Expanding the Genetic Repertoire of the Methylotrophic Yeast *Pichia pastoris*[†]

Travis S. Young,[‡] Insha Ahmad,[‡] Ansgar Brock,[§] and Peter G. Schultz*,[‡]

Department of Chemistry and Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, and Genomics Institute of the Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, California 92121

Received November 25, 2008; Revised Manuscript Received January 13, 2009

ABSTRACT: To increase the utility of protein mutagenesis with unnatural amino acids, a recombinant expression system in the methylotrophic yeast *Pichia pastoris* was developed. Aminoacyl-tRNA synthetase/suppressor tRNA (aaRS/$tRNA_{CUA}$) pairs previously evolved in *Saccharomyces cerevisiae* to be specific for unnatural amino acids were inserted between eukaryotic transcriptional control elements and stably incorporated into the *P. pastoris* genome. Both the *Escherichia coli* tyrosyl- and leucyl-RS/$tRNA_{CUA}$ pairs were shown to be orthogonal in *P. pastoris* and used to incorporate eight unnatural amino acids in response to an amber codon with high yields and fidelities. In one example, we show that a recombinant human serum albumin mutant containing a keto amino acid (*p*-acetylphenylalanine) can be efficiently expressed in this system and selectively conjugated via oxime ligation to a therapeutic peptide mimetic containing an ϵ-(2-(aminooxy)acetyl)-L-lysine residue. Moreover, unnatural amino acid expression in the methylotrophic host was systematically optimized by modulation of aaRS levels to express mutant human serum albumin in excess of 150 mg/L in shake flasks, more than an order of magnitude better than that reported in *S. cerevisiae*. This methodology should allow the production of high yields of complex proteins containing unnatural amino acids whose expression is not practical in existing systems.

Recently, we developed methodology that makes it possible to genetically encode a wide variety of unnatural amino acids with novel properties (including fluorophores, metal ion chelators, photocaged and photo-cross-linking groups, NMR, crystallographic and IR probes, and posttranslationally modified amino acids) in both prokaryotic and eukaryotic organisms (*1*−*3*). This is accomplished through the evolution of an orthogonal aminoacyl-tRNA[1] synthetase/suppressor tRNA (aaRS/$tRNA_{CUA}$) pair, designed to selectively insert a desired unnatural amino acid in response to a nonsense or frameshift codon. Thus far, this methodology has been used to add more than 40 unnatural amino acids to the genetic repertoires of *Escherichia coli*, *Saccharomyces cerevisiae*, and several mammalian cell lines (*2*−*4*). Orthogonality in these systems is achieved by transplanting an orthogonal aaRS/$tRNA_{CUA}$ pair with distinct tRNA identity elements into the host organism such that no cross-aminoacylation occurs between the host aminoacylation machinery and the transplanted aaRS/tRNA pair (while still maintaining function in translation). In the current systems, this has proven most successful using aaRS/$tRNA_{CUA}$ pairs derived from the *Methanococcus jannaschii* tyrosyl-RS/$tRNA_{CUA}$ pair in *E. coli* (*5*) and the *E. coli* tyrosyl- or leucyl-RS/$tRNA_{CUA}$ pairs in *S. cerevisiae* (*2, 6*) or mammalian cells (*4*). Directed evolution is then used to alter the specificity of the orthogonal aaRS so that it recognizes the unnatural amino acid of interest and not an endogenous amino acid.

To apply this methodology to the production of large quantities of proteins that are not easily expressed in bacterial hosts, a recombinant system is desired with low cost, scalability, and the ability to produce complex, posttranslationally modified proteins. One such host is *Pichia pastoris,* which is capable of producing mammalian proteins in quantities comparable to those of *E. coli* (*7*). Therapeutic proteins such as tumor necrosis factor (TNF), tetanus toxin fragment C (TTC), and human serum albumin (HSA) have afforded expression levels >10 g/L in high-density fermentations (*8*−*11*). *P. pastoris*' ability to produce proteins in such yields is attributed to its alcohol oxidase 1 promoter (P_{AOX1}), one of the most highly regulated and strongest promoters known (*12*). In addition, *P. pastoris* lacks endotoxins which can contaminate therapeutic proteins expressed in *E. coli* and does not produce antigenic α1,3 glycan linkages as does *S. cerevisiae* (*13*). Additionally, it is now possible to modulate glycosylation patterns in *P. pastoris*, including control of sialylation (*14*). For these reasons, we undertook the development of methodology to allow unnatural amino acids to be genetically encoded in *P. pastoris*. Here we report that eight unnatural amino acids were site-specifically introduced

[†] Supported by the U.S. Department of Energy, Division of Materials Sciences (Grant DE-FG03-00ER46051). T.S.Y. is supported by an Achievement Rewards for College Scientists scholarship. This is paper 19876 of The Scripps Research Institute.

* To whom correspondence should be addressed. Telephone: (858) 784-9300. Fax: (858) 784-9440. E-mail: Schultz@scripps.edu.

[‡] The Scripps Research Institute.

