Expanding the Genetic Code of Yeast for Incorporation of Diverse Unnatural Amino Acids via a Pyrrolysyl-tRNA Synthetase/tRNA Pair

Susan M. Hancock,† Rajendra Uprety,‡ Alexander Deiters,‡ and Jason W. Chin†,*

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 0QH, U.K., and North Carolina State University, Department of Chemistry, Raleigh, North Carolina 27695

Received May 27, 2010; E-mail: chin@mrc-lmb.cam.ac.uk

Abstract: We report the discovery of a simple system through which variant pyrrolysyl-tRNA synthetase/tRNA_CUA pairs created in Escherichia coli can be used to expand the genetic code of Saccharomyces cerevisiae. In the process we have solved the key challenges of producing a functional tRNA_CUA in yeast and discovered a pyrrolysyl-tRNA synthetase/tRNA_CUA pair that is orthogonal in yeast. Using our approach we have incorporated an alkyne-containing amino acid for click chemistry, an important post-translationally modified amino acid and one of its analogs, a photocaged amino acid and a photo-cross-linking amino acid into proteins in yeast. Extensions of our approach will allow the growing list of useful amino acids that have been incorporated in E. coli with variant pyrrolysyl-tRNA synthetase/tRNA_CUA pairs to be site-specifically incorporated into proteins in yeast.

Introduction

The pyrrolysyl-tRNA synthetase/tRNA_CUA (PylRS/tRNA_CUA) pairs from Methanosarcina barkeri (Mb) and M. mazei (Mm) are orthogonal in Escherichia coli.1 These pairs have been evolved to direct the site-specific incorporation of a range of unnatural amino acids, including amino acids that are post-translationally modified, amino acids containing bio-orthogonal chemical handles, and amino acids protected with light- and acid-sensitive groups, into proteins in E. coli in response to the amber codon.1–6 In contrast to other aminocyl-tRNA synthetase/tRNA pairs for the incorporation of unnatural amino acids, which are orthogonal in either eukaryotic or prokaryotic hosts, the PylRS/tRNA_CUA pairs are orthogonal in both E. coli and mammalian cells.2,6,7 Several unnatural amino acids have been site-specifically incorporated into proteins in mammalian cells by evolving the synthetase/tRNA pair in E. coli and subsequently transferring it to mammalian cells. This approach has the advantage of bypassing the requirement to evolve the amino acid specificity of the synthetase directly in a eukaryotic host.8–10

Many biological processes are more effectively addressed in the yeast Saccharomyces cerevisiae than in mammalian cells. Yeast has a rapid doubling time, bar-coded libraries of gene knockouts exist, protein interaction and transcriptome data is most complete, tap-tagged strains are readily available and powerful genetic approaches can be simply implemented. However, the requirement to evolve the current orthogonal pairs directly in yeast has limited the scope of unnatural amino acids that have been incorporated in yeast.

Preliminary work by Yokoyama and co-workers introduced a PylRS/tRNA_CUA pair into yeast and reported very weak phenotypes consistent with poor incorporation of Nε-tert-butyloxycarbonyl-L-lysine,7 but a properly characterized system for incorporating amino acids using PylRS/tRNA_CUA pairs has not been reported. Here we report the creation and characterization of a functional and orthogonal PylRS/tRNA_CUA pair in yeast.
and demonstrate the incorporation of several useful unnatural amino acids using variants of this pair created in *E. coli* (Figure 1).

**Results and Discussion**

To investigate the amber suppressor activity and potential orthogonality of the *MbPylRS/tRNACUA* pair in *S. cerevisiae* we used MaV203:pGADGAL4(2TAG) cells.8,9 This yeast strain contains a *GAL4* transcriptional activator gene bearing amber codons, is auxotrophic for histidine, and contains *HIS3* and *LacZ* genes on *GAL4*-activated promoters. When a functional amber suppression system, such as the *EcTyrRS/tRNACUA* Tyr pair,8,9 is transformed into this strain, full length *GAL4* is produced, leading to activation of *LacZ* and *HIS3* genes. Transcription of these genes allows cells to grow in the absence of histidine and turn blue in the presence of X-Gal.

