping histidyl-tRNA function from the *E. coli* HisRS•tRNA^{His} pair to that of *C. crescentus* in a genomically modified *E. coli* strain in which the original *E. coli* HisRS paired with its cognate amber suppressor tRNA^{His} would serve as an orthogonal histidine translational system.

2. Materials and Methods

2.1. Reagents and Strains

Restriction enzymes, T4 DNA ligase, and the Gibson Assembly mix were obtained from New England Biolabs. All amino acids were purchased from Sigma-Aldrich, Chem-Impex and Bachem. The Keio collection *E. coli* His-auxotroph strain JW2004-1 (*hisB270*) was obtained from the Coli Genetic Stock Center (Yale University).

2.2. Plasmid construction

The *C. crescentus* histidyl-tRNA synthetase (*hisS* gene) and tRNA^{His} (*hisT* gene) expression cassette (CchisST) was constructed by first cloning *hisT* into the EcoRI and BamHI sites of the plasmid pGFib under the *lpp* promoter and *rrnC* terminator, and *hisS* (codon-optimized for expression in *E. coli*) under the constitutive promoter *glnS* into the NdeI and HindIII sites of the plasmid pGLN (Supplementary Information). The *lpp.CchisT.rrnC* fragment was then PCR-amplified from CchisT.pGFib, digested with NotI, and cloned into CchisS.pGLN. The resulting CchisRS-tRNA^{His} expression cassette was later cloned into the pUC18 and pKD13 vectors to generate CchisST.pUC18 and CchisST.pKD13, respectively, using Gibson assembly. The T7 promoter of the multiple cloning sites of pRSFDuet1 and pCDFDuet1 (Novagen) were replaced with the *tac* promoter using inverse PCR resulting in *tac.pRSF* and *tac.pCDF*. Additionally, the *lac* repressor and the T7 promoter of pCDFDuet1 were replaced with the constitutive *lpp* promoter (*lpp.pCDF*). WTsfGFP was amplified and cloned into *tac.pCDF*. The *E. coli* tRNA^{His}_{CUA} gene (*EchisR_{am}*) was cloned into the EcoRI and BamHI sites of pGFib. The *lpp.EchisR_{am}.rrnC* cassette from EchisR_{am}.pGFib was cloned downstream of the T7 terminator of WTsfGFP.*tac.pCDF*. *EchisR_{am}* was also cloned into the pACYC184-derived plasmid pCAT [10], which encodes the chloramphenicol acetyl-transferase gene (*cat*) with an amber stop codon at position 112. The *E. coli* HisRS gene (*EchisS*) was amplified from EchisS.pCA24N [11] and cloned into plasmids *tac.pRSF* and *lpp.pCDF*. EchisR_{am} was cloned into pTECH to generate EchisR_{am}.pTECH. Construction of sfGFP2TAG.pBAD was previously described [12]. Site-directed mutagenesis was carried out using the QuikChange kit from Agilent. A list of all plasmid used is shown in Supplementary Table 1.

2.3. *E. coli* MEOV1 strain construction

The *E. coli* MEOV0.1 strain was constructed by integrating the CchisST expression cassette into the chromosome of *E. coli* TOP10 cells using the λ -Red recombination system from the plasmid pKD46 as described by Datsenko and Wanner [13]. The CchisST cassette was integrated at the *yohCI yohD* locus together with a kanamycin (Kan) marker from pKD13 [13]. The Kan marker was later removed using the FLP recombinase from the plasmid