[§] Genomics Institute of the Novartis Research Foundation.

[1] Abbreviations: tRNA, transfer ribonucleic acid; $tRNA_{CUA}$, suppressor transfer ribonucleic acid; PCR, polymerase chain reaction; OD_{600}, optical density at 600 nm; miRNA, microribonucleic acid; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; EDTA, ethylenediaminetetraacetic acid; HPLC, high-performance liquid chromatography; FPLC, fast-protein liquid chromatography; TFA, trifluoroacetic acid; RP LC-MS/MS, reverse-phase liquid chromatography tandem mass spectrometry; TCEP, tris(2-carboxyethyl)phosphine; MALDI-TOF, matrix-assisted laser desorption ionization time of flight.

10.1021/bi802178k CCC: $40.75
Published on Web 03/05/2009

into recombinant human serum albumin (rHSA) expressed in this host with high yields and fidelities.

MATERIALS AND METHODS

Materials. DNA primers were purchased from Integrated DNA Technology (San Diego, CA) and sequences listed in Supporting Information, Table T1. Restriction enzymes were purchased from New England Biolabs (Beverly, MA). The pPIC3.5k vector (map available at http://tools.invitrogen.com/content/sfs/manuals/ppic3.5kpao_man.pdf) and protocols for yeast competency, transformation, and media recipes (Multi-Copy *Pichia* Expression Kit Version F (*15*) manual available at http://tools.invitrogen.com/content/sfs/manuals/pichmulti_man.pdf) were purchased from Invitrogen (Carlsbad, CA). DNA was amplified in *E. coli* DH10B (Invitrogen) or, when noted, by PCR using Platinum Pfx (Invitrogen). The rHSA gene (accession BC034023) was obtained from the Mammalian Gene Collection (National Institutes of Health, Bethesda, MD). All DNA constructs were confirmed by DNA sequencing (Genomics Institute of the Novartis Research Foundation, La Jolla, CA). The double auxotrophic strain, GS200 (*his4, arg4*), and pBLARG vector were gracious gifts from the James Cregg laboratory at the Keck Graduate Institute, Claremont, CA. Transformations of *P. pastoris* and *E. coli* were carried out on a GenePulser Xcell (Bio-Rad, Hercules, CA) using 2 and 1 mm electroporation cuvettes (Fisher Scientific, Rochester, NY). Tris−glycine (4−20%) SDS−PAGE gels for protein analysis were purchased from Invitrogen. RNA was harvested via the protocols and reagents accompanying the Purelink miRNA isolation kit (Invitrogen) or Ribo-pure−yeast kit (Ambion, Austin, TX). All relative gel band densities were determined using Photoshop CS2 (Adobe, San Jose, CA).

Construction of pPIC3.5k-rHSA. For compatibility with pPIC3.5k linearization (*15*), *Bgl*II sites were removed from rHSA by a modified QuikChange mutagenesis (Stratagene, La Jolla, CA) protocol (*16*) using primers −*Bgl*II 1F and −*Bgl*II 1R for $BglII_{781}$, and −*Bgl*II 2F and −*Bgl*II 2R for $BglII_{817}$ to create $rHSA_{WT}$. Glu37 was replaced by the amber codon TAG using the modified QuikChange protocol and the primers Glu37 F′ and Glu37 R′ to create $rHSA_{E37X}$. $rHSA_{WT}$ and $rHSA_{E37X}$ were amplified using primers HSA forward and HSA reverse, digested with *Eco*RI and *Bam*HI, and ligated into the similarly digested pPIC3.5k vector to create pPIC3.5k-$rHSA_{WT}$ or pPIC3.5k-$rHSA_{E37X}$.

Construction of pREAV. The pPR1-P_{PGK1} + 3*SUP4*-$tRNA^{Tyr}_{CUA}$ vector (*17*) harboring the pApaRS was amplified by PCR, excluding the *TRP* and 2μ origin regions, to add restriction sites *Kpn*I and *Hin*dIII with primers pESC F and pESC R and digested with *Hin*dIII and *Kpn*I. The *ARG4* coding region was amplified from pBLARG with primers ARG4 F new and ARG4 R new, digested with *Kpn*I and *Hin*dIII, and ligated into the similarly digested pPR1-P_{PGK1} + 3*SUP4*-$tRNA^{Tyr}_{CUA}$ PCR product to create the recombinant eukaryotic *ARG4* vector, pREAV-P_{ADH1}-pApaRS. To create pREAV-P_{AOX1}-pApaRS, the *AOX1* promoter and terminator sequences were derived from pPIC3.5k. The pApaRS was amplified with primers KETO-Koz-F and KetoRS R 6×His, digested with *Eco*RI and *Not*I, and ligated into the similarly digested pPIC3.5k to create pPIC3.5k-pApaRS. The pREAV-P_{ADH1}-pApaRS was amplified by PCR, excluding the P_{ADH1}-pApaRS-T_{ADH1} region, to add restriction sites *Asc*I and *Afl*II with primers pESC-AOX-KETO F and pESC-AOX-KETO R. The P_{AOX1}-pApaRS-T_{AOX1} coding region was amplified from pPIC3.5k-pApaRS with primers pPIC-keto AOX5 F and pPIC-keto AOXTT R, digested with *Asc*I and *Afl*II, and ligated into the similarly digested pREAV-P_{ADH1}-pApaRS PCR product to create pREAV-P_{AOX1}-pApaRS. Constructs were confirmed by size mapping and sequencing.

