

## Ribosomal Synthesis of Unnatural Peptides

Kristopher Josephson, Matthew C. T. Hartman, and Jack W. Szostak\*

Contribution from the Howard Hughes Medical Institute, and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114

Received March 11, 2005; E-mail: szostak@molbio.mgh.harvard.edu.

**Abstract:** Combinatorial libraries of nonbiological polymers and drug-like peptides could in principle be synthesized from unnatural amino acids by exploiting the broad substrate specificity of the ribosome. The ribosomal synthesis of such libraries would allow rare functional molecules to be identified using technologies developed for the in vitro selection of peptides and proteins. Here, we use a reconstituted *E. coli* translation system to simultaneously reassign 35 of the 61 sense codons to 12 unnatural amino acid analogues. This reprogrammed genetic code was used to direct the synthesis of a single peptide containing 10 different unnatural amino acids. This system is compatible with mRNA-display, enabling the synthesis of unnatural peptide libraries of  $10^{14}$  unique members for the in vitro selection of functional unnatural molecules. We also show that the chemical space sampled by these libraries can be expanded using mutant aminoacyl-tRNA synthetases for the incorporation of additional unnatural amino acids or by the specific posttranslational chemical derivitization of reactive groups with small molecules. This system represents a first step toward a platform for the synthesis by enzymatic tRNA aminoacylation and ribosomal translation of cyclic peptides comprised of unnatural amino acids that are similar to the nonribosomal peptides.

### Introduction

The nonribosomal peptides (NRPs) are a class of microbial natural products that includes several valuable small molecule therapeutics, such as the antibiotics penicillin and vancomycin, and the immunosuppressant cyclosporin A.<sup>1,2</sup> NRPs are synthesized nonribosomally by enormous multi-subunit complexes comprised of catalytic modules dedicated to the ordered addition and modification of each residue. The success of NRPs as small molecule therapeutics has been attributed to the fact that they typically contain nonproteinogenic amino acid analogues with modified side-chains and backbones that expand the limited chemical diversity of proteinogenic amino acids and can increase protease resistance and membrane permeability. Additionally, the structures of NRPs are frequently constrained by macrocyclization and side-chain heterocyclization, which reduces conformational entropy allowing for specific high affinity target binding as well as contributing to protease resistance and membrane permeability.

Several groups have proposed adapting in vitro bacterial translation for the coded ribosomal synthesis of drug-like peptides composed of nonproteinogenic or unnatural amino acids.<sup>3–6</sup> Because the synthesis of these compounds is coded for by mRNA, it will be possible to select functional molecules using in vitro selection technologies developed for the selection of peptides and proteins from complex mRNA libraries of up

to  $10^{14}$  unique sequences.<sup>7</sup> This approach will require rewriting the genetic code for the simultaneous translation of multiple unnatural amino acids and is contingent on the ability of the ribosome to accept a large number of analogues for the synthesis of libraries with good chemical diversity. However, methods for manipulating the genetic code have already been developed and employed to demonstrate that the ribosome is able to incorporate a large number of diverse side-chain analogues and will tolerate at least a few backbone modifications.

The most common approach to translating unnatural amino acids is to expand the genetic code by utilizing the amber nonsense (stop) codon for incorporation of an unnatural 21st amino acid. The unnatural amino acid is chemically esterified<sup>8</sup> to an amber suppressor-tRNA, and incorporated in otherwise natural proteins opposite amber (UAG) codons using in vitro translation systems derived from cellular extracts.<sup>9,10</sup> This approach has been widely used to demonstrate that the ribosome will translate chemically diverse amino acid analogues, including bulky side-chain groups and modified main-chains, such as *p*-benzoyl-L-Phe, (aminoxy)acetic acid, and *N*-methyl amino acids.<sup>11–13</sup> Schultz and co-workers have also engineered orthogonal pairs of suppressor-tRNAs and mutant aminoacyl-tRNA-synthetases (AARSs) for the in vivo incorporation of

- (1) Walsh, C. T. *Science* **2004**, *303*, 1805–1810.
- (2) Sieber, S. A.; Marahiel, M. A. *J. Bacteriol.* **2003**, *185*, 7036–7043.
- (3) Forster, A. C.; Tan, Z.; Nalam, M. N.; Lin, H.; Qu, H.; Cornish, V. W.; Blacklow, S. C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6353–6357.
- (4) Merryman, C.; Green, R. *Chem. Biol.* **2004**, *11*, 575–582.
- (5) Frankel, A.; Millward, S. W.; Roberts, R. W. *Chem. Biol.* **2003**, *10*, 1043–1050.
- (6) Bjorklund, M.; Koivunen, E. *Lett. Drug Des. Discovery* **2004**, *1*, 163–167.

- (7) Frankel, A.; Li, S.; Starck, S. R.; Roberts, R. W. *Curr. Opin. Struct. Biol.* **2003**, *13*, 506–512.
- (8) Heckler, T. G.; Chang, L. H.; Zama, Y.; Naka, T.; Chorghade, M. S.; Hecht, S. M. *Biochemistry* **1984**, *23*, 1468–1473.
- (9) Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182–188.
- (10) Bain, J. D.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R.; Diala, E. S. *J. Am. Chem. Soc.* **1989**, *111*, 8013–8014.
- (11) Hoshaka, T.; Kajihara, D.; Ashizuka, Y.; Murakami, H.; Sisido, M. *J. Am. Chem. Soc.* **1999**, *121*, 34–40.
- (12) Ellman, J. A.; Mendel, D.; Schultz, P. G. *Science* **1992**, *255*, 197–200.
- (13) Eisenhauer, B. M.; Hecht, S. M. *Biochemistry* **2002**, *41*, 11472–11478.

unnatural amino acids in response to a nonsense codon.<sup>14,15</sup> Nonsense suppression approaches are limited, however, to the simultaneous incorporation of only two different unnatural amino acids, because there are only three nonsense codons, and one must be reserved for proper translation termination. To circumvent this limitation, several groups have expanded the genetic code beyond the 64 three-residue codons by developing four- or even five-residue codons or three-residue codons with novel base pairs.<sup>16–21</sup>

An alternative to expanding the genetic code is to reassign sense codons to unnatural amino acids. In vivo, sense codons have been reassigned during recombinant protein expression by forcing bacterial expression hosts that are metabolically incapable of synthesizing a particular amino acid to use a chemically similar amino acid analogue provided in the culture media.<sup>22–25</sup> Sense codon reassignment with backbone modified analogues has also been demonstrated in vitro using chemically charged tRNAs. For example, Frankel et al. used chemically charged tRNAs to reassign a Val and an Ala codon to *N*-methyl-Phe and demonstrated the synthesis of protease resistant poly-*N*-methyl-Phe peptides.<sup>5</sup> Tan et al. used sense codon reassignment and chemically charged tRNAs to test the translation of several different backbone analogues of Phe and Ala and found that the *N*-methyl and  $\alpha$ -hydroxy analogues of these amino acids are efficiently incorporated at a single codon in a tripeptide.<sup>26</sup> An alternative approach to the chemical charging of unnatural amino acids is to chemically transform the natural amino acids after they have been enzymatically charged onto tRNA. For example, Phe-tRNA<sup>Phe</sup> has been chemically deaminated to  $\alpha$ -hydroxy-Phe-tRNA<sup>Phe</sup> for the ribosomal synthesis of the Phe-polyester.<sup>27</sup> Using a similar approach, Merryman and Green chemically methylated the amino groups of the 20 canonical amino acids while attached to their cognate tRNA, and demonstrated translation of *N*-methyl peptides.<sup>4</sup>

Recently, reconstituted translation systems have been assembled from highly purified *E. coli* ribosomes and tRNA plus recombinant translation factors and AARSs.<sup>28,29</sup> These purified systems are capable of translating entire proteins with yields comparable to S30 translation extracts. However, unlike crude

extracts, reconstituted systems offer complete control of the components that are critical for the manipulation of the genetic code, that is, tRNA, AARSs, amino acids, and release factors. Therefore, multiple sense codons can be simultaneously reassigned to unnatural amino acids by simply reconfiguring the components of the translation system. For example, by omitting amino acids and AARSs and providing three different chemically charged tRNAs, Forster et al. used a reconstituted translation system to simultaneously reassign three sense codons to unnatural amino acids.<sup>3</sup>

Ultimately, we envision using the PURE reconstituted *E. coli* translation system<sup>28</sup> and an expanded genetic code to synthesize chemically diverse libraries of NRP-like molecules in a fully enzymatic ribosome-based system. To explore the feasibility of using AARSs for the in situ aminoacylation of tRNAs with unnatural amino acids, followed by ribosomal synthesis of unnatural peptides, our initial efforts have focused on 12 unnatural amino acid analogues that are known AARS substrates. Here, we simultaneously translate 10 of the 12 analogues and show that this system is compatible with the synthesis of puromycin-mediated molecular fusions between the unnatural peptide and its mRNA for the in vitro selection of drug-like molecules by mRNA-display.<sup>30</sup> Finally, we show that the chemical diversity of these peptides can be expanded by reconfiguring the system with mutant AARSs or by the specific posttranslational chemical ligation of small molecules to one of the unnatural side-chains.