We replaced the functional *EcTyrRS/tRNACUA* pair with the *MbPylRS/MbPyltRNA* pair (Figure 2B construct 1) and supplemented with *N*-[(2-propynyloxy)carbonyl]-L-lysine (I) (Figure 1A, a known substrate for *MbPylRS*) in MaV203:pGADGAL4(2TAG). These cells were unable to grow in media lacking histidine and did not turn blue in the presence of X-Gal, suggesting that this original construct is not functional (Figure 2D). We demonstrated by western blot that the yeast codon-optimized *MbPylRS* was expressed in yeast cells (data not shown). However, analysis of northern blots indicated that *MbtRNA* was not transcribed from our initial construct (Figure 2C). Since the *EctRNA* gene contains the consensus A- and B-box RNA polymerase III promoter sequences within the structural gene, suggesting that this original construct is not functional (Figure 2D), it seemed likely that additional promoter elements would be required to direct the transcription of *MbtRNA*.

To address the challenge of creating new promoter elements to direct the transcription of *MbtRNA* we investigated strategies to introduce A- and B-box sequences into our tRNA expression construct. We first mutated the sequence of the *MbtRNA* gene to contain either near-consensus A box sequences (A11C/U15G/T24G, Figure 2B construct 2) or B box sequences (A56C, Figure 2B construct 3). Northern blot analysis demonstrated that the A56C mutation in the B box, led to very low but detectable levels of the mutant *MbtRNA* (Supporting Information Figure 1), while expression of the (A11C/U15G/T24G) mutant tRNA was not detectable by northern blot. However, when the A56C mutant of *MbtRNA* and *MbPylRS* were transferred to MaV203:pGADGAL4(2TAG) in the presence of 1, we did not observe phenotypes consistent with amber suppression (Figure 2D). This implies that the tRNA is transcribed but not correctly folded or processed, or that the mutation abolishes synthetase recognition. Combining the A- and B-box mutations (Figure 2B construct 4) led to low levels of detectable tRNA production (Supporting Information Figure 1), but did not give phenotypes consistent with amber suppression (Figure 2D).

Since enhancing the transcription of *MbtRNA* by mutation of the A- and B-box sequences within the structural gene did not produce a functional amber suppressor, we next investigated the potential of constructs that might augment the transcription of *MbtRNA* using extragenic sequences. The 5'-leader sequence of the yeast SNR52 primary transcript contains A- and B-box promoters that are post-transcriptionally removed to

---

(11) Edwards, H.; Schimmel, P. *Mol. Cell. Biol.* 1990, 10, 1633.
produce mature SNR52 snoRNA.12 A previous report suggested that adding 5′-SNR52 and 3′-SUP4 flanking sequences to EctDNAUCU′_Cua and EctDNAUCU′_Cua enhanced their amber suppression in yeast. When MbtRNAUCU′_Cua was cloned between 5′-SNR52 and 3′-SUP4 flanking sequences (Figure 2B construct 5), we could detect weak MbtRNAUCU′_Cua transcription by northern blot (Figure 2C), and when the cassette was transformed into MaV203;pGADGAL4(2TAG) containing MbtPyIRS and grown in the presence of 1, we observed blue coloration on X-Gal plates but not growth in the absence of histidine in the presence of 40 mM 3-aminotriazole (3AT) (Figure 2D). These data suggest that addition of extragenic A- and B-box sequences via the 5′-SNR52 and 3′-SUP4 flanking sequences can partially compensate for the absence of consensus A- and B-box sequences in MbtRNAUCU′_Cua. However, since the EcTyrRS/tRNAUCU′_Cua orthogonal pair supports growth on media lacking histidine and containing 40 mM 3AT,9,9 this system does not, we decided that the system was suboptimal and opted to explore further extragenic sequences.