Transformation of Cassettes into P. pastoris. Briefly, to create GS200-$rHSA_{E37X}$ (*HIS4*, *arg4*, Gen^R, mut^s), 20 μg of pPIC3.5k-$rHSA_{E37X}$ was linearized with *Bgl*II, concentrated to 10 μL by ethanol preciptiation, added to 80 μL of freshly competent GS200 in a 2 mm electroporation cuvette, and electroporated with the *P. pastoris* settings (2000 V, 25 μF, 200 Ω). Cells were recovered in 1 mL of cold 1 M sorbitol; 250 μL of recovered cells was plated on regeneration dextrose Bacto agar (RDB) plates (15 cm) supplemented with 0.4 mg/mL L-arginine (Arg) and incubated at 30 °C. After 3 days, colonies were picked into 96-well 2 mL blocks (Nunc, Rochester, NY) with 1 mL of yeast peptone dextrose (YPD) media and grown overnight (29.2 °C, 300 rpm). The cultures were diluted 1:100 and 1−2 μL of replica plated on YPD agar plates containing 0.25 mL of Geneticin (Invitrogen) and incubated at 30 °C. After 4 days, colony G3 showed good growth, was picked, and made competent. Transformations to create GS200-$rHSA_{E37X}$/pREAV-P_{ADH1}-pApaRS and GS200-$rHSA_{E37X}$/pREAV-P_{AOX1}-pApaRS (*HIS4*, *ARG4*, Gen^R, mut^s) were carried out using the aforementioned protocol with competent G3, except recovered cells were plated on RDB plates lacking L-histidine (His) and Arg. After 3 days colonies were picked into 96-well 2 mL blocks and rescreened as above for resistance to 0.25 mg/mL Geneticin. GS200-$rHSA_{WT}$/pREAV-P_{ADH1}-pApaRS (*HIS4*, *ARG4*, Gen^R, mut^s) was created in identical fashion to isolate colony F2. GS200-pREAV-P_{AOX1}-pApaRS (*his4*, *ARG4*, Gen^R, mut^s) was created by transforming pREAV-P_{AOX1}-pApaRS into GS200 but plated on RDB plates supplemented with 0.4 mg/mL His and not further screened for Geneticin resistance.

Test Protein Expression. All protein expression experiments followed protocols for mut^s found in the Multi-Copy *Pichia* Expression Kit (*15*). Briefly, 14 colonies for GS200-$rHSA_{E37X}$/pREAV-P_{ADH1}-pApaRS, GS200-$rHSA_{E37X}$/pREAV-P_{AOX1}-pApaRS, or GS200-$rHSA_{WT}$/pREAV-P_{ADH1}-pApaRS were picked from plates containing 0.25 mg/mL Geneticin and grown to near saturation ($OD_{600} \approx 12-18$) in 10 mL of buffered glycerol-complex medium (BMGY) (29.2 °C, 300 rpm). Cultures were centrifuged at 1500*g* (10 min) and resuspended in 2 mL of buffered methanol-complex media (BMMY) with 2 mM pApa amino acid (SynChem, Des Plaines, IL). Growth was continued for 6 days, with methanol supplementation to 0.5% every 24 h. Media (200 μL, 10% culture volume) or sterile water was added every 24 h to account for evaporation. Media (50 μL) was removed every 24 h and cleared of cells by centrifugation at 3000*g* (5 min). Twenty-five microliters of the cleared media was added to 12.5 μL of SDS loading buffer, heated for 1 min at 95 °C, and run on a SDS−PAGE gel (150 V, 1 h). Bands at 66.5 kDa in GS200-$rHSA_{E37X}$/pREAV-P_{AOX1}-pApaRS and GS200-$rHSA_{WT}$/pREAV-P_{ADH1}-pApaRS expression were clearly visible by Coomassie stain (40% methanol, 10% acetic acid, 50% water, 0.1% (w/v) Coomassie Brilliant Blue R250 (Sigma-Aldrich, St. Louis, MO)) after 3 days and peaked

after 6 days. Clone G3-2 for GS200-rHSA$_{E37X}$/pREAV-P$_{AOX1}$-pApaRS and F2-wt for GS200-rHSA$_{WT}$/pREAV-P$_{ADH1}$-pApaRS showed highest expression and were used in further comparisons. No clones from GS200-rHSA$_{E37X}$/pREAV-P$_{ADH1}$-pApaRS showed expression by Coomassie stain. To confirm that amber suppression was specific for pApa, clones from GS200, G3, GS200-pREAV-P$_{AOX1}$-pApaRS, G3-2, and F2-wt were expressed as described above in the presence or absence of 2 mM pApa and 0.5% methanol (Figure 2B).