## Materials and Methods

**Amino Acids.** Natural amino acids, L-azetidine-2-carboxylic acid (P<sub>a</sub>), 2-fluoro-L-phenylalanine (F<sub>a</sub>), DL-7-azatryptophan (W<sub>a</sub>), DL- $\beta$ -(1,2,4-triazol-3-yl)-alanine (H<sub>a</sub>), DL-4-thiaisoleucine (I<sub>a</sub>), DL- $\beta$ -hydroxy-norvaline (T<sub>a</sub>), S-(2-aminoethyl)-L-cysteine (K<sub>a</sub>), L-canavanine (R<sub>a</sub>), L-allylglycine (L<sub>b</sub>), and L-*p*-iodo phenylalanine (F<sub>b</sub>) were purchased from Sigma. 5',5',5'-Trifluoroleucine (L<sub>a</sub>), 3-fluoro-L-tyrosine (Y<sub>a</sub>), and L-threo- $\beta$ -hydroxy aspartic acid (D<sub>a</sub>) were purchased from Lancaster, Fluka, and ICN, respectively. DL-2-Amino-hex-5-ynoic acid (M<sub>a</sub>) was kindly provided by J. Link and D. Tirrell. All amino acids were dissolved in water at 1, 10, or 40 mM, and the pH was adjusted to 7.0–7.5 with KOH and filtered. Isotopically labeled amino acids <sup>35</sup>S-Met (2600 dpm/fmol) and <sup>3</sup>H-His (84 dpm/fmol) were from Perkin-Elmer.

**Translation Factors, Enzymes, and Ribosomes.** PCR primers and plasmids for the expression of hexa-histidine tagged IF1, IF2, IF3, EF-Tu, EF-G, EF-Ts, RF1, RF3, RRF, MTF, MetRS, GluRS, PheRS, AspRS, SerRS, and ThrRS are described in the Supporting Information. Expression plasmids for ArgRS, GlnRS, IleRS, LeuRS, TrpRS, AsnRS, HisRS, and ProRS<sup>28</sup> were from T. Ueda, and TyrRS (Supporting Information) and LysRS<sup>31</sup> were from P. Schimmel. The PheRS Ala294Gly mutant was generated from the wild-type expression plasmid using the quick-change mutagenesis kit (Stratagene) and a pair of mutagenic primers. To construct the Asp345Ala mutant of LeuRS, the LeuRS gene was excised from pET21a with *Nde*I and *Xho*I and moved to the pET24a backbone, which lacks *Dra*I sites. The Asp345Ala mutation was created by PCR amplification of the *Age*I/*Dra*I fragment using a mutagenic forward primer spanning the *Age*I site and a reverse primer spanning the *Dra*I site, which were used to reintroduce the mutant fragment into the parental pET24a-LeuRS construct. The plasmid for the expression of tRNA-nucleotidyltransferase as a fusion with maltose binding protein was constructed by U. RajBahndary and provided by C. Merryman. The expression, purification, and storage

- (14) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498–500.
- (15) Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G. *J. Am. Chem. Soc.* **2004**, *126*, 14306–14307.
- (16) Anderson, J. C.; Wu, N.; Santoro, S. W.; Lakshman, V.; King, D. S.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7566–7571.
- (17) Hohsaka, T.; Ashizuka, Y.; Taira, H.; Murakami, H.; Sisido, M. *Biochemistry* **2001**, *40*, 11060–11064.
- (18) Magliery, T. J.; Anderson, J. C.; Schultz, P. G. *J. Mol. Biol.* **2001**, *307*, 755–769.
- (19) Hohsaka, T.; Ashizuka, Y.; Murakami, H.; Sisido, M. *Nucleic Acids Res.* **2001**, *29*, 3646–3651.
- (20) Bain, J. D.; Switzer, C.; Chamberlin, A. R.; Benner, S. A. *Nature* **1992**, *356*, 537–539.
- (21) Hirao, I.; Ohtsuki, T.; Fujiwara, T.; Mitsui, T.; Yokogawa, T.; Okuni, T.; Nakayama, H.; Takio, K.; Yabuki, T.; Kigawa, T.; Kodama, K.; Nishikawa, K.; Yokoyama, S. *Nat. Biotechnol.* **2002**, *20*, 177–182.
- (22) Hortin, G.; Boime, I. *Methods Enzymol.* **1983**, *96*, 777–784.
- (23) Wilson, M. J.; Hatfield, D. L. *Biochim. Biophys. Acta* **1984**, *781*, 205–215.
- (24) Link, A. J.; Mock, M. L.; Tirrell, D. A. *Curr. Opin. Biotechnol.* **2003**, *14*, 603–609.
- (25) Budisa, N.; Minks, C.; Alefelder, S.; Wenger, W.; Dong, F.; Moroder, L.; Huber, R. *FASEB J.* **1999**, *13*, 41–51.
- (26) Tan, Z. P.; Forster, A. C.; Blacklow, S. C.; Cornish, V. W. *J. Am. Chem. Soc.* **2004**, *126*, 12752–12753.
- (27) Fahnestock, S.; Rich, A. *Science* **1971**, *173*, 340–343.
- (28) Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. *Nat. Biotechnol.* **2001**, *19*, 751–755.
- (29) Forster, A. C.; Weissbach, H.; Blacklow, S. C. *Anal. Biochem.* **2001**, *297*, 60–70.

- (30) Roberts, R. W.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12297–12302.
- (31) Steer, B. A.; Schimmel, P. *Biochemistry* **1999**, *38*, 4965–4971.

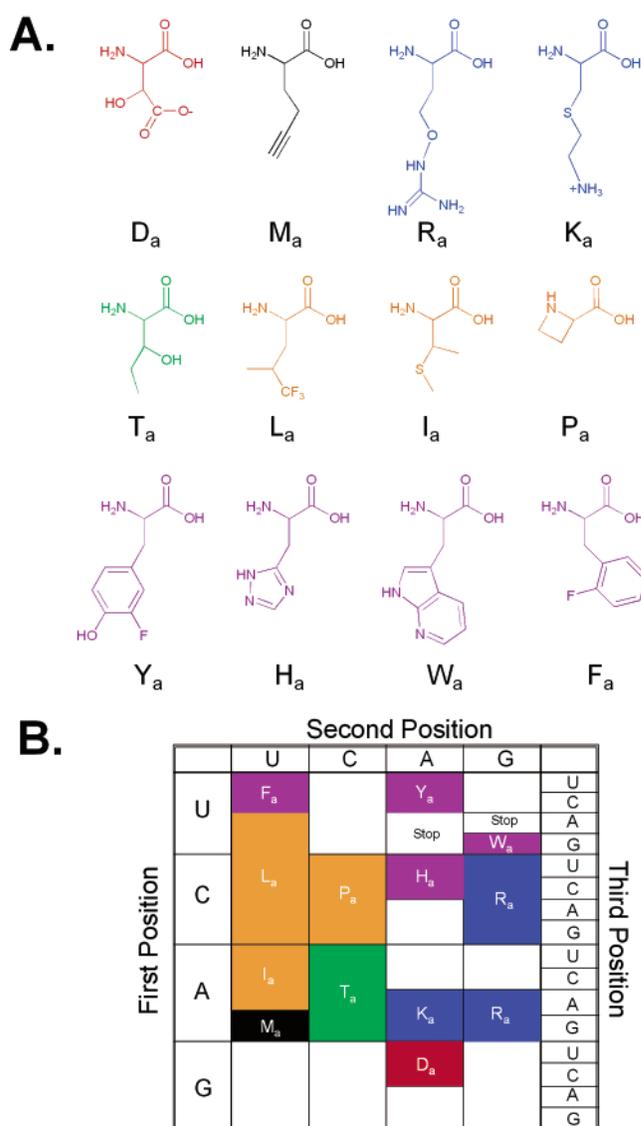
of all factors and enzymes followed a previously described protocol<sup>28</sup> with the following modifications: cells were lysed with B-PER reagent (Pierce) and purified using a batch procedure with Ni-NTA resin (Qiagen) as suggested by the manufacturer. Enzyme concentrations were determined from the UV absorbance at 280 nm and extinction coefficients calculated from amino acid compositions. MetRS required storage in buffer containing 50% glycerol at  $-20^{\circ}\text{C}$ .