pCP20. The Integration of CchisST and removal of the Kan marker was confirmed by PCR amplification of the corresponding genomic regions. Next, the plasmid CchisST.pUC18 was introduced into the resulting *E. coli* strain, TOP10.CchisST (or MEOV0.1), followed by transformation of the plasmid pSIM5 [14]. Chromosomal deletion of the *EchisS* from the *E. coli* MEOV0.1 strain was mediated by the λ -Red recombination system encoded in pSIM5 following a previously published protocol [14]. Briefly, MEOV0.1 cells with CchisST.pUC18 and pSIM5 were grown at 30 °C to an optical density at 600 nm (OD₆₀₀) of 0.6, shifted to 42 °C for 20 min and the cells were made electrocompetent. MEOV0.1 cells were transformed with an *EchisS* knockout cassette amplified from pKD13 containing 50 base-pair homologous regions flanking the *EchisS* gene, targeting the genomic *hisS* region. After successful knock out of Ec HisRS, the Kan resistance marker was removed using pCP20. The resulting strain, TOP10.CchisST. *hisS*, was named MEOV0.2. Finally, a mutagenic cassette targeting the only chromosomal copy of tRNA^{His} (*hisR*) in *E. coli* was used to mutate its anticodon sequence from GTG to CTA. The tRNA^{His} mutagenic cassette was fused to the Kan resistance cassette of pKD13 via overlapping PCR using three PCR fragments constructed as follows. PCR fragments 1 and 2 were amplified from *E. coli* genomic DNA and with primers (HISR_KFN and hisR.CTA_QR and hisR.CTA_QF and HISR_CAS_R, respectively, Supplementary Table 2) introducing the *hisR* anticodon change from GTG to CTA. Both fragments were digested with DpnI to remove residual template DNA and gel-purified. Fragment 3 was amplified from the plasmid pKD13 with primers PKD13.F2 and HISR_KRN. All 3 fragments were mixed and used for overlapping PCR with HISR_KFN and HISR_KRN to produce the mutagenic knock-in cassette. MEOV0.2 cells with pSIM5 were made λ -Red- and electrocompetent as described above, and transformed with the mutagenic knock-in cassette. Colony PCR was used to screen for the chromosomal GTG-to-CTA mutation indicated by a PCR product using the primers HISR_KFN and hisR.CTA_SCR. After pCP20-mediated removal of the Kan marker, the GTG to CTA mutation in *hisR* was verified by DNA sequencing. This strain was named MEOV1.

2.4. MEOV1 cells growth rate

Single colonies for MEOV1 and TOP10 cells were grown overnight at 37 °C in LB media without antibiotics. The overnight cultures were then diluted in LB media to an OD₆₀₀ of ~0.01. 125 μ L of the diluted cultures were transferred to a 96-well assay plate (Corning) and cell growth was monitored for 24 h at 37 °C in a Synergy HT plate reader (BioTek) without antibiotics. The final growth curves are the average of three independent experiments using three different colonies.

2.5. sfGFP reporter assays

MEOV1 cells harboring the plasmid *EchisS.tac.pRSF* were transformed with sfGFP2TAG.*EchisRam.tac.pCDF*, WTsfGFP.*EchisRam.tac.pCDF*, WTsfGFP.*tac.pCDF*, or sfGFP2TAG.*tac.pCDF*. Single colonies were grown overnight in LB medium with the corresponding antibiotics at 37 °C. Overnight cultures were diluted in LB media to an OD₆₀₀ of ~0.01, and 125 μ L of the cultures were then transferred to a 96-well assay plate. Protein expression was induced with 0.2 mM isopropyl β -D-1-thiogalactopyranoside. Additionally, 3 μ L of the overnight cell cultures were spotted on LB-agar plates containing

antibiotics and isopropyl β -D-1-thiogalactopyranoside. The plates were incubated for 24 h at 37 °C. Cell growth and sfGFP fluorescence ($\lambda_{\text{ex}} = 485 \text{ nm}$ and $\lambda_{\text{em}} = 528 \text{ nm}$) were monitored for 24 h at 37 °C in a Synergy HT plate reader (BioTek). Relative fluorescence was calculated by dividing the fluorescence arbitrary unit (a.u.) by the OD₆₀₀ at a specific time point. The bars and standard deviations are the average of three assays from three distinct colonies.

2.7. sfGFP purification

0.5 ml of an overnight culture of MEOV1 cells expressing sfGFP2TAG, *E. coli* tRNA^{His} and HisRS was used to inoculate 1 L of LB media containing the appropriate antibiotics. Cells were grown to an A₆₀₀ of 0.6 at 37 °C. sfGFP overexpression was induced with 0.1 mM IPTG for ~14 h at 25 °C. Cells were harvested and lysed with buffer containing 50 mM Tris (pH 8), 300 mM NaCl and 10 mM imidazole and sonication. Lysed cells were centrifuged at 11000 rpm for 35 min at 4 °C, and the supernatant was loaded on 1 mL of TALON (Takara) metal affinity resin. The resin was washed with lysis buffer and the protein was eluted with 250 mM imidazole.

2.6. Chloramphenicol resistance assay

MEOV1 cells were co-transformed with EchisR_{am}-pCAT and EchisS.*lpp*-pCDF. Colonies were then grown overnight in 3 ml LB medium supplemented with the corresponding antibiotics at 37°C. Serial dilutions of the overnight cultures were prepared with OD₆₀₀ of 0.1, 0.01, and 0.001 with fresh LB medium. 2 μ L of the diluted cells were spotted on LB-agar plates with 100 μ g/ml ampicillin, 10 μ g/ml tetracycline, 100 μ g/ml spectinomycin, and varying concentration of chloramphenicol. The plates were incubated at 37°C for 48 hours. Negative controls were carried out using the empty pCDFDuet1.