The yeast U6 (SNR6) gene assembles the same RNA polymerase III transcriptional machinery as tRNA genes but possesses an additional TATA-box promoter element 30 base pairs upstream of the transcription start site that binds TFIIIB.14 The TATA-box enables TFIIIC-independent RNA polymerase III recruitment and is proposed to overcome the large separation (240 bp) of the A- and B-box promoter elements of this gene.15 Several yeast tRNAs, some of which contain large introns between the A- and B-boxes, have TATA boxes that allow TFIIIC-independent RNA polymerase transcription.15 We reasoned that by incorporating the flanking sequences of these genes into our tRNA cassettes it may be possible to compensate for the poor A- and B-box consensus of MbtRNAUCU′_Cua. We created constructs where the 5′-flanking region of SNR6, Ile[TAT]_LRI, Pro[TGG]_FL, and Asp[GTC]_KR and the 3′-flanking region of SNR6 sandwich MbtRNAUCU′_Cua (Figure 2B constructs 6a–d). We also added a consensus sequence16 found at the transcription start site of yeast tRNAs to the SNR6 construct. Northern blots revealed low-level tRNA production from construct 6a (Supporting Information Figure 1). However, we did not observe phenotypes consistent with amber suppression when any of these constructs were transformed into MaV203:pGADGAL4(2TAG) containing MbtPyIRS and grown in the presence of 1 (Figure 2D). These data suggested that, while these promoter elements may compensate for increases in the A- and B-box spacing, they cannot efficiently compensate for defects in the A- and B-box sequence in MbtRNAUCU′_Cua.

Yeast possess an unusual dicistronic tDNAUCU′_Cua−tDNAUCU′_Cua gene in which the two mature tRNAs are generated from a single precursor RNA.17 The A- and B-box promoter in the tDNAUCU′_Cua gene directs the transcription of the precursor, and the transcription of tDNAUCU′_Cua is entirely dependent on the promoter elements of tDNAUCU′_Cua. This suggests that tDNAUCU′_Cua may provide the A- and B-box sequences required to transcribe tDNAs inserted in place of tDNAUCU′_Cua.18 Indeed replacing tDNAUCU′_Cua with human initiator tDNA or a transcriptionally inactive yeast tyrosine suppressor tDNA allows these tDNAs to be transcribed and processed to produce functional tRNAs in yeast.19

To test this system for the transcription of MbtRNAUCU′_Cua we constructed a SetDNAUCU′_Cua−MbtDNAUCU′_Cua cassette containing the natural 5′-, 3′-, and 10-base pair linker sequences (Figure 2B construct 7). Northern blot analysis revealed that MbtRNAUCU′_Cua was transcribed from this construct much more efficiently than any other construct tested (Figure 2C). When transformed into MaV203:pGADGAL4(2TAG) in the presence of MbtPyIRS and 1, the SetDNAUCU′_Cua−MbtDNAUCU′_Cua cassette conferred survival on media lacking histidine and containing 40 mM 3AT, and produced the strongest blue color of any construct tested when incubated with X-Gal (Figure 2D).

The tRNA constructs we discovered that are both transcribed (as judged by northern blot) and functional (as judged by phenotyping (constructs 5 and 7)) showed amber suppression
To begin to demonstrate the range of amino acids that can be incorporated in yeast using our approach, we incorporated the important post-translational modification \( \text{N}_{\text{r}} \)-acetyl-l-lysine (2) and its analog \( \text{N}_{\text{r}} \)-trifluoroacetyl-l-lysine (3), a photocaged lysine derivative \( \text{N}_{\text{r}} \)-[1-(6-nitrobenzo[1,3]dioxol-5-yl)ethoxy]-carbonyl]-l-lysine (4), and photo-cross-linker \( \text{N}_{\text{r}} \)-[2-(3-methyl-3H-diazirin-3-yl)ethoxy]carbonyl]-l-lysine (5) into hSOD-His\( \text{u} \), produced in \( \text{S. cerevisiae} \) (Figure 4A) using \( \text{MbpPyIRS} \) and variants of \( \text{MbpPyIRS} \) we have previously evolved in \( \text{E. coli} \).\(^{1,3} \) While we have not specifically evolved a synthetase for \( \text{N}_{\text{r}} \)-trifluoroacetyl-l-lysine, we have found that AcKRS2,\(^{1} \) previously evolved for incorporating \( \text{N}_{\text{r}} \)-acetyl-l-lysine, efficiently incorporates this amino acid. We demonstrated the incorporation of each amino acid by western blot (Figure 4A). We carried out large-scale expression and purification of hSOD in the presence of 1, 2, and 3 (Figure 4B), which unlike 4 and 5 are not photosensitive and are available in gram quantities, to further confirm the site and identity of amino acid incorporation by ESI-MS and MS/MS sequencing (Figure 4C–H). We have demonstrated the specific incorporation of an amino acid into SOD in the presence of 4 and 5. In addition we have reported MS and MS/MS data for the incorporation of amino acids 4 and 5 into proteins in other organisms.\(^{2,23} \) However, we have not yet obtained MS data directly in yeast and cannot rule out the possibility that an aspect of yeast metabolism—that is not conserved in either other eukaryotes or bacteria—leads to the selective post-translational modification of these amino acids in \textit{vivo}. Purified hSOD yields were 30–100 \( \mu \)g/L of yeast culture which is similar to the 50 \( \mu \)g/L yield reported for incorporating \( p \)-acetyl-l-phenylalanine into hSOD using the \( \text{E. coli} \) yeast RNA synthetase in yeast.\(^{9} \)