Ribosomes were prepared at  $4^{\circ}\text{C}$  from *E. coli* strain A19. The cells from a 3 L log-phase culture were harvested by centrifugation at 5000g for 30 min and washed in 300 mL of buffer A (10 mM Tris-HCl pH 7.5, 10 mM Mg(OAc)<sub>2</sub>, 100 mM NH<sub>4</sub>Cl, 0.25 mM EDTA, 7 mM  $\beta$ -ME). The pelleted cells were resuspended in 25 mL of B-PER cell lysis solution (Pierce) supplemented with 4 mM Mg(OAc)<sub>2</sub>, 100 mM NH<sub>4</sub>Cl, 0.25 mM EDTA, 7 mM  $\beta$ -ME, and 200  $\mu\text{g}/\text{mL}$  lysozyme. The S30 fraction was completely clarified by three successive 30 min spins at 30 000g. Ribosomes were isolated from the approximately 20 mL of S30 fraction by layering 5 mL aliquots onto 7 mL of 30% sucrose in buffer B (10 mM Tris-HCl pH 7.5, 10 mM Mg(OAc)<sub>2</sub>, 500 mM NH<sub>4</sub>Cl, 7 mM  $\beta$ -ME) and pelleting for 18 h at 100 000g in a Beckman SW40Ti swinging bucket rotor. This salt washing procedure was repeated two more times by resuspending ribosomes in buffer B and pelleting through the 30% sucrose solution. Following the third salt wash, the ribosome pellets were resuspended in buffer C (10 mM Tris-HCl pH 7.5, 10 mM Mg(OAc)<sub>2</sub>, 60 mM NH<sub>4</sub>Cl, 0.5 mM EDTA, 3 mM  $\beta$ -ME), aliquoted for storage, and flash frozen in liquid nitrogen.

**tRNA Labeling and AARS Assay.** Total *E. coli* tRNA was labeled according to the method of Wolfson and Uhlenbeck<sup>32</sup> with several modifications. The labeling reaction (100  $\mu\text{L}$ ) contained 20 mM Tris pH 8.7, 20 mM MgCl<sub>2</sub>, 50  $\mu\text{M}$  sodium pyrophosphate, 1  $\mu\text{M}$  total tRNA (Roche), and 660 nM [ $\alpha$ -<sup>32</sup>P]ATP (227  $\mu\text{Ci}$ ) and 34 ng/mL *E. coli* tRNA nucleotidyl transferase. The reaction was incubated at  $37^{\circ}\text{C}$  for 2 h and quenched by the addition of 10  $\mu\text{L}$  of 100 units/mL yeast inorganic pyrophosphatase (Sigma). After 2 min at  $37^{\circ}\text{C}$ , CTP was added to a final concentration of 4.5  $\mu\text{M}$ . The solution was phenol:chloroform extracted and the tRNA purified from the aqueous layer on a NAP-5 column (Amersham Biosciences) eluted with water. A YM-10 centricon (Millipore) was used to concentrate the labeled tRNA, and residual [ $\alpha$ -<sup>32</sup>P]ATP was removed by two exchanges into water. The final tRNA labeling yield was 50  $\mu\text{Ci}$ .

Each AARS assay (10  $\mu\text{L}$ ) contained 40 mM HEPES-KOH pH 7.4, 45 mM KCl, 25 mM MgCl<sub>2</sub>, 3.4 mM BME, 6 mM ATP, 0.02 units/mL yeast inorganic pyrophosphatase, 245  $\mu\text{M}$  [ $\alpha$ -<sup>32</sup>P]tRNA (0.67  $\mu\text{Ci}$ ), BSA (0.09 mg/mL), 2 mM of the appropriate amino acid and AARS (2.1  $\mu\text{M}$  AspRS, 220 nM PheRS, 1.23  $\mu\text{M}$  HisRS, 280 nM IleRS, 1.07  $\mu\text{M}$  LysRS, 11.5 nM LeuRS, 910 nM MetRS, 230 nM ProRS, 550 nM ArgRS, 240 nM ThrRS, 92 nM TrpRS, and 290 nM TyrRS). After 5 min, a 1  $\mu\text{L}$  aliquot was removed and added to an ice-cold 4  $\mu\text{L}$  solution of nuclease P1 (0.066 unit/ $\mu\text{L}$ ) in 200 mM NaOAc pH 5.0. The quenched reactions were incubated for 20 min at ambient temperature. The solutions were spotted on a 10 cm PEI-cellulose TLC plate (Sigma, pre-run with distilled water), developed with 85:10:5 water:saturated boric acid:acetic acid, dried, and exposed overnight to a phosphorimager screen. TLC spots were quantified using a phosphorimager, and background was accounted for by subtracting a spot of equal size from a blank TLC lane.

**mRNA Templates.** DNA templates for the in vitro transcription of mRNAs that code for the peptides in Table 1 and Figure 4 were constructed by ligating complementary oligonucleotides with the appropriate overhangs into pET12b (Novagen), which had been opened with *Nde*I and *Bam*HI or *Nde*I and *Psi*II (Supporting Information). Plasmid DNA was prepared from ampicillin resistant transformants and used as the template in PCR reactions with primers that correspond to the T7 RNA polymerase promoter and terminator sequences. PCR products were purified using the QIAquick PCR purification kit (Qiagen) and used as the template for standard T7 RNA polymerase in vitro



**Figure 1.** Unnatural amino acids known to be translatable. (A) The structures are colored red (polar positive), blue (polar positive), purple (nonpolar aromatic), green (polar neutral), orange (nonpolar aliphatic), and black (unique reactivity) according to side-chain properties at physiological pH. The imidazole side-chain of His has a  $\text{pK}_a$  of  $\sim 6.0$  and is typically considered polar positive; however, because H<sub>a</sub> has a  $\text{pK}_a$  of  $\sim 3.3$ , it is classified as nonpolar aromatic. (B) The universal genetic code is shown with the sense codons reassigned to 12 unnatural amino acids shaded according to symbols and coloring in (A).

transcription reactions.<sup>33</sup> The desired transcript was purified by denaturing preparative polyacrylamide gel electrophoresis. For mRNA-display, purified mRNA was photochemically ligated at 366 nm for 15 min to a synthetic linker possessing a 5' psoralen and a 3' puromycin [5'-psoralen-UAGCCGGUG-dA<sub>15</sub>-(spacer)<sub>2</sub>-dA-rCrC-puromycin-3' (Glen Research), the italicized sequence corresponds to 2'-methoxy nucleosides].<sup>34</sup>

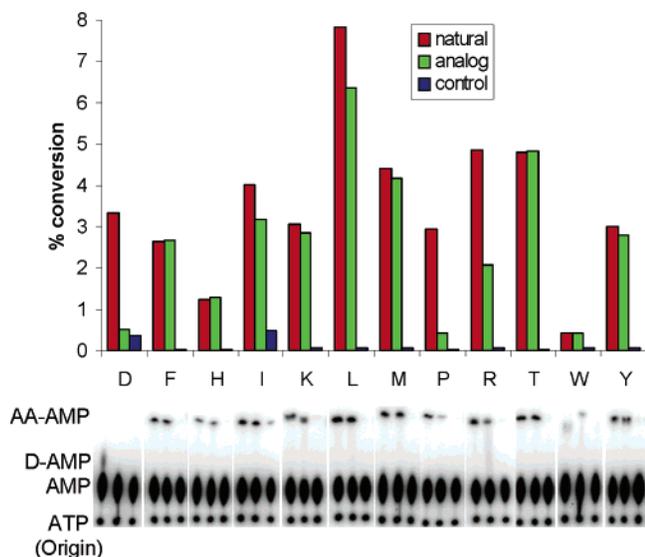
**PURE Translation System.** The PURE translation system was as previously described with slight modification.<sup>28</sup> Freshly prepared standard polymix buffer<sup>35</sup> was supplemented with 2 mM ATP, 2 mM GTP, 10 mM creatine phosphate, and 30  $\mu\text{M}$  10-formyl-5,6,7,8-tetrahydrofolic acid. Translation reactions (50  $\mu\text{L}$ ) contained final concentrations of 0.2  $\mu\text{M}$  MTF, 1.0  $\mu\text{M}$  IF1, 0.3  $\mu\text{M}$  IF2, 0.7  $\mu\text{M}$  IF3,

(33) Milligan, J. F.; Uhlenbeck, O. C. *Methods Enzymol.* **1989**, *180*, 51–62.