2.8. Amino acid library screening

The His-auxotrophic *E. coli* strain JW2004-1 (*hisB*⁻) from the Keio collection were co-transformed with the plasmids sfGFP2TAG.pBAD and EchisR_{am}-pTECH [12, 15]. Cells were grown overnight at 37°C in LB medium with the corresponding antibiotics (kanamycin, ampicillin, and chloramphenicol). The cells were pelleted and washed three times with 1x M9 minimal salts (BD), and finally resuspended with M9 minimal media (1x M9 minimal salts, 0.1 mM CaCl₂, 1mM MgSO₄, and 1% glycerol), the appropriate antibiotics, and 0.2% arabinose. 50- μ L aliquots of cell suspension were transferred to a 384-well plate containing 1 mM of different amino acids in each well [12]. Cell growth and sfGFP expression were monitored by measuring OD₆₀₀ and fluorescence intensity, respectively, in a Synergy HT plate reader for 12 h at 37°C.

2.9. In vitro aminoacylation assays

E. coli tRNA^{His} was *in vitro* transcribed, purified, and ³²P-labeled following previously described protocols [16]. *E. coli* HisRS was expressed and purified as previously described [17]. Aminoacylation reactions were performed using trace amounts of ³²P-labeled tRNA^{His} with 2 μ M unlabeled tRNA^{His} and 5 mM amino acid in buffer containing 50 mM HEPES (pH 7.5), 4 mM ATP, 10 mM MgCl₂, 0.1 mg/mL BSA, and 1 mM Dithiothreitol. Reactions

were initiated by addition of 50 nM *E. coli* HisRS. Time points were taken by quenching 2 μ L of the reaction mixture into 5 μ L of a solution containing 0.4 units/ μ L nuclease P1 (Sigma) in 200 mM sodium acetate (pH 5). 1 μ L of quenched solution was run on polyethyleneimine-cellulose TLC plates to separate the aminoacyl-[32 P]AMP and [32 P]AMP products.

3. Results and discussion

3.1. Functional replacement of *E. coli* His-tRNA^{His} biosynthesis

HisRS•tRNA^{His} pairs from *E. coli* and *C. crescentus* have orthogonal mechanisms of tRNA recognition that has been demonstrated *in vitro* and *in vivo* [7, 8]. A key distinction between *E. coli* and *C. crescentus* HisRS•tRNA recognition is the strong specificity of *C. crescentus* HisRS for the anticodon sequence GUG of tRNA^{His} [7]. This property of *C. crescentus* HisRS renders the enzyme unfit for use in genetic code expansion as it fails to aminoacylate *C. crescentus* tRNA^{His} with an altered anticodon [7]. In contrast, converting *E. coli* tRNA^{His} to an amber suppressor is well tolerated by *E. coli* HisRS, making it a more suitable candidate for the development of an *E. coli* HisRS•tRNA pair for GCE [8, 18]. We sought to genetically modify *E. coli* to free *E. coli* HisRS from its essential role in translation by functionally replacing it with *C. crescentus* HisRS•tRNA^{His}. To accomplish this, we first integrated a copy of *C. crescentus* HisRS and tRNA^{His} under constitutive promoters (*glsS* and *lpp*, respectively) into a permissive site of the chromosome of *E. coli* TOP10 using a λ -Red recombination system to generate the *E. coli* MEOV0.1 strain. However, attempts to delete the *E. coli* HisRS genomic copy failed. A possible explanation is that the expression level of the *C. crescentus* HisRS•tRNA^{His} pair was not sufficient to sustain deletion of *E. coli hisS*. To overcome this, we introduced the *C. crescentus* HisRS•tRNA^{His} expression cassette in the high-copy plasmid pUC18, and using λ Red recombination, *E. coli* HisRS was deleted in *E. coli* MEOV0.1 to create *E. coli* MEOV0.2. We then tried to remove the single gene (*hisR*) encoding tRNA^{His} in *E. coli*, which is found in an operon with three other tRNA genes (*argX*, *leuT* and *proM*). We reasoned that deleting *hisR* may disrupt the expression of the entire locus, potentially leading to unintended consequences. To avoid this, *E. coli hisR* was converted to encode an amber suppressor tRNA^{His} by mutating the anticodon sequence (GTG to CTA), which produced the *E. coli* MEOV1 strain that exclusively relies on the activity of *C. crescentus* HisRS•tRNA^{His} for growth (Figure 1A).