Conclusions

In summary, we have solved the key challenges of producing a functional and orthogonal tRNA\( \text{u} \)tRNA\( \text{C} \) pair that is orthogonal in yeast, and described a simple system through which variant \( \text{MbpPyIRS} \) pairs created in \( \text{E. coli} \) can be transplanted to expand the genetic code of yeast for a wide range of unnatural amino acids. Using our approach we have incorporated the alkyl-containing amino acid \( \text{N}_{\text{r}} \)-[(2-propionyloxy)carbonyl]-l-lysine (1), an important post-translationally modified amino acid \( \text{N}_{\text{r}} \)-acetyl-l-lysine (2), and an analog of \( \text{N}_{\text{r}} \)-acetyl-l-lysine, trifluoroacetyl-l-lysine (3), a photocaged lysine derivative \( \text{N}_{\text{r}} \)-[(1-(6-nitrobenzo[1,3]dioxol-5-yl)ethoxy]carbonyl]-l-lysine (4), and a photo-cross-linker \( \text{N}_{\text{r}} \)-[2-(3-methyl-3H-diazirin-3-yl)ethoxy]carbonyl]-l-lysine (5) into proteins in yeast. Amino acid 1 may be used for bio-orthogonal [3 + 2] cycloadditions.

References

(20) Giege, R.; Sissler, M.; Florentz, C. \textit{Nucleic Acids Res.} 1998, 26, 5017.
(21) (a) Hou, Y. M.; Schimmel, P. \textit{Nature} 1988, 335, 140. (b) McClain, W. H.; Foss, K. \textit{Science} 1988, 240, 793.
(22) Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G. \textit{Proc. Natl. Acad. Sci. U.S.A.} 2006, 103, 9785.
(23) Chou, C.; Upreat, R.; Davis, L.; Chin, J. W.; Deiters, A. Submitted 2010.
Amino acid 2 may be used for producing acetylated proteins directly in yeast and synthetically controlling processes normally regulated by acetylation in yeast. Amino acid 3 is a very poor substrate for sirtuins, but not for HDACs, and should allow us to install irreversible acetylation at sites directly regulated by sirtuins in Vivo. It should allow us to probe the deacetylases that act on a given site in a protein. Amino acid 4 is a photocaged lysine with demonstrated utility for controlling protein function in eukaryotic cells, and we anticipate that genetically encoded photocontrol of proteins in yeast will be a powerful approach for gaining a temporal and spatial understanding of cellular processes. Amino acid 5 is a photocross-linking amino acid with demonstrated utility for mapping protein interactions in E. coli, and we believe that this will find wide utility in mapping protein–protein interactions in yeast. Given the growing list of amino acids that can be incorporated using MbPylRS and its variants, we anticipate that our approach will allow the introduction of a wide range of chemical functional groups into yeast. Finally, the strategies we have explored for creating and expressing heterologous, orthogonal tRNAs in yeast may be useful for improving other orthogonal aminoacyl-tRNA synthetase/tRNA CuA pairs.