(34) Kurz, M.; Gu, K.; Lohse, P. A. *Nucleic Acids Res.* **2000**, *28*, E83.

(35) Jelenc, P. C.; Kurland, C. G. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 3174–3178.

(32) Wolfson, A. D.; Uhlenbeck, O. C. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5965–5970.

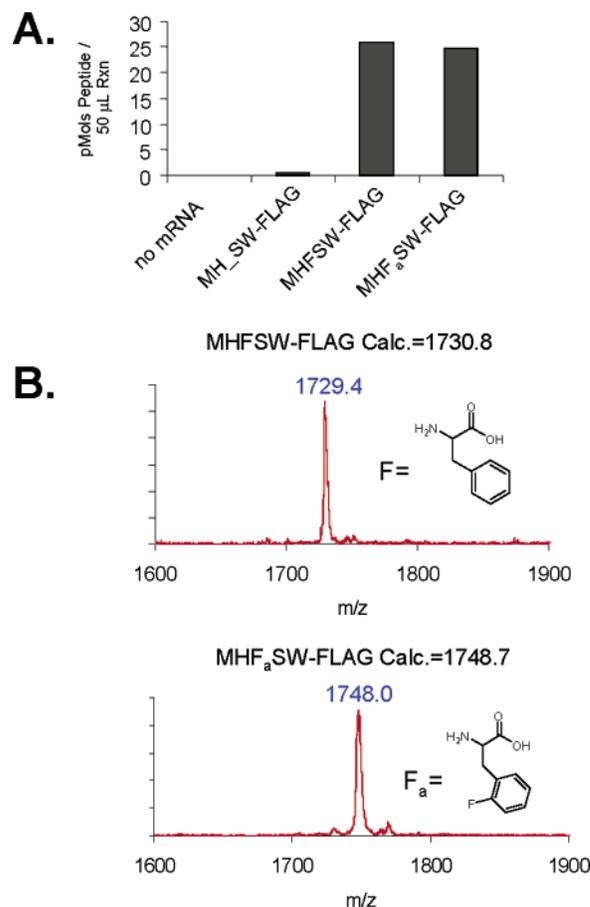


**Figure 2.** TLC AARS assay. The histogram plot indicates the percent of total tRNA converted to aminoacyl-tRNA after 5 min as measured by TLC analysis of the AARS reaction for natural amino acids (red bars), analogues (green bars), and a negative control lacking amino acid (blue bars). Below, three TLC lanes are shown for each AARS reaction containing the natural amino acid (left), the analogue (middle), or the negative control lacking amino acid (right). Percent conversion, the ratio of aminoacyl-tRNA (AA-AMP) to total tRNA (AA-AMP+AMP), is calculated from the integration of background corrected TLC spots. It should be noted that the observed percent tRNA converted reflects the use of total *E. coli* tRNA rather than a single in vitro transcribed tRNA, which can be almost completely aminoacylated.<sup>32</sup>

3.2  $\mu\text{M}$  EF-Tu, 0.6  $\mu\text{M}$  EF-Ts, 0.5  $\mu\text{M}$  EF-G, 0.3  $\mu\text{M}$  RF1, 0.4  $\mu\text{M}$  RF3, 0.1  $\mu\text{M}$  RRF, 0.5  $\mu\text{M}$  ribosomes, and 0.05  $A_{260}$  units/ $\mu\text{L}$  total tRNA. In addition, the reactions contained only the amino acids (0.4 mM unless specified otherwise in figure legends) and AARSs (0.1  $\mu\text{M}$  MetRS, 0.3  $\mu\text{M}$  LeuRS, 0.6  $\mu\text{M}$  GluRS, 0.2  $\mu\text{M}$  ProRS, 1.0  $\mu\text{M}$  GlnRS, 1.0  $\mu\text{M}$  HisRS, 0.3  $\mu\text{M}$  PheRS, 1.5  $\mu\text{M}$  TrpRS, 0.2  $\mu\text{M}$  SerRS, 0.2  $\mu\text{M}$  IleRS, 0.4  $\mu\text{M}$  ThrRS, 0.6  $\mu\text{M}$  AsnRS, 0.6  $\mu\text{M}$  AspRS, 0.5  $\mu\text{M}$  TyrRS, 0.5  $\mu\text{M}$  LysRS, 0.4  $\mu\text{M}$  ArgRS) necessary to translate each peptide. The reactions were assembled on ice and started by addition of mRNA to 1.0  $\mu\text{M}$  followed by incubation for 1 h at 37  $^{\circ}\text{C}$ . Peptides were isotopically labeled by including either 0.2  $\mu\text{M}$   $^{35}\text{S}$ -Met plus 20  $\mu\text{M}$  Met, or 0.5  $\mu\text{M}$   $^3\text{H}$ -His.

**Peptide Translation Assay.** Translated peptides were purified from 50  $\mu\text{L}$  reactions using prewashed FLAG M2 agarose (Sigma) or Ni-NTA agarose (Qiagen). Following two buffer washes (50 mM Tris-HCl, 300 mM NaCl, pH 8.0), the peptides were eluted from both matrices with 0.2% TFA, and the peptide yield was determined by liquid scintillation counting and the specific activity of  $^{35}\text{S}$ -Met or  $^3\text{H}$ -His. For mass analysis, peptides were concentrated and desalted by reverse phase micro-chromatography using C18 zip tips (Millipore) and eluted with a 50% acetonitrile, 0.1% TFA solution saturated with the matrix  $\alpha$ -cyano-4-hydroxycinnamic acid. Mass measurements were made using an Applied Biosystems Voyager MALDI-TOF with delayed extraction operated in the positive mode and externally calibrated with angiotensin II (1046.54 da), ACTH fragment 18–19 (2465.20 da), and the insulin B chain (3494.65 da) standards (Sigma).

**mRNA-Display.** Peptide-mRNA fusions were produced by translating mRNA photochemically ligated to the puromycin linker in 0.5 mL translation reactions configured with either natural or unnatural amino acids. Following the translation reaction, the KCl and  $\text{Mg}(\text{OAc})_2$  concentrations were adjusted to 550 and 50 mM to facilitate fusion formation, and incubated an additional 15 min at room temperature.<sup>36</sup> The mRNA-puromycin-peptide fusions were detected by separation on 10–20% SDS-PAGE gradient gels (Biorad) and autoradiography and were purified by oligo(dT) affinity chromatography. For MALDI-TOF



**Figure 3.** Unnatural amino acid peptide incorporation assay. (A) The peptide yields from four translation reactions were used to determine that  $F_a$  is efficiently translated into peptides. The control translation reactions lacking mRNA or Phe (F) or its analogue ( $F_a$ ) were compared to the reactions containing F or  $F_a$ . (B) MALDI-TOF MS analysis of the peptide products from translation reactions containing F or  $F_a$  confirmed the incorporation of the analogue. Calc. is the calculated average mass for the singly charged species,  $[M + H]^+$ .

analysis of the mRNA-fused peptide, the puromycin-peptide moiety was liberated from the mRNA by digestion with nuclease P1 and purified for mass analysis by nickel affinity and C18 reverse phase micro-chromatography, as described above.

**Posttranslational Chemical Derivatization of Peptides.** Reactions (10  $\mu\text{L}$ ) contained 100 nM peptide, 0.1 M Tris pH 8.0, 0.5 mM  $\text{CuSO}_4$ , 1 mM tris(triazolyl)amine,<sup>37</sup> 1 mM TCEP, and 1 mM either AZT (Sigma) or  $\beta$ -D-glucopyranosyl azide (Sigma) in an aqueous solution with 10% DMSO. After 1 h at ambient temperature, 0.5% TFA (50  $\mu\text{L}$ ) was added to quench, and the peptide products were purified using Zip-Tips and analyzed by MALDI-TOF MS as described above.