3.2. Cell growth evaluation of MEOV1

A previous attempt to functionally replace an endogenous *E. coli* aaRS•tRNA pair with an orthogonal pair produced an *E. coli* strain with a severe growth defect [19]. Therefore, to ensure that the MEOV1 strain is a viable system for GCE applications, we compared its growth to that of the parental TOP10 strain. MEOV1 and TOP10 cells showed very similar doubling times (Figure 1B), indicating the ability of the *C. crescentus* HisRS•tRNA^{His} to effectively replace the *E. coli* pair in the engineered MEOV1 strain.

3.3. Probing the orthogonality of the Ec HisRS•tRNA^{His} pair in MEOV1 cells

To evaluate the orthogonality of *E. coli* HisRS•tRNA^{His} in the MEOV1 strain, we monitored the suppression of an amber codon at position 2 of super-folder green fluorescence protein

(sfGFP) mRNA by *E. coli* tRNA^{His}_{CUA}. Because MEOV1 encodes a copy of *E. coli hisR* with amber anticodon (tRNA^{His}_{CUA}), we wondered whether its background expression was sufficient to promote suppression of the amber codon in sfGFP (sfGFP2TAG). Therefore, a plasmid harboring sfGFP2TAG was introduced into MEOV1 together with a plasmid containing *E. coli* HisRS (Figure 2A). Surprisingly, induction of sfGFP2TAG and *E. coli* HisRS showed no expression of sfGFP was observed (Figure 2B), indicating that the expression of the chromosomal *E. coli hisR_{am}* (tRNA^{His}_{CUA}) gene in MEOV1 is not sufficient to promote suppression of the amber codon in sfGFP. To probe this hypothesis, a copy *E. coli hisR_{am}* under the constitutive promoter *lpp* was cloned into the plasmid containing sfGFP2TAG, and the resulting plasmid was introduced into MEOV1 cells together with *E. coli* HisRS. Expression of tRNA^{His}_{CUA} under the *lpp* promoter together with *E. coli* HisRS resulted in expression of sfGFP (Figure 2B). In addition to sfGFP, we used a chloramphenicol (Cm)-resistance-based assay to confirm the orthogonality of the *E. coli* HisRS•tRNA^{His} pair. In the chloramphenicol acetyltransferase (*cat*) gene of pACYC184, an amber stop codon at the permissive site 112 was introduced by site-directed mutagenesis, and the *E. coli* tRNA^{His}_{CUA} expression cassette was cloned into the same plasmid. MEOV1 cells expressing *E. coli* HisRS and tRNA^{His}_{CUA} grew to a Cm concentration of 50 µg/mL (Figure 2C), while cells expressing only tRNA^{His}_{CUA} did not survive even in low concentration of Cm. Together, these results demonstrate that the *E. coli* HisRS•tRNA^{His}_{CUA} pair is orthogonal in *E. coli* MEOV cells.

3.4. Suppression efficiency of *E. coli* HisRS•tRNA^{His}_{CUA}

Previously reported suppression efficiencies mediated by the *E. coli* HisRS•tRNA^{His}_{CUA} pair vary from 4% in *C. crescentus* [8] to 26–41% in *E. coli* [18]. To assess the *E. coli* HisRS•tRNA^{His}_{CUA} efficiency in MEOV cells, we compared the fluorescence intensity from MEOV1 cells expressing wild-type sfGFP to cells expressing sfGFP2TAG in the presence of the *E. coli* HisRS•tRNA^{His}_{CUA} pair (Figure 2A). Our results show that *E. coli* HisRS•tRNA^{His}_{CUA} suppresses the amber codon in sfGFP with 50% efficiency relative to wild type (Figure 3), producing up to 7.5 mg of sfGFP per liter of MEOV1 culture. Additionally, we investigated the possibility of using *E. coli* tRNA^{His} as an opal (UCA) or ochre (UUA) suppressor. Using sfGFP with either TGA or TAA at position 2, we found that *E. coli* tRNA^{His}_{UUA} suppresses an ochre codon with 33% efficiency relative to WTsfGFP fluorescence (Table 1). In contrast, tRNA^{His}_{UCA} only poorly suppressed the opal codon (~6% efficiency). Based on these results, it appears that anticodon base U34 plays an important role in tRNA^{His} recognition by *E. coli* HisRS. Moreover, the anticodon changes did not affect the orthogonality of *E. coli* tRNA^{His} in MEOV1 as sfGFP fluorescence was only observed when both HisRS and the corresponding tRNA^{His} suppressor were expressed. However, while we observed high suppression efficiency by the *E. coli* HisRS•tRNA pair in the sfGFP experiments (Figure 3), the suppression efficiency appeared to be weaker in the chloramphenicol-based assays (Figure 2C). We sought to improve the suppression efficiency by introducing a mutation in the anticodon loop (U38 to A) of tRNA^{His}, which was previously reported to increase its suppression efficiency by up to 20 fold [20]. We introduced this mutation in *E. coli* tRNA^{His}_{CUA} and found that cells with the mutant tRNA were able to grow to chloramphenicol concentrations two times higher than the “wild-type” tRNA^{His}_{CUA} (up to 100 µg/mL). Overall, these results show that *E. coli* HisRS•tRNA^{His}_{CUA}