**Experimental Section**

**General Methods.** Nε-[2-(Propynloyx)carbonyl]-L-lysine, Nε-[1-(6-nitrobenzod[d][1,3]dioxol-5-yl)ethoxy]carbonyl]-L-lysine and Nε-[2-(3-methyl-3H-diazimin-3-yl)ethoxy]carbonyl]-L-lysine were synthesized as previously reported. Nε-Acetyl-L-lysine and Nε-trifluoroacetyl-L-lysine were purchased from Bachem.

**Northern Blot Analysis.** Total RNA was purified from yeast cells using TRI reagent (Sigma) and ethanol precipitated. The RNA was denatured, separated on a 6% Novex TBE-urea gel (Invitrogen), blotted onto Bio-dyne B modified nylon membrane (Thermo Scientific), and cross-linked by UV fixation. The membrane was hybridized overnight at 55°C with a biotinylated probe S'-.
GGAAACCCGGGAATCTAACCCGGCTGAACGGATTTAG-
AG, which is specific for \( MbtDNA_{\text{CUL}}^{\text{Pyl}} \). The hybridized probe was
detected with North2South chemiluminescent hybridization and
detection kit (Pierce). The number of cells was used to control the
total amount of RNA loaded.

**Phenotyping Yeast Cells.** Phenotyping was performed as
described in Chin et al.\(^8\) Briefly, \( S. cerevisiae \) MaV203 (Invitrogen)
was transformed by the lithium acetate method with the pGADGAL4-
(2TAG) reporter, \( pMbp_{\text{Pyl}} \) and \( tDNACUL^{\text{Pyl}} \) constructs. Overnight
cultures were serially diluted and replica plated onto selective media
in the presence or absence of 2 mM \( N_{\varepsilon}-[2\text{-propynyl}oxy]c\text{arbonyl}]-l\)-lysine (1). X-Gal assays were performed using the agarose overlay
method.

**Protein Expression, Purification, western blot Analysis, and
Mass Spectrometry.** Appropriate selective medium ± unnatural
amino acid was inoculated with a stationary phase culture to give
an \( O_{D_{600}} \approx 0.2 \). Cultures were grown at 30°C for 24–48 h. Proteins
were extracted from yeast cells using Y-PER reagent (Thermo
Scientific) containing complete, EDTA-free inhibitor cocktail
(Roche). Clarified supernatants were separated by SDS-PAGE, and
western blots were performed using anti-His6 (Qiagen). Human
superoxide dismutase was purified using \( Ni^{2+}\)NTA resin (Qiagen)
as previously described.\(^26\) For expressions with \( N_{\varepsilon}\)-acetyl-l-lysine
(2), 20 mM nicotinamide was added to the cultures and to lysis
buffers; for expressions with \( N_{\varepsilon}\)-trifluoroacetyl-l-lysine (3), 10 mM
sodium butyrate was added to the cultures and to lysis buffers.
Protein concentration was determined using the Biorad Protein
Assay in comparison to IgG standard. Total mass analysis was
performed on a LCT time-of-flight mass spectrometer with elec-
trospray ionization (Micromass) with protein solutions in 20 mM
ammonium bicarbonate and mixed 1:1 with 1% formic acid in 50%
MeOH. Samples were injected at 10 µL·min\(^{-1}\), and calibration was
performed in positive ion mode using horse heart myoglobin. MS/
MS analysis was performed on a LTQ-Orbitrap mass spectrometer
on protein samples that were in-gel digested with Glu-C (Roche).

**Acknowledgment.** We are grateful to Farida Begum and Sew
Peak-Chew (MRC-LMB) for mass spectrometry, and Heinz Neu-
mann and Sebastian Greiss for assistance. This work was supported
by the Medical Research Council (MRC) UK and a European
Research Council (ERC) StG to J.W.C.

**Supporting Information Available:** Supplementary figures,
details of plasmid construction, and construct sequences. This
material is available free of charge via the Internet at http://
pubs.acs.org.

\(^{26}\) Chen, S.; Schultz, P. G.; Brock, A. *J. Mol. Biol.* 2007, 371, 112.