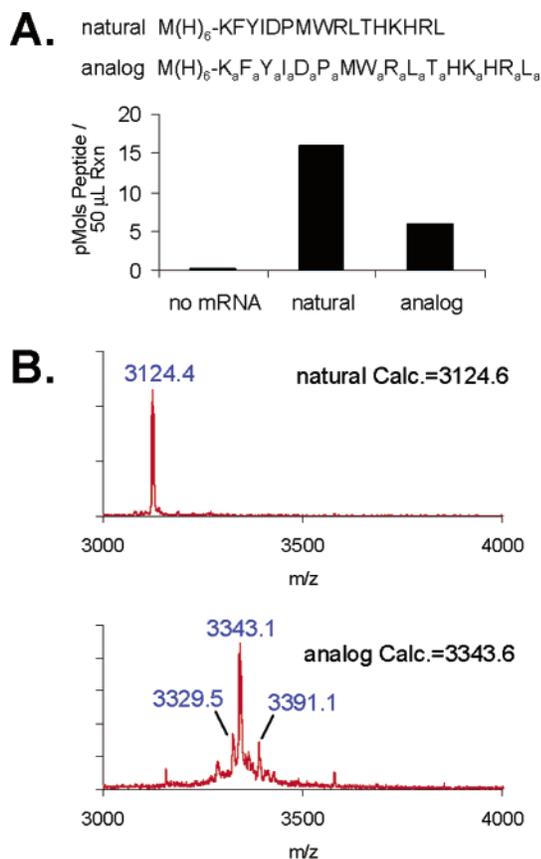
## Results

**Selection of Unnatural Amino Acids.** A literature survey identified unnatural amino acid substrates (Figure 1) for 12 of the 20 AARSs that in most cases were known to be incorporated into proteins by the *E. coli* translation machinery.<sup>22,38</sup> Therefore, by exploiting the natural substrate promiscuity of the AARSs, this set of side-chain analogues could be used to test the ability of a reconstituted translation system to simultaneously incor-

(36) Liu, R.; Barrick, J. E.; Szostak, J. W.; Roberts, R. W. *Methods Enzymol.* **2000**, *318*, 268–293.

(37) Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G. *J. Am. Chem. Soc.* **2003**, *125*, 3192–3193.

(38) Lea, P. J.; Fowden, L. *Phytochemistry* **1973**, *12*, 1903–1916.



**Figure 4.** (A) The amino acid sequences and yields of the peptides resulting from translation of the 23 amino acid mRNA with either 12 natural amino acids (0.15 mM) or 10 unnatural amino acids ( $R_a$ ,  $P_a$ ,  $F_a$ , and  $K_a$ , 0.15 mM;  $D_a$ ,  $W_a$ , and  $T_a$ , 0.2 mM;  $Y_a$ , 0.4 mM;  $I_a$ , 0.8 mM;  $L_a$ , 2.0 mM) plus His (0.15 mM) and Met (4  $\mu$ M). (B) MALDI-TOF analysis of the natural peptide (top) and the analogue peptide (bottom). The basis for the minor peaks observed in the analogue translation is unknown. However, based on their molecular weights, they may result from incomplete substitution of Thr with  $T_a$  (Calc. = 3329.6) and the use of contaminating Trp by HisRS (Calc. = 3392.6).

porate multiple unnatural amino acids. Translation of this set with the requisite AARSs and total tRNA effectively reassigns 35 of the 61 sense codons (Figure 1). Like the 20 proteinogenic amino acids, this set contains hydrophobic, polar, and positively and negatively charged side-chains. However, the properties of several of these unnatural amino acids are significantly different from the natural amino acids that they replace. For instance, the  $pK_a$  of the oxyguanidino group ( $pK_a \approx 7.0$ ) of  $R_a$  is about five units lower than the guanidino group of Arg ( $pK_a \approx 12.0$ ), indicating that the side-chain of  $R_a$  will ionize within the physiological pH range.<sup>39</sup> Conversely, unlike the imidazole side-chain of His ( $pK_a \approx 6.0$ ), which ionizes within the physiological pH range, the 1,2,4-triazole ( $pK_a \approx 3.3$ ) of  $H_a$  is predicted to be completely deprotonated at physiological pH.<sup>40,41</sup> The Trp analogue,  $W_a$ , is more soluble and has a red-shifted fluorescence maximum relative to Trp, a property that has been exploited for FRET studies of protein folding and protease cleavage studies.<sup>42,43</sup>  $D_a$  is one of the invariant Fe(III) coordinating

residues in the aquachelins, a class of marine siderophores.<sup>44</sup>  $P_a$ , the four-membered-ring proline analogue, constrains backbone geometry, but increases local flexibility by increasing rotational clearance (relative to Pro) with the side-chains of neighboring residues in the polypeptide chain.<sup>45</sup>  $F_a$ ,  $Y_a$ , and  $L_a$  are fluorinated amino acid analogues that are more hydrophobic than the natural amino acids from which they are derived and have a propensity for self-association.<sup>46,47</sup> In fact, the replacement of multiple Leu residues with  $L_a$  in the core of the GCN4 leucine zipper increases thermal and chemical stability.<sup>48</sup> Finally, the alkyne side-chain of  $M_a$  provides a functional group with a unique reactivity that is orthogonal to those in biological polymers.<sup>37</sup>

**Testing Unnatural Amino Acids with AARSs.** The 12 unnatural amino acids selected from the literature were verified to be substrates for *E. coli* AARSs using a highly sensitive aminoacylation assay developed by Wolfson and Uhlenbeck.<sup>32</sup> This assay does not depend on isotopically labeled amino acids for detection, but instead uses tRNA terminal nucleotidyl transferase to specifically exchange the 3' terminal adenosine nucleotide of tRNA for [ $\alpha^{32}$ P]-AMP. Following aminoacylation, the labeled tRNA is digested with nuclease P1 and the liberated aminoacyl- $[\alpha^{32}$ P]-AMP is detected by TLC and autoradiography. Analysis of the unnatural amino acids selected from the literature revealed that 11 of the 12 analogues were aminocylated onto total *E. coli* tRNA by purified *E. coli* AARSs (Figure 2). The formation of  $D_a$ - $[\alpha^{32}$ P]-AMP could not be detected by this TLC system, presumably because its side-chain causes it to have the same  $R_f$  as AMP or ATP. Indeed D-AMP is more difficult to separate from AMP than other AA-AMPs (Figure 2). However, the ability of  $D_a$  to serve as a substrate for AspRS was verified using a recently developed MALDI-TOF AARS assay (M. Hartman, unpublished results).

**Testing Translation of Unnatural Amino Acids.** A peptide translation assay was used to test whether the unnatural aminoacyl-tRNAs are (1) properly delivered to the A-site of the ribosome by EF-Tu, (2) efficient peptidyl acceptors in the A-site, and (3) efficient peptidyl-donors in the P-site. The assay was designed to direct the incorporation of each unnatural amino acid by a single test codon placed between an initiator AUG and an affinity tag in one of five mRNA templates (Figure 3 and Table 1). In addition to assessing peptide yields as a measure of translation efficiency, incorporation of the unnatural amino acids was verified by MALDI-TOF mass spectrometry. Both components of this analysis, product yield and mass, are important for assessing the overall fidelity and processivity of the system and confirming the incorporation of the unnatural amino acids in the translated peptide products.

Peptides were purified from the translation reactions on the basis of the C-terminal FLAG or His<sub>6</sub> tag, and the yield was determined by scintillation counting of <sup>35</sup>S-Met (<sup>3</sup>H-His was used for analysis of  $M_a$  containing peptides). The incorporation

(39) Boyar, A.; Marsh, R. E. *J. Am. Chem. Soc.* **1982**, *104*, 1995–1998.  
 (40) Burke, L. A.; Elguero, J.; Leroy, G.; Sana, M. *J. Am. Chem. Soc.* **1976**, *98*, 1685–1690.  
 (41) Soumillion, P.; Fastrez, J. *Protein Eng.* **1998**, *11*, 213–217.  
 (42) De Filippis, V.; De Boni, S.; De Dea, E.; Dalzoppo, D.; Grandi, C.; Fontana, A. *Protein Sci.* **2004**, *13*, 1489–14502.

(43) Anderson, R. D.; Zhou, J.; Hecht, S. M. *J. Am. Chem. Soc.* **2002**, *124*, 9674–9675.  
 (44) Barbeau, K.; Rue, E. L.; Bruland, K. W.; Butler, A. *Nature* **2001**, *413*, 409–413.  
 (45) Zagari, A.; Nemethy, G.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 951–959.  
 (46) Gough, C. A.; Pearlman, D. A.; Kollman, P. *J. Chem. Phys.* **1993**, *99*, 9103–9110.  
 (47) Bilgicer, B.; Xing, X.; Kumar, K. *J. Am. Chem. Soc.* **2001**, *123*, 11815–11816.  
 (48) Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, I. W.; DeGrado, W. F.; Tirrell, D. A. *Biochemistry* **2001**, *40*, 2790–2796.