is an effective pair for stop codon suppression, although its suppression effectiveness varies depending on the targeted stop codon. However, low suppression efficiency can be countered by introducing a mutation in the anticodon loop of tRNA^{His}.

3.5. Dissecting the *E. coli* HisRS active site

After demonstrating that the *E. coli* HisRS•tRNA^{His} pair is orthogonal in MEOV1 cells, it can now be subjected to directed evolution methods to potentially generate novel HisRS variants with desired amino acid specificities. This process requires the randomization of HisRS residues involved in amino acid binding. Although there is a wealth of structural and biochemical information regarding the active site of HisRS that can be used to target specific residues, most of the data was obtained using *in vitro* experiments [21, 22]. Thus, we decided to explore the contribution of different *E. coli* HisRS residues *in vivo* using the Cm-based reporter assay. An analysis of the crystal structure of *E. coli* HisRS bound to His revealed a series of residues that has been predicted to be involved in binding (Figure 4A) [21]. E83 forms a salt bridge with α -amino group of His, and mutating the residue to Ala resulted in a 4500-fold decrease in the aminoacylation rate *in vitro* [22]. However, *in vivo*, the E83A HisRS mutant only showed a small growth defect in MEOV cells (Figure 4B). T85 helps to maintain the position of E131 through hydrogen bonds. However, its role aminoacylation seems dispensable since a T85A HisRS mutant grew as well as WT HisRS (Figure 4B). In accordance with previous work showing that E131 and Y264 determined the specificity of the His imidazole ring, we found that E131A and Y264A HisRS mutants were inactive (Figure 4B). Intrigued by these results, we decided to investigate whether these mutations in HisRS have an impact in its amino acid specificity. To test this hypothesis, the Cm-based assays was used in the presence of the His analogs: 2-thio-His (2SH), D-His, and 3-methyl-His (3MeH) (Figures 4C and 5A). Interestingly, neither E131 nor Y264 HisRS were able to support growth on Cm (Figure 4C), suggesting that these residues are essential for enzymatic activity of HisRS.

3.6. Substrate selectivity of *E. coli* WT HisRS

To gain more insights into the amino acid substrate specificity of *E. coli* HisRS, we adapted a recently established amino acid screen that allows high-throughput assessment of aaRS specificity *in vivo* [12, 15]. This approach is based on the suppression of an amber codon in sfGFP by plasmid-encoded *E. coli* tRNA^{His}_{CUA} and endogenously expressed HisRS in a His-auxotrophic *E. coli* strain grown in minimal media in the presence of a particular amino acid. The library consists of 313 amino acids, including canonical AAs and several His analogs, distributed in a 384-well plate (Figure 5A, Supplementary Information) [12]. To validate this experimental setup, we tested the expression of WT sfGFP and *E. coli* tRNA^{His}. As expected, the strongest fluorescence signal was observed from cells grown with His (Supplementary Figure 1). Only one additional fluorescence peak appeared when cells expressed sfGFP2TAG and tRNA^{His}_{CUA} (Figure 5B), which originated from cells grown with 2SH. These results suggest that *E. coli* HisRS has a stringent amino acid selection mechanism (Figure 5B). However, our *in vivo* results contrast with previous reports that indicated that other His analogs present in our library are substrates of HisRS [23–25]. Among those His analogs are 1-methyl- (1MeH) and 3-methyl-His. To corroborate the results from the amino acid library, we carried out *in vitro* aminoacylation assays.