tRNA Northern Blot. Two *P. pastoris* clones, G3-2 and GS200, and two *S. cerevisiae* clones, SCY4-pPR1-P$_{PGK1}$ + 2SUP4-tRNA, and SCY4, were grown under their respective expression conditions, and micro-RNA (miRNA) harvested. Two micrograms of RNA from each sample was loaded onto two 6% Novex TBE-urea gel (Invitrogen) and run at 180 V for 1 h. RNA was transferred to a Biodyne B nylon membrane (Pall Life Science, East Hills, NY) using an XCell surelock mini-cell (Invitrogen) in 0.5× TBE buffer (Invitrogen) and accompanying protocols. The membranes were auto-cross-linked with UV Stratalinker 2400 (Stratagene, La Jolla, CA). Hybridization and detection were carried out with protocols and reagents found in the North2South chemiluminescent hybridization and detection kit (Pierce, Rockford, IL). Briefly, one blot was incubated with biotinylated probes specific for tRNA^{Ser} (tRNaser cere 1, 5′-/5Biosg/CAT TTC AAG ACT GTC GCC TTA ACC ACT CGG CCA T-3′; tRNaser cere 2, 5′-/5Biosg/GAA CCA GCG CGG GCA GAG CCC AAC ACA TTT CAA G-3′; tRNaser pich 1, 5′-/5Biosg/CTG CAT CCT TCG CCT TAA CCA CTC GGC CAT CGT A-3′; tRNaser pich 2, 5′-/5Biosg/ACA CGA GCA GGG TTC GAA CCT GCG CGG GCA GAG C-3′) and the second blot incubated with biotinylated probes specific for tRNA$^{Tyr}_{CUA}$ (tRNA 5′ biot, 5′-/5Biosg/GGA AGG ATT CGA ACC TTC GAA GTC GAT GAC GG-3′; tRNA 3′ biot, 5′-/5Biosg/TCT GCT CCC TTT GGC CGC TCG GGA ACC CCA CC-3′). Probes were incubated overnight at 55 °C, bound to a streptavidin−horseradish peroxidase (HRP) conjugate, and detected with a luminol/enhancer stable peroxide solution (Pierce) (Figure 2A). Relative tRNA amounts were determined by band density.

Scaled Expression, Purification, and Mass Spectrometry of rHSA$_{E37X}$. For scaled expression of rHSA$_{E37pApa}$, the test expression protocol was modified. One liter of BMGY was inoculated with 20 mL of saturated G3-2 culture in YPD and grown (∼24 h, 29.2 °C, 300 rpm) to OD$_{600}$ ≈ 12−18. The culture was centrifuged at 1500*g* and resuspended in 200 mL of buffered minimal methanol (BMM) supplemented with 10% BMMY and 2 mM pApa. After 6 days of growth (29.2 °C, 300 rpm, with methanol and volume supplementation) the culture was centrifuged at 3000*g*, cells were discarded, and media were passed through a 0.22 μm filter (Millipore, Billerica, MA). The media were ammonium sulfate (NH_4SO_4) precipitated by addition of NH_4SO_4 with slow stirring at 4 °C to 50% of saturation (58.2 g), centrifugation at 20000*g* for 20 min, and again by addition of NH_4SO_4 to 75% of saturation (31.8 g), and centrifugation at 20000*g* for 20 min. The second precipitation contained rHSA$_{E37pApa}$ and was resuspended in FPLC buffer A (25 mM Tris-HCl, 25 mM sodium chloride, 1 mM EDTA, 1× protease inhibitor cocktail (Roche, Basel, CH), pH = 8.5). The resolubilized protein was purified with a MonoQ 5/5 column (GE Healthcare, Giles, U.K.) on an AKTA purifier FPLC (Amersham Biosciences, Piscataway, NJ) (elution at 20−35% buffer B (buffer A + 1 M NaCl)). Fractions were analyzed by SDS−PAGE gel, combined, dialyzed with a 30 MWCO dialysis cassette (Pierce) to PBS, and purified with a Superdex 200 10/300 GL (GE Healthcare) on an AKTA purifier FPLC (elution after 14 min in PBS at 0.5 mL/min). Fractions were analyzed by SDS−PAGE gel, combined, and purified with a C8 Vydac HPLC column (300 mm, 200 Å, 5 μm; Grace) on a Dynamax HPLC (Rainin, Oakland, CA) (elution at 40−46% MeCN in water, 0.1% TFA). Fractions were analyzed by SDS−PAGE gel, and rHSA$_{E37pApa}$-containing fractions were flash frozen and lyophilized to a white powder. Purification of rHSA$_{WT}$ from F2-wt was done in similar fashion.