**Table 1.** MALDI-TOF MS Analysis of Unnatural Amino Acid Incorporation

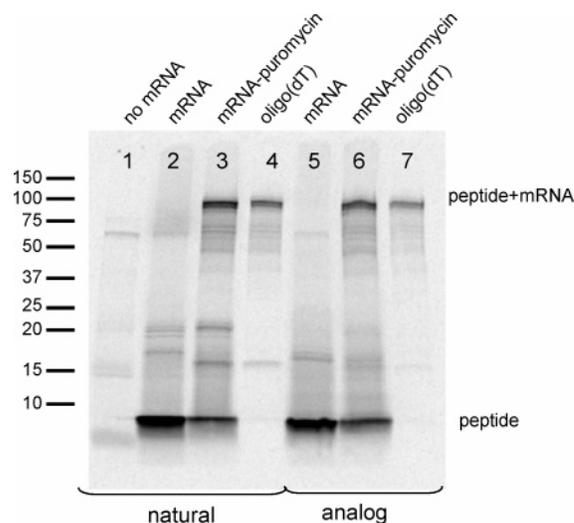
| peptide                                  | % yield <sup>a</sup> | calc.  | m/z    |
|--|----------------------|--------|--------|
| MLEPQDYKDDDDK                            | 100                  | 1640.7 | 1640.6 |
| ML <sub>a</sub> EPQDYKDDDDK              | 100                  | 1694.7 | 1694.9 |
| MLEP <sub>a</sub> QDYKDDDDK              | 88                   | 1626.7 | 1626.8 |
| MHFSWDYKDDDDK                            | 100                  | 1730.8 | 1729.4 |
| MH <sub>F</sub> <sub>a</sub> SWDYKDDDDK  | 95                   | 1748.7 | 1748.0 |
| MHFSW <sub>a</sub> DYKDDDDK              | 100                  | 1731.7 | 1731.5 |
| MH <sub>F</sub> <sub>a</sub> FSWDYKDDDDK | 92                   | 1731.7 | 1731.0 |
| MITNPDYKDDDDK                            | 100                  | 1598.7 | 1597.9 |
| MI <sub>a</sub> TNPDYKDDDDK              | 95                   | 1616.7 | 1615.8 |
| MIT <sub>a</sub> NPDYKDDDDK              | 100                  | 1612.7 | 1611.4 |
| MDYKMHHHHHHH                             | 100                  | 1538.7 | 1538.1 |
| M <sub>a</sub> DYKM <sub>a</sub> HHHHHHH | 100                  | 1494.7 | 1493.8 |
| MD <sub>a</sub> YKMHHHHHHH               | 93                   | 1554.7 | 1555.0 |
| MDY <sub>a</sub> KMHHHHHHH               | 84                   | 1556.7 | 1556.5 |
| MDYK <sub>a</sub> MHHHHHHH               | 100                  | 1556.7 | 1556.3 |
| MTINRDYKDDDDK                            | 100                  | 1557.7 | 1557.9 |
| MTINR <sub>a</sub> DYKDDDDK              | 100                  | 1559.7 | 1560.0 |

<sup>a</sup> The percent yield of the analogue containing peptides relative to the respective natural peptide.

efficiency of each analogue was evaluated by comparing the yields of four translation reactions, a control reaction lacking mRNA and three mRNA programmed reactions with different combinations of amino acids. For example, in addition to the no mRNA control, the three reactions used to analyze F<sub>a</sub> incorporation were a positive control containing all natural amino acids, a negative control lacking Phe, and a test reaction containing F<sub>a</sub> (Figure 3). The peptide yield from the reaction containing F<sub>a</sub> was compared to those of the positive control, which provides a benchmark for efficient translation, and to the negative control, which establishes the background level of misincorporation at a particular codon.

In general, the positive control translation reactions containing a complete complement of the natural amino acids required to translate a given template yielded 25 pmol of peptide from a 50  $\mu$ L reaction. Comparable yields ( $\geq 84\%$ ) were observed for translation reactions containing a single unnatural amino acid, suggesting that all 12 analogues were efficiently incorporated into peptides (Figure 3 and Table 1). The subsequent mass analysis of the affinity purified peptide products confirmed the fidelity of the translation system and revealed that each of the selected unnatural amino acids was faithfully incorporated (Figure 3B and Table 1). Additionally, mass analysis revealed that the M<sub>a</sub> containing peptide was properly formylated and that M<sub>a</sub> was incorporated opposite the initiator AUG as well as within the open reading frame. Therefore, M<sub>a</sub>-tRNA<sup>Met</sup> is a substrate for methionyl-tRNA transformylase (MTF) and its dedicated initiation factor (IF2).

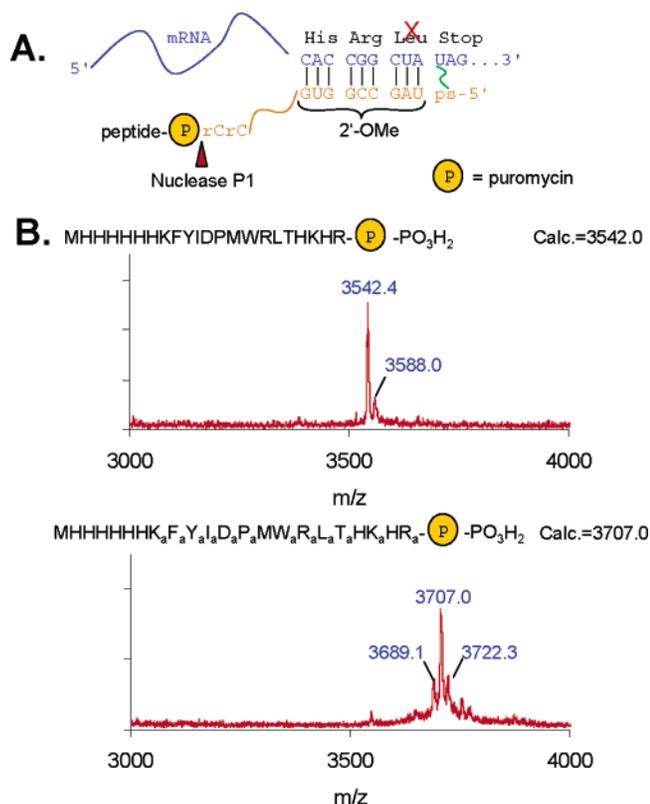
**Unnatural Peptide Translation.** The ability of the reconfigured PURE translation system to simultaneously incorporate multiple unnatural amino acids was tested using an mRNA that coded for a 23-residue peptide with an N-terminal His<sub>6</sub> tag and at least one codon for each of the 12 unnatural amino acids in Figure 1. This template was translated with either all natural amino acids or 10 of the 12 unnatural amino acids in Figure 1 (His and <sup>35</sup>S-Met/Met were included in place of H<sub>a</sub> and M<sub>a</sub> to allow epitope purification and detection of peptide products, respectively). The final concentration of the unnatural amino acid analogues was adjusted to minimize the incorporation of natural amino acids that contaminate some of the components of the translation system or the analogues at low levels (Figure



**Figure 5.** mRNA-display of natural and unnatural peptides. 10–20% gradient SDS-PAGE analysis of translation reactions and oligo(dT) purified peptide-mRNA fusions.

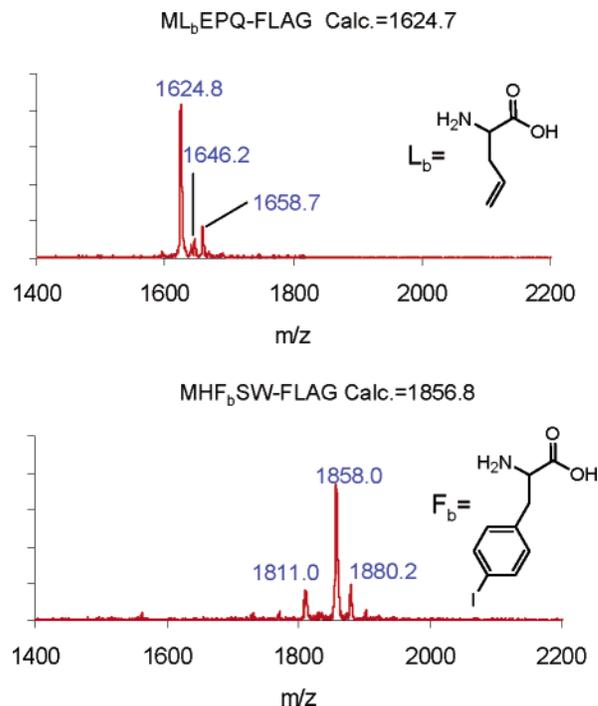
4, legend). Comparison of the peptide yields from 50  $\mu$ L translation reactions containing all natural amino acids (15.9 pmol) or 10 unnatural amino acids plus His and Met (5.9 pmols) indicated that the unnatural peptide is translated with about 37% of the efficiency of the natural peptide (Figure 4). MALDI-TOF analysis of the peptide products from the translation reaction containing natural amino acids revealed a single peak with the predicted molecular weight for the 23-residue peptide. The same analysis of the products from the translation reaction containing 10 amino acid analogues revealed a major peak corresponding to the fully substituted analogue peptide and several minor peaks. Possible explanations for these minor peaks include the presence of contaminating natural amino acids, or incomplete aminoacyl-tRNA synthetase specificity (Figure 4, legend); nevertheless, the fidelity of translation with multiple analogues is clearly sufficient for successful *in vitro* selection experiments.