Interestingly, in contrast to the *in vivo* results, purified *E. coli* HisRS aminoacylated tRNA^{His} with both 1MeH and 3MeH with similar efficiency as cognate His (Figure 5C). This discrepancy between *in vivo* and *in vitro* experiments might be explained by how well *E. coli* cells can tolerate widespread incorporation of these analogs into the proteome.

4. Conclusions

In this work, we were able to delete the essential *hisS* (HisRS) gene from *E. coli* by functionally replacing its activity with that of *C. crescentus* HisRS•tRNA^{His}. In addition, we showed that the *E. coli* HisRS•tRNA^{His} pair is orthogonal in MEOV1 cells demonstrating that MEOV1 can serve as a platform for the development of the *E. coli* HisRS•tRNA^{His} pair for genetic code expansion applications. Construction of *E. coli* HisRS libraries are being developed, which will be subjected to multiple rounds of selection and screening aimed at finding HisRS variants with desired ncAA-specificity. A drawback of our engineering system is that the resulting *E. coli* HisRS variants may only be used in a limited number of organisms (*C. crescentus* and MEOV1), due to the frequent occurrence of organisms encoding G₁-containing tRNA^{His} genes. To counter this limitation, the tRNA^{His} substrate specificity of *E. coli* HisRS could be altered, which has been an useful approach in the development of other aaRS•tRNA pairs [26].

Acknowledgments

We are grateful to Drs. Takahito Mukai and Ana Crnković for enlightening discussions. This work was supported by a grant from the National Institute of General Medical Sciences [GM22854 to DS].

References

1. Chin JW. Expanding and reprogramming the genetic code of cells and animals. *Annu Rev Biochem.* 2014; 83:379–408. [PubMed: 24555827]
2. O'Donoghue P, Ling J, Wang Y-S, Söll D. Upgrading protein synthesis for synthetic biology. *Nature chemical biology.* 2013; 9:594–598. [PubMed: 24045798]
3. Dumas A, Lercher L, Spicer CD, Davis BG. Designing logical codon reassignment—Expanding the chemistry in biology. *Chem Sci.* 2015; 6:50–69. [PubMed: 28553457]
4. Wan W, Tharp JM, Liu WR. Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics.* 2014; 1844:1059–1070. [PubMed: 24631543]
5. Wang K, Fredens J, Brunner SF, Kim SH, Chia T, Chin JW. Defining synonymous codon compression schemes by genome recoding. *Nature.* 2016; 539:59–64. [PubMed: 27776354]
6. Ostrov N, Landon M, Guell M, Kuznetsov G, Teramoto J, Cervantes N, Zhou M, Singh K, Napolitano MG, Moosburner M, Shrock E, Pruitt BW, Conway N, Goodman DB, Gardner CL, Tyree G, Gonzales A, Wanner BL, Norville JE, Lajoie MJ, Church GM. Design, synthesis, and testing toward a 57-codon genome. *Science.* 2016; 353:819–822. [PubMed: 27540174]
7. Yuan J, Gogakos T, Babina AM, Soll D, Randau L. Change of tRNA identity leads to a divergent orthogonal histidyl-tRNA synthetase/tRNA^{His} pair. *Nucleic Acids Res.* 2011; 39:2286–2293. [PubMed: 21087993]
8. Ko JH, Llopis PM, Heinritz J, Jacobs-Wagner C, Soll D. Suppression of amber codons in *Caulobacter crescentus* by the orthogonal *Escherichia coli* histidyl-tRNA synthetase/tRNA^{His} pair. *PLoS One.* 2013; 8:e83630. [PubMed: 24386240]
9. Himeno H, Hasegawa T, Ueda T, Watanabe K, Miura K, Shimizu M. Role of the extra G-C pair at the end of the acceptor stem of tRNA^{His} in aminoacylation. *Nucleic acids research.* 1989; 17:7855–7863. [PubMed: 2678006]