Tryptic Digest, Nano-RP LC-MS/MS. Purified rHSA$_{E37X}$ was digested overnight with trypsin under reducing conditions (10 mM TCEP, 1 M guanidine hydrochloride, 100 mM triethanolamine hydrochloride, pH = 7.8). The digest was purified by reversed-phase solid-phase extraction with a Sep-Pak, C18 (Waters, Milford, MA), and lyophilized. Oxidation of cysteines to cysteic acid and methionine to methionine sulfone was performed by incubation of lyophilized peptides with performic acid (9 parts concentrated formic acid + 1 part 30% H_2O_2) (*18*) for 1 h on ice. The reaction was quenched by addition of an excess of mercaptoethanol and 20× dilution with water. Nano-RP LC-MS/MS was performed with a HPLC system (Agilent Technologies, Santa Clara, CA) equipped with an LTQ Orbitrap hybrid mass spectrometer (ThermoElectron, Rochester, NY). Tryptic digests were loaded onto the precolumn (4 cm, 100 μm i.d., 5 μm, Monitor C18; Column Engineering, Chicago, IL) of a vented column setup (*19*) at a flow rate of ∼2 μL/min. After a load/wash period of 10 min gradient elution was started by switching the precolumn in line with the analytical column (10 cm, 75 μm i.d., 5 μm C18). The chromatographic profile was from 100% solvent A (0.1% aqueous acetic acid) to 50% solvent B (0.1% acetic acid in acetonitrile) in 40 min at ∼100 nL/min. Data-dependent MS/MS acquisitions were performed following a top 10 scheme in which the mass spectrometer was programmed to first record a high-resolution Orbitrap scan (*m*/*z* 500−2000) followed by 10 data-dependent MS/MS scans (relative collision energy = 35%; 3 Da isolation window). The raw data were searched against the SwissProt 51.6 database using MASCOT (Matrixscience, London, U.K.) for protein identification with pApa as a variable modification.

Creation of mut$^+$ Phenotype. To create GS200-rHSA$_{E37X}$/pREAV-P$_{AOX1}$-pApaRS (*HIS4*, *ARG4*, Gen^R, mut$^+$), 20 μg of pPIC3.5k-rHSA$_{E37X}$ was linearized with *SacI* or *Sal*I and transformed into freshly competent GS200 as previously described. Cells were recovered in 1 mL of cold 1 M sorbitol and plated on RDB plates supplemented with 0.4 mg/mL Arg. Colonies were picked into a 2 mL 96-well block with 1 mL of YPD, grown to saturation (29.2 °C, 300 rpm), diluted 1:100, and replica plated on plates containing 0−3.0 mg/mL Geneticin. Clone 1D12, which survived up to 1.0 mg/mL Geneticin, was made competent, transformed with pREAV-P$_{AOX1}$-pApaRS as previously described, and plated on RDB plates lacking Arg or His. Colonies were picked into a 1 mL 96-well block, grown to saturation, diluted 1:100, and rescreened on Geneticin 1.0 mg/mL plates. Fourteen surviving clones were picked and tested for rHSA$_{E37pApa}$

expression in the presence of pApa amino acid and methanol. For continuity, the mut^s protocol was used as described above. Clone K5 showed greatest protein expression and was compared to G3-2 in test expressions (Supporting Information, Figure S3). Relative amounts of protein were determined by band density.

Construction of pREAV-P$_{promoter}$-pApaRS. Five promoters, P$_{AOX2}$, P$_{YPT1}$, P$_{ICL1}$, P$_{FLD1}$, and P$_{GAP}$, were separately amplified by PCR from genomic DNA (*P. pastoris* GS200) using the following primers: PAOX2 F and PAOX2 R, PYPT1 F and PYPT1 R, PICL1 F and PICL1 R, PFLD1 F and PFLD1 R, and PGAP1 F and PGAP1 R, respectively. The PCR-amplified fragments were digested with *Afl*II and *Nco*I and ligated into the similarly digested pREAV-P$_{AOX1}$-pApaRS (after removal of the P$_{AOX2}$ coding region via agarose gel purification) to create the pREAV-P$_{promoter}$-pApaRS. After sequence confirmation, the plasmids (including the previously constructed pREAV-P$_{AOX1}$-pApaRS) were linearized with *Aat*II, transformed into freshly competent GS200-rHSA$_{E37X}$ (clone 1D12), and plated on RDB plates lacking Arg or His as previously described to create GS200-rHSA$_{E37X}$/pREAV-P$_{promoter}$-pApaRS (*HIS4*, *ARG4*, Gen^R, mut$^+$). Surviving clones were screened for Geneticin resistance at 0.75 and 1.0 mg/mL. Forty-eight clones corresponding to each promoter were picked into 1 mL 96-well blocks containing BMGY and grown to saturation (29.2 °C, 24 h, 300 rpm). The saturated cultures were centrifuged at 1500*g* for 10 min, and cells were resuspended in 200 μL of BMMY + 2 mM pApa amino acid. After 6 days (29.2 °C, 300 rpm, with supplementation), the media were cleared by centrifugation at 3000*g* for 10 min, and 1−2 μL of the cleared media was spotted on a 0.45 μm nitrocellulose membrane (Bio-Rad) using a 96-well pin tool. The membrane was probed with the HSA antibody [1A9] HRP conjugate (Abcam, Cambridge, MA) using standard Western blotting techniques (*20*) and detected with ECL HRP chemiluminescence detection reagents and protocols (GE Healthcare). The two highest expressing clones corresponding to each promoter (*AOX2*, A6 and B7; *YPT1*, D11 and B7; *ICL1*, E5 and H3; *FLD1*, E11 and F3; *GAP*, B7 and B10; *AOX1*, E3 and E7) were chosen for parallel test expressions (Figure 4).