**mRNA-Display of Natural and Unnatural Peptides in the PURE System.** The *in vitro* selection of ribosomally synthesized peptides requires a physical link between the peptide (phenotype) and the mRNA that codes for its synthesis (genotype). The mRNA-display approach uses the peptidyl-accepting antibiotic puromycin, chemically appended to the 3' of the mRNA, to provide a covalent link between the nascent peptide and its coding mRNA prior to release from the ribosome.<sup>30</sup> Because most mRNA-display experiments have been done with eukaryotic translation extracts, it was important to establish that the reconstituted bacterial PURE translation system and the translation of unnatural peptides are compatible with the formation of puromycin-linked fusions. For this purpose, we modified our 23 amino acid test mRNA with a synthetic oligonucleotide linker ending with a 3' puromycin. Examination of the fusions between this mRNA and peptides translated with natural or unnatural amino acids revealed that in both cases about 35% of the translated peptide was coupled to its mRNA (Figure 5, lanes 3 and 6). As expected, no peptide was detected in control reactions lacking mRNA (Figure 5, lane 1), and no mRNA-peptide fusion was detected in translations programmed with mRNA lacking the 3' puromycin (Figure 5, lanes 2 and 5). The mRNA-peptide fusions can be purified from the translation reaction on the basis



**Figure 6.** MALDI-TOF analysis of the mRNA-displayed peptide. (A) A schematic diagram detailing the interaction between the puromycin linker and the mRNA is shown. The last three codons of the mRNA open reading frame (His, Arg, Leu) form a stable duplex with a complementary 2'-OMe sequence in the puromycin linker. This 2'-OMe stretch is preceded by a psoralen (ps) that intercalates in the mRNA and is predicted to covalently cross-link (shown in green) to the uracil (U) in the stop codon upon photoactivation. The mass analysis of the mRNA-displayed peptide below showed that the final Leu codon is not decoded by the ribosome presumably due to the proximity of the covalent psoralen cross-link. (B) The peptide-puromycin moiety from oligo(dT) purified mRNA fusions with natural (top) or unnatural (bottom) peptides was liberated by nuclease P1 and analyzed by MALDI-TOF. The spectra indicate that both natural and unnatural peptides lack the C-terminal Leu or  $L_a$  and are linked to puromycin, indicating that they are properly displayed on mRNA in the PURE system. The basis for the minor peaks observed is unknown. However, based on their molecular weights, the minor peak observed in the all natural amino acid translation reaction may result from the use of Trp by HisRS (Calc. 3591.0). The minor peaks in the analogue translation may result from the incomplete substitution of Phe, Ile, or Tyr with  $F_a$ ,  $I_a$ , or  $Y_a$  (Calc. = 3689.0) and the use of  $D_a$  by ThrRS (Calc. = 3723.0).

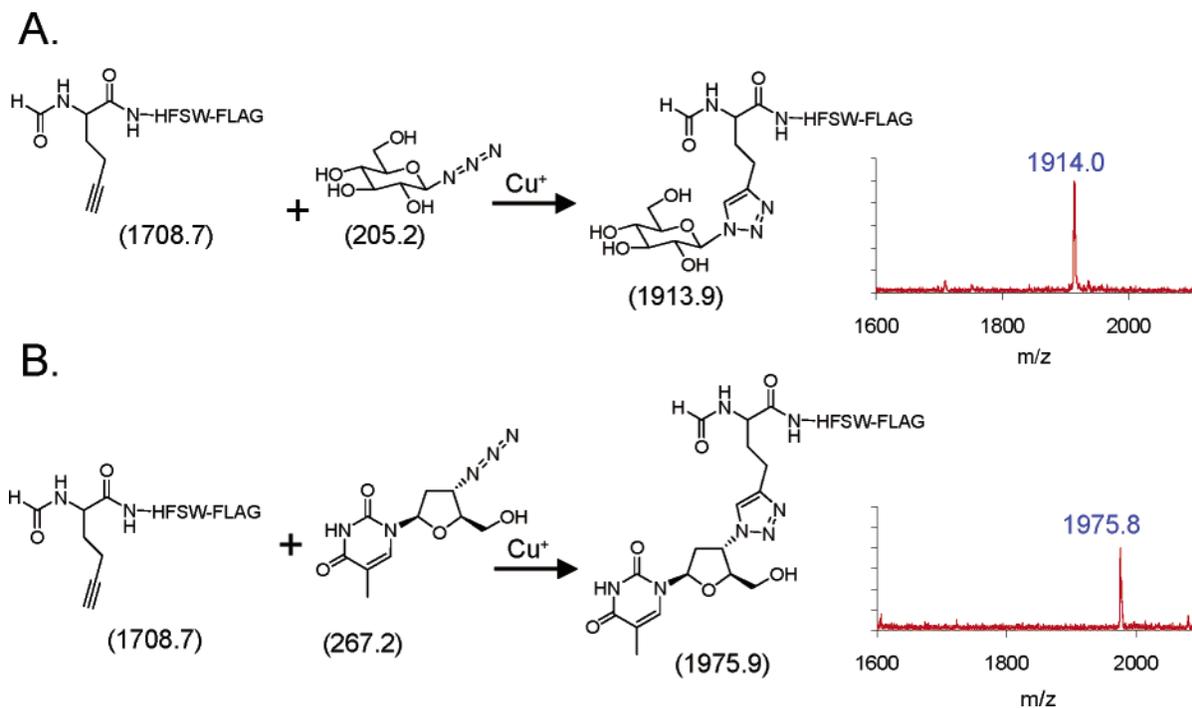
of a poly(dA) stretch in the puromycin linker (Figure 5, lanes 4 and 7). Oligo(dT) purification from 500  $\mu$ L translation reactions containing natural or unnatural amino acids yielded  $\sim$ 28 pmol of mRNA-peptide fusion. The reason the yield of mRNA-peptide fusions is reduced  $\sim$ 10-fold as compared to the free Ni-purified peptides is that the puromycin reaction limits each translating ribosome to a single turnover. The complexity of mRNA-displayed libraries is limited by the amount of mRNA-peptide fusion that can be synthesized in translation reactions of practical volumes. Assuming linear scalability, our observed yield of mRNA-peptide fusion from a 500  $\mu$ L reaction predicts that a 10 mL PURE translation reaction would yield about 560 pmol of fusion, or an unnatural peptide library of approximately  $10^{14}$  unique members. However, because the highest yields in the PURE system were realized with  $\sim$ 5-fold more ribosomes than used in this experiment, it may be possible to increase the library size by a further factor of at least 5.



**Figure 7.** Additional analogues incorporated with mutant AARSs. MALDI-TOF spectra are shown for the incorporation of allylglycine ( $L_b$ , 1.2 mM) and *p*-iodo-Phe ( $F_b$ , 0.8 mM) using mutant LeuRS and PheRS into FLAG-tagged test peptides. In both cases, the desired peptide is the major product. There are two additional products from the reaction reconfigured with LeuRS Asp345Ala. Based on their molecular weights, one of these may be the sodium adduct of the desired product (Calc. = 1646.7), and the second is likely to represent incorporation of Met rather than  $L_b$  (Calc. = 1658.7) due to the promiscuity of the editing-inactivated LeuRS.<sup>51</sup> The additional peaks observed in the spectra of the  $F_b$  containing peptide may be the sodium adduct of the desired product (Calc. = 1878.8) and the incorporation of *p*-bromo-Phe (Calc. = 1809.8), which presumably enters the translation reaction as an impurity in *p*-iodo-Phe.