10. Umehara T, Kim J, Lee S, Guo LT, Soll D, Park HS. N-acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess N-acetyl lysine specificity in vitro and in vivo. *FEBS Lett.* 2012; 586:729–733. [PubMed: 22289181]
11. Kitagawa M, Ara T, Arifuzzaman M, Ioka-Nakamichi T, Inamoto E, Toyonaga H, Mori H. Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res.* 2005; 12:291–299. [PubMed: 16769691]
12. Fan C, Ho JM, Chirathivat N, Soll D, Wang YS. Exploring the substrate range of wild-type aminoacyl-tRNA synthetases. *Chembiochem.* 2014; 15:1805–1809. [PubMed: 24890918]
13. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A.* 2000; 97:6640–6645. [PubMed: 10829079]
14. Datta S, Costantino N, Court DL. A set of recombineering plasmids for gram-negative bacteria. *Gene.* 2006; 379:109–115. [PubMed: 16750601]
15. Ko, J-h, Wang, Y-S., Nakamura, A., Guo, L-T., Söll, D., Umehara, T. Pyrrolysyl-tRNA synthetase variants reveal ancestral aminoacylation function. *FEBS Lett.* 2013; 587:3243–3248. [PubMed: 23994531]
16. Das M, Vargas-Rodriguez O, Goto Y, Suga H, Musier-Forsyth K. Distinct tRNA recognition strategies used by a homologous family of editing domains prevent mistranslation. *Nucleic Acids Res.* 2014; 42:3943–3953. [PubMed: 24371276]
17. Guth EC, Francklyn CS. Kinetic discrimination of tRNA identity by the conserved motif 2 loop of a class II aminoacyl-tRNA synthetase. *Mol Cell.* 2007; 25:531–542. [PubMed: 17317626]
18. Yan W, Francklyn C. Cytosine 73 is a discriminator nucleotide in vivo for histidyl-tRNA in *Escherichia coli*. *J Biol Chem.* 1994; 269:10022–10027. [PubMed: 8144499]
19. Iraha F, Oki K, Kobayashi T, Ohno S, Yokogawa T, Nishikawa K, Yokoyama S, Sakamoto K. Functional replacement of the endogenous tyrosyl-tRNA synthetase-tRNA^{Tyr} pair by the archaeal tyrosine pair in *Escherichia coli* for genetic code expansion. *Nucleic Acids Res.* 2010; 38:3682–3691. [PubMed: 20159998]
20. Kleina LG, Masson JM, Normanly J, Abelson J, Miller JH. Construction of *Escherichia coli* amber suppressor tRNA genes. II. Synthesis of additional tRNA genes and improvement of suppressor efficiency. *J Mol Biol.* 1990; 213:705–717. [PubMed: 2193162]
21. Arnez JG, Augustine JG, Moras D, Francklyn CS. The first step of aminoacylation at the atomic level in histidyl-tRNA synthetase. *Proc Natl Acad Sci U S A.* 1997; 94:7144–7149. [PubMed: 9207058]
22. Guth E, Connolly SH, Bovee M, Francklyn CS. A substrate-assisted concerted mechanism for aminoacylation by a class II aminoacyl-tRNA synthetase. *Biochemistry.* 2005; 44:3785–3794. [PubMed: 15751955]
23. Freist W, Verhey JF, Ruhlmann A, Gauss DH, Arnez JG. Histidyl-tRNA synthetase. *Biol Chem.* 1999; 380:623–646. [PubMed: 10430027]
24. Ikeda Y, Kawahara Si, Taki M, Kuno A, Hasegawa T, Taira K. Synthesis of a novel histidine analogue and its efficient incorporation into a protein in vivo. *Protein eng.* 2003; 16:699–706. [PubMed: 14560056]
25. Schlesinger S, Schlesinger MJ. The effect of amino acid analogues on alkaline phosphatase formation in *Escherichia coli* K-12 : I. Substitution of triazolealanine for histidine. *J Biol Chem.* 1967; 242:3369–3372. [PubMed: 5339728]
26. Chatterjee A, Xiao H, Schultz PG. Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in *Escherichia coli*. *Proc Natl Acad Sci U S A.* 2012; 109:14841–14846. [PubMed: 22927411]

Highlights

- The *C. crescentus* HisRS•tRNA^{His} pair functionally complements an *E. coli* *hisS* strain.
- The *E. coli* HisRS•tRNA^{His} pair is orthogonal and mediates suppression of UAG stop codons in *E. coli* MEOV1 cells.
- *E. coli* HisRS active site residues essential for *in vivo* activity were identified.

Figure 1.

Design and growth rate of *E. coli* MEOV1. **A.** Schematic representation of MEOV1 cells. A copy of the *C. crescentus* HisRS and tRNA^{His} expression cassette (shown in the blue box as *Cc hisST*) was introduced into the *E. coli* TOP10 chromosome to give MEOV0.1. The high-copy plasmid pUC18 harboring a copy of the *C. crescentus* HisRS and tRNA^{His} expression cassette was transformed into MEOV0.1 followed by deletion of the *E. coli* HisRS gene (*hisS*, shown in dark green box) from the chromosome and mutagenesis of anticodon (GUG to CUA) of the *E. coli* tRNA^{His} gene (*hisR*, shown in light green box). **B.** Growth curves of *E. coli* TOP10 (black circles) and MEOV1 (red squares).