pApaRS Northern Blot. Top expressing clones, *AOX2*, B7; *YPT1*, D11; *ICL1*, H3; *FLD1*, E11; *GAP*, B7; and *AOX1*, E3, were grown under test expression conditions for 6 days. Cells (3×10^8) (2.5 mL at OD$_{600}$ = 1.0) were collected, and total RNA was isolated via the RiboPure yeast kit (Ambion) reagents and protocols. Thirteen micrograms of each RNA sample was loaded onto a 2% formaldehyde gel (2% agarose, 20 mM MOPS, 8 mM sodium acetate, 2.2 mM formaldehyde, pH = 7.0). Three volumes of NorthernMax formaldehyde load dye (Ambion) was mixed with 1 volume of RNA, heated to 65 °C for 15 min, and chilled on ice for 5 min before loading. The gel was electrophoresed (50 V for 2 h) and equal loading confirmed via ethidium bromide straining of 18S and 28S rRNA (Figure 3C, top). The RNA was drawn onto a Biodyne B nylon membrane (Pall Life Science) in 10× SSC buffer (1.5 M sodium chloride, 0.15 M sodium citrate, pH = 7.0) via a standard blotting apparatus. The membrane was rinsed in 2× SSC buffer, dried, and auto-cross-linked with a UV Stratalinker 2400 (Stratagene). Hybridization and detection were carried out via protocols and reagents from the North2South chemiluminescent hybridization and detection kit (Pierce). Briefly, 400−500 μg of biotinylated probes, ketoRS3 biot 5′-/5Biosg/TGA GAC GCT GCT TAA CCG CTT C-3′ and ketoRS4 biot 5′-/5Biosg/TAA AGA AGT ATT CAG GAT CGG ACT G-3′, were incubated overnight at 55 °C, bound to a streptavidin−HRP conjugate, and detected with a luminol/enhancer stable peroxide solution (Pierce) (Figure 3C, bottom). Relative mRNA titers were determined by band density.

pApaRS Western Blot. Clones *AOX2*, A6 and B7, *YPT1*, D11 and B7, *ICL1*, E5 and H3, *FLD1*, E11 and F3, *GAP*, B7 and B10, and *AOX1*, E3 and E7 were cultured under test expression conditions, pelleted (3000*g*, 10 min), and lysed with 2 mL of YeastBuster (Novagen, Gibbstown, NJ) + 10 mM β-mercaptoethanol and Complete protease inhibitor tablets. Samples were cleared at 20000*g*, and 15 μL of the lysate was run on a 4−20% SDS−PAGE gel (1:15 h, 150 V). The protein was transferred to a 0.45 μm nitrocellulose membrane (Bio-Rad) using a Trans-Blot SD semidry transfer cell (Bio-Rad) in Tobin's transfer buffer (24 mM Tris base, 192 mM glycine, 20% ethanol) (2 h, 20 V, 100 mA). Residual protein on gel was stained with Coomassie (Figure 3B, top) to ensure equal loading. The membrane was blotted using standard Western blotting techniques (*20*) with an anti-His$_{6x}$-HRP conjugated antibody (Sigma-Aldrich) and detected with ECL (GE Healthcare) HRP chemiluminescence detection reagents and protocols (Figure 3B, bottom). Relative expression rates were determined by band density.

rHSA$_{E37X}$-ABT-510 Oxime Ligation. An ABT-510 peptide mimetic was synthesized (Anaspec, San Jose, CA) with a ϵ-(2-(aminooxy)acetyl)-L-lysine replacing the sixth L-norvaline residue (sequence: Ac-Sar-Gly-Val-D-aloIle-Thr-Lys(Aoa)-Ile-Arg-Pro-NEt, MW = 1097.3 Da). The peptide (2.25 mM, 0.5 mg) was added to 75 μM rHSA$_{E37pApa}$ or rHSA$_{WT}$ (1.0 mg) in 200 μL of oxime ligation buffer (1.5 M sodium chloride, 500 mM sodium acetate, pH = 4.4) and incubated overnight at 37 °C. The reactions were purified with a C8 Vydac HPLC column (300 mm, 200 Å, 5 μm; Grace) on a Dynamax HPLC (Rainin) (elution 40−46% acetonitrile in water, 0.1% TFA). Fractions were collected, combined, and analyzed via Coomassie-stained SDS−PAGE gel. Intact protein mass measurements were performed using a linear MALDI-TOF MS Biflex III (Bruker Daltonics, Billerica, MA) instrument with a sinapinic acid matrix. The mass difference between rHSA$_{WT}$ + peptide and rHSA$_{E37pApa}$ + peptide, less 60 Da owing to the E37pApa mutation (905 Da less 60 Da = 845 Da), was used to determine ligation efficiency (~77%). The mass of rHSA$_{WT}$ changed negligibly before and after treatment with the protein (data not shown).

pREAV-P$_{FLD1}$-(Synthetase$_{Tyr}$) Construction and Transformation. Unnatural aaRSs specific for tyrosine (wt), pBpa, pAzapa, pPpa, pMpa, and pIpa were amplified by PCR using the primers KETO-Koz-F and KetoRS R 6×His (described above), digested with *Nco*I and *Eag*I, and ligated into the similarly digested pREAV-P$_{FLD1}$-pApaRS (after removal of the pApaRS region via agarose gel purification). After sequence confirmation, the plasmids were transformed into GS200-rHSA$_{E37X}$ clone 1D12 as previously described to create GS200-rHSA$_{E37X}$/pREAV-P$_{FLD1}$-(*synthetase$_{Tyr}$*) (*HIS4*, *ARG4*, Gen^R, mut$^+$). Twelve clones were chosen from each transformation and screened via dot blot in 96-well format as previously described. The best producer was chosen from

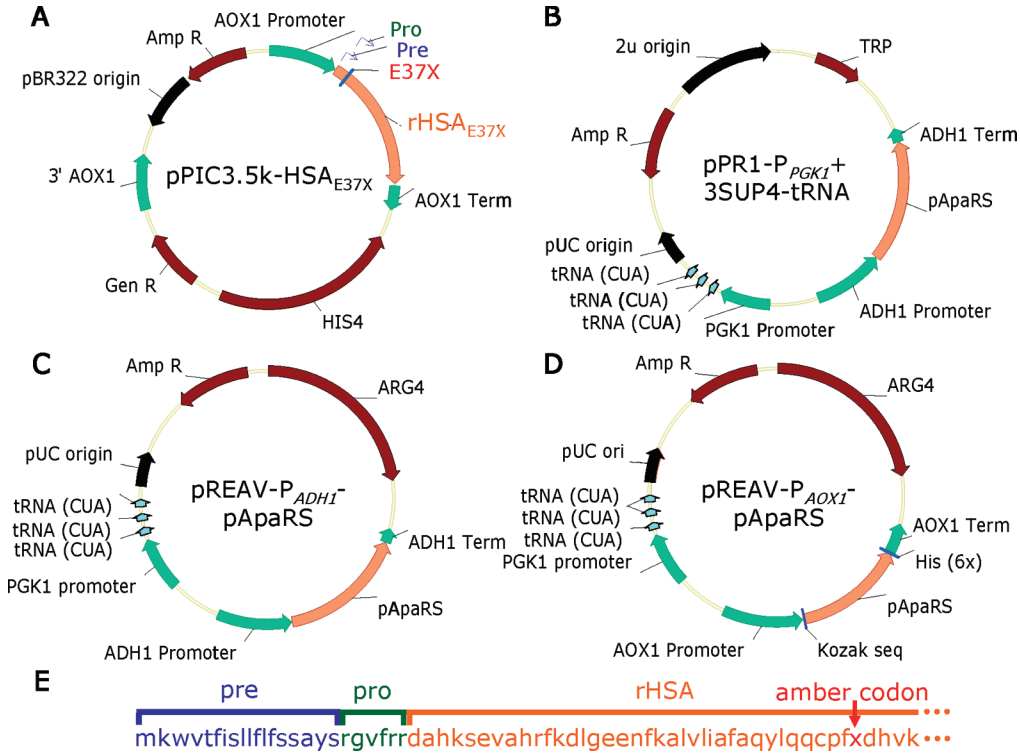

FIGURE 1: Vectors for amber suppression in eukaryotes illustrating markers (maroon), replication origins (black), target proteins (orange), control elements (green), and suppressor tRNAs ("tRNA (CUA)", light blue). (A) Map of the commercially available pPIC3.5k shuttle vector (*15*) for *in vivo* multicopy incorporation and expression in *P. pastoris*. rHSA$_{E37X}$ (orange) is subcloned between the *AOX1* promoter and terminator. (B) Optimized amber suppression vector for *S. cerevisiae* (*17*) harboring the pApaRS/tRNA pair under P$_{ADH1}$ control. tRNA$_{CUA}$ repeats are separated by regions from the *SUP4* gene (not labeled) and driven by P$_{PGK1}$. (C) Modified pPR1-P$_{PGK1}$ + 3SUP4-tRNA plasmid where the 2μ eukaryotic origin and *TRP* marker were replaced by *ARG4* to create pREAV-P$_{ADH1}$-pApaRS. (D) P$_{ADH1}$ and T$_{ADH1}$ were replaced by their *AOX1* counterparts to create pREAV-P$_{AOX1}$-pApaRS. (E) The first 61 amino acids of rHSA$_{E37X}$. The prepro leader peptide (blue, green) allows export of rHSA$_{E37X}$ into the media and is cleaved during transport to yield the mature protein (rHSA, orange) beginning with an aspartic acid. The 37th residue (X, red) of the mature rHSA denotes the unnatural amino acid incorporated in response to the amber codon.

each (Tyr, A9; pBpa, B7; pAzapa C9; pPpa, D6; pMpa, E6; and pIpa, F6) and compared to *FLD1*, E11, in test expressions (Figure 6C). Relative protein yields were determined by band density.

pREAV-P$_{FLD1}$-(Synthetase$_{Leu}$) Construction and Transformation. To create pREAV$_{Leu}$-P$_{FLD1}$, a section corresponding to three tandem repeats of tRNA$^{Leu5}_{CUA}$ lacking the 5′ CCA and separated by *SUP4* segments was synthesized (DNA 2.0, Menlo Park, CA) and PCR amplified using primers Leu tRNA F and Leu tRNA R. The resulting 643 bp product was digested with *Nhe*I and *Mlu*I and ligated into the similarly digested pREAV-P$_{FLD1}$-pApaRS (after removal of the tyrosyl tRNA via agarose gel purification) to create pREAV$_{Leu}$-P$_{FLD1}$-pApaRS. aaRSs with specificity for the DMNB-C and dansyl unnatural amino acids were amplified using primers LeuRS F and LeuRS R, digested with *Nco*I and *Not*I, and ligated into the similarly digested pREAV$_{Leu}$-P$_{FLD1}$-pApaRS (after removal of the pApaRS coding region via agarose gel purification) to create pREAV$_{Leu}$-P$_{FLD1}$-DMNB-S or pREAV$_{Leu}$-P$_{FLD1}$-dansyl (Figure 6D). After sequence confirmation, the plasmids were transformed into GS200-rHSA$_{E37X}$ (clone 1D12) and screened in 96-well dot blot format as described. Clones A:A5 (DMNB-C) and B:G12 (dansyl) were identified as successful producers grown under test expression conditions for 3 days postinduction in buffered minimal methanol (BMM) media. rHSA$_{WT}$ was expressed for 3 days in BMMY for comparison (Figure 6F). Incorporation of the dansyl and DMNB-C (analogue amino acid) was further confirmed via LC-MS/MS as described above (Supporting Information, Figure S7).

RESULTS AND DISCUSSION

Design of a Two-Gene Cassette Expression System. Due to the relative instability of autonomously replicating plasmids in *P. pastoris* (*21*), a system was devised in which the target gene of interest and the aaRS/tRNA$_{CUA}$ pair were encoded in cassettes on two separate plasmids and stably integrated into the genome. The double auxotroph, GS200 (*arg4*, *his4*), was used as the host strain for protein expression, and the gene of interest was inserted into the commercially available pPIC3.5k plasmid (*HIS4*, Gen^R) (Figure 1A) (*15*). rHSA was used as a model protein given its utility in producing fusion proteins or peptide bioconjugates that enhance the serum half-life of short-lived therapeutic polypeptides (*22*−*24*). Expression of rHSA in *E. coli* and *S. cerevisiae* is not practical due to the protein's complex disulfide cross-linkages. A Glu37TAG mutant rHSA (rHSA$_{E37X}$) was generated by PCR mutagenesis and expressed under the *AOX1* promoter and terminator to create pPIC3.5k-rHSA$_{E37X}$. Glu37 is in a solvent-accessible helix which should facilitate the conjugation of peptides to a chemically reactive unnatural amino acid (i.e., *p*-acetylphenylalanine, **1**) introduced at this site and ensure that incorporation of relatively bulky groups minimally disrupt native protein structure and folding. The 24 amino acid

mammalian "pre-pro" leader sequence of rHSA (Figure 1E) is fully compatible with expression in *P. pastoris* and allows export of the mature protein into the media (*25*). As a positive control for protein expression, wild-type rHSA ($rHSA_{WT}$) was used to create pPIC3.5k-$rHSA_{WT}$ in a similar fashion. Linearization of this plasmid in the 5′ *AOX1* promoter allows genomic integration of one or more copies of the cassette; generally more copies result in higher overall yields of target protein (*26*). Integration in this manner leaves the *AOX1* gene intact, retaining the yeast's ability to rapidly utilize methanol (mut^+ phenotype). Alternatively, gene replacement can be carried out by linearization on either side of the *AOX1* cassette, resulting in replacement of the *AOX1* gene by the pPIC3.5k vector (*15*). Yeast lacking *AOX1* relies on the weaker *AOX2* gene for methanol utilization and is phenotypically mut^s. Because expression of rHSA is commonly carried out with mut^s yeast (*27*), pPIC3.5k-$rHSA_{E37X}$ was linearized and used to replace the *AOX1* gene to yield GS200-$rHSA_{E37X}$ (*HIS4*, *arg4*, Gen^R, mut^s). Successful transformants grew normally on minimal media plates lacking histidine and on rich media plates containing up to 0.25 mg/mL aminoglycoside antibiotic