Although these results were encouraging, it was important to verify the mass of the peptide from the peptide-mRNA fusions to ensure that it was composed of the desired natural or unnatural amino acids and that it was full-length. In particular, if the incorporation of an amino acid analogue is sufficiently slow, then the puromycin fusion reaction might occur prematurely, giving rise to truncated peptide-mRNA fusions that may not be resolved by gel electrophoresis. Therefore, the peptide-puromycin moiety of the oligo(dT) purified peptide-mRNA fusions was liberated by extensive digestion with nuclease P1, purified and concentrated by nickel affinity chromatography, and desalted by C18 micro-chromatography for mass analysis. In both cases, fusions made by translation with natural or unnatural amino acids, the displayed peptides contained the expected amino acids and unnatural analogues. However, the peptides were only 22 amino acids long and lacked the C-terminal Leu or  $L_a$  residue (Figure 6). This presumably results from the inability of the ribosome to decode this codon, which is immediately followed by the 5' psoralen used to covalently photo-cross-link the linker to the mRNA. Surprisingly, the ribosome is able to read through the mRNA/2'-OMe duplex to translate through the two previous codons (His and Arg).

**Expanding Analogue Diversity with Mutant AARSs.** Several engineering or selection-based approaches have been developed to relax or alter the substrate specificity of AARSs for the translation of unnatural amino acids that are excluded



**Figure 8.** Expanding chemical diversity by posttranslational derivatization. Copper(I) catalyzed [3 + 2] cycloaddition between the alkyne of  $M_a$  incorporated at the N-terminus of a peptide and glucopyranosyl azide (A) or AZT (B).

by the natural translation machinery. For example, the mutation of Ala294 to Gly in the Phe binding pocket of *E. coli* PheRS relaxes the specificity of the enzyme, allowing numerous para-substituted Phe analogues to be charged onto tRNA<sup>Phe</sup> and incorporated into proteins.<sup>49</sup> In another example, a mutation that inactivates the editing domain of LeuRS<sup>50</sup> greatly expands the range of substrates that can be incorporated into proteins by tRNA<sup>Leu</sup>.<sup>51,52</sup>

To demonstrate that the chemical diversity of the unnatural amino acids translated in the PURE translation system can be expanded using mutant AARSs, we reconfigured two translation reactions with PheRS Ala294Gly or LeuRS Asp345Ala. As expected, the system was now capable of faithfully incorporating *p*-iodo-Phe ( $F_b$ ) and allylglycine ( $L_b$ ) into peptide products (Figure 7). Not only does this demonstrate that mutant Phe and Leu AARSs can be used for the incorporation of an aryl-iodide ( $F_b$ ) and a terminal alkene ( $L_b$ ), but it also validates the use of mutant AARSs to expand the chemical diversity of our peptide libraries. In fact, Schultz and co-workers have selected mutant AARSs that enable the incorporation of over 30 chemically useful unnatural amino acid analogues, including photocleavable amino acids, fluorescent and glycosylated side-chains, chemical groups with unique and photocaged reactivity, and groups that can be photo-cross-linked both in vitro and in vivo.<sup>53</sup>

**Expanding Analogue Diversity by Posttranslational Chemical Derivatization.** An additional approach to expanding the chemical diversity of products of the PURE translation system is to incorporate unnatural amino acids with unique reactivity

for posttranslational chemical modification. This strategy was recently combined with mRNA-display to select for peptide-drug conjugates between 10-residue peptides of natural amino acids and penicillin, that displayed 100-fold higher activity than free penicillin.<sup>54</sup> This approach can also be adapted to libraries built from unnatural amino acids by exploiting the orthogonal reactivity of the terminal alkyne in the side-chain of  $M_a$ , which can be specifically ligated with azide-containing small molecules by copper(I)-catalyzed [3 + 2] cycloaddition.<sup>37,55</sup> To demonstrate this principle the azide derivatives of the nucleoside thymidine and the sugar glucose, AZT and glucopyranosyl azide, respectively, were posttranslationally appended to the terminal alkyne of the  $M_a$  side-chain at the N-terminus of a peptide. The expected thymidine and glucopyranosyl ligated products were detected by MALDI-TOF MS (Figure 8). Similar, aqueous derivatization reactions catalyzed by Pd(II) and Cu(I) are also available for the allylglycine (above) and *p*-azido-Phe (unpublished results) side-chains incorporated with mutant LeuRS and PheRS.<sup>56</sup>

## Discussion

As a step toward the development of a system for the ribosomal synthesis and in vitro selection of novel NRP-like molecules, we have adapted the PURE reconstituted translation system for the simultaneous translation of 10 unnatural amino acid analogues and demonstrated the potential of this system to synthesize libraries of peptide-mRNA fusions with up to  $10^{14}$  members. Moreover, the chemical diversity of these peptide libraries can be further expanded by the use of mutant AARSs or by the specific posttranslational derivatization of a reactive side-chain group with useful small molecules. We are currently

(49) Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. *ChemBioChem* **2002**, *3*, 235–237.

(50) Mursinna, R. S.; Martinis, S. A. *J. Am. Chem. Soc.* **2002**, *124*, 7286–7287.

(51) Tang, Y.; Tirrell, D. A. *Biochemistry* **2002**, *41*, 10635–10645.

(52) Lincecum, T. L., Jr.; Tukalo, M.; Yaremchuk, A.; Mursinna, R. S.; Williams, A. M.; Sproat, B. S.; Van Den Eynde, W.; Link, A.; Van Calenbergh, S.; Grotli, M.; Martinis, S. A.; Cusack, S. *Mol. Cells* **2003**, *11*, 951–963.

(53) Wang, L.; Schultz, P. G. *Angew. Chem., Int. Ed.* **2004**, *44*, 34–66.

(54) Li, S.; Roberts, R. W. *Chem. Biol.* **2003**, *10*, 233–239.

(55) Lin, H.; Walsh, C. T. *J. Am. Chem. Soc.* **2004**, *126*, 13998–14003.

(56) Kanan, M. W.; Rozenman, M. M.; Sakurai, K.; Snyder, T. M.; Liu, D. R. *Nature* **2004**, *431*, 545–549.

extending both of these approaches by using additional mutant AARSs and by incorporating several unnatural amino acids with orthogonal reactivities for the derivatization of multiple amino acid side-chains with different small molecules. We have also begun extensively screening amino acid analogues with modified side-chains and backbones that have not previously been tested in translation. Last, we are testing different strategies for the posttranslational macrocyclization of our unnatural peptides to make our linear unnatural peptides more similar to NRPs.

Typical small molecule libraries created using combinatorial chemistry and screened by high-throughput robots generally do not exceed  $\sim 10^6$  molecules. As the members of these libraries are intended to be orally bioavailable, they are typically less than 500 Da and are designed to include many different chemical groups, including those not found in biology. Recently, “chemical translation” systems have been described that are capable of synthesizing small molecule and peptide libraries comprised of chemically diverse amino acid and non-amino acid building blocks.<sup>57–60</sup> These methods combine the powerful *in vitro* selection techniques normally restricted to the synthesis and selection of biological polymers with the chemical versatility of small molecule libraries. Although still only in their infancy,

the size of the libraries created by these systems has thus far not exceeded  $10^6$ , but theoretically could approach  $10^{12}$ .

By combining chemically diverse building blocks compatible with translation with strategies for expanding the genetic code, we hope to create highly modified peptide libraries of high complexity. For example, the theoretical complexity of a 10-residue library translated using an expanded genetic code with 30 amino acid analogues is  $\sim 10^{15}$ . As currently configured, our system can synthesize  $10^{14}$  peptide-mRNA fusion molecules. While complete coverage of  $10^{15}$  sequences may not be possible with our system, the chemical space sampled by  $10^{14}$  of these sequences dwarfs traditional small molecule libraries and is comparable to that sampled by chemical translation systems.

**Acknowledgment.** We thank Takuya Ueda, Paul Schimmel, and Chuck Merryman for expression plasmids, David Tirell and Jamie Link for amino acid analogues, and Chuck Merryman for helpful discussions. J.W.S. is an Investigator and K.J. and M.C.T.H. are Research Associates of the Howard Hughes Medical Institute. This work was supported by the HHMI.

**Supporting Information Available:** PCR primers (Table S1) for construction of translation factor and AARS expression plasmids (Table S2); mRNA sequences (Figure S1). This material is available free of charge via the Internet at <http://pubs.acs.org>.

(57) Halpin, D. R.; Lee, J. A.; Wrenn, S. J.; Harbury, P. B. *PLoS Biol.* **2004**, *2*, 1031–1038.

(58) Halpin, D. R.; Harbury, P. B. *PLoS Biol.* **2004**, *2*, 1022–1030.

(59) Halpin, D. R.; Harbury, P. B. *PLoS Biol.* **2004**, *2*, 1015–1021.

(60) Gartner, Z. J.; Tse, B. N.; Grubina, R.; Doyon, J. B.; Snyder, T. M.; Liu, D. R. *Science* **2004**, *305*, 1601–1605.

JA0515809