Figure 2.

Reintroducing *E. coli* HisRS•tRNA^{His} as an orthogonal pair in MEOV1 cells. **A.** *E. coli* tRNA^{His}_{CUA} was introduced in the same plasmid with sfGFP (reporter), while *E. coli* HisRS was introduced in a separate plasmid. **B.** MEOV1 cells harboring *E. coli* HisRS and either wild-type (WT) or sfGFP with a TAG codon at position 2. Cells were grown overnight and spotted on LB-agar. The plate was incubated overnight at 37 °C. “UV” indicates sfGFP fluorescence, whereas “white” corresponds to the growth on the plate visualized with white light. **C.** MEOV1 cells harboring a plasmid encoding the *cat* gene with an amber codon at position 112 were transformed with a plasmid with *E. coli* HisRS. Serial dilutions (OD₆₀₀ 0.1, 0.01, and 0.001) of overnight grown cell cultures were spotted on LB/agar plates with different concentration of chloramphenicol (Cm). **D.** Cloverleaf structure of *E. coli* tRNA^{His}. The arrow indicates position 37, which was mutated to adenosine.

Figure 3. MEOV1 cells harboring *E. coli* HisRS and either WT or sfGFP with an amber codon at position 2 (sfGFP2TAG) were grown overnight, diluted to an OD₆₀₀ of 0.01, and grown for 24 h at 37 °C. Cell grow and fluorescence were monitored using a plate reader.

Figure 4.

Mutational analysis of *E. coli* HisRS active site residues. **A.** The structure of *E. coli* HisRS active site with Histidine (PDB: 1KMM). The HisRS residues that interact with the imidazole ring and amino group of His are shown. These residues were mutated to Ala. **B.** *In vivo* activity of *E. coli* HisRS mutants using the Cm-based assay as described for Figure 2C. **C.** Substrate specificity of HisRS mutants. Cm-based was used to explore potential changes in the substrate specificity of HisRS upon mutations. 1 mM 3mHis, 1 mM 2SH and 2 mM D-His were used. “Empty” and “-” indicates plasmid without HisRS and no amino acid added, respectively.

Figure 5.

E. coli HisRS substrate specificity *in vitro* and *in vivo*. **A.** Molecular structure of histidine analogs. aMeH (DL- α -methyl-His), 1MeH (1-methyl-His), 3MeH (3-methyl-His), BzH (Benzyl-L-His), 4ThzA (3-[4-Thiazolyl]-L-alanine), 2ThzA (2-[Thiazolyl]-L-alanine), TriAza (1,2,4-Triazole-alanine), 2FUA (2-Furyl)-L-alanine, Cyc5A (3-Cyclopentane-L-alanine), 2SH (2-Thio-L-His), and 3ThiA (3-[3-Thienyl]-DL-alanine). The letter and number on each amino acid indicate the position in the 384-well plate. **B.** *E. coli* HisRS substrate specificity profile. 313 distinct amino acids (1 mM) were distributed in a 384-well plate. 50 μ L of a His-auxotrophic *E. coli* strain JW2004-1 (*hisB*⁻) culture harboring sfGFP2TAG and *E. coli* tRNA^{His}_{CUA} in minimal medium were added to each well and the plate was incubated at 37 °C with constant shaking. Cell growth and fluorescence were monitored and used to estimate relative fluorescence after 24 h. **C.** *In vitro* aminoacylation assays were carried out with *E. coli* HisRS and tRNA^{His} at 37 °C with 1 mM amino acid concentration. Each time point represents the average of three independent trials with the corresponding standard deviation shown.

Table 1

Suppression efficiency of amber (UAG), ochre (UAA), and opal (UGA) codons at position 2 of sfGFP by the *E. coli* HisRS and tRNA^{His} with the corresponding anticodon sequence.

tRNA ^{His} _{XXX}	sfGFP2XXX	Suppression efficiency (%) [*]
GUG	CAC	-
CUA	UAG	44.3 ± 11
UUA	UAA	33.1 ± 2.5
UCA	UGA	5.9 ± 2.6

^{*} Suppression efficiency is relative to wild-type sfGFP. The percentage values are the result of three independent assays, and the standard deviation is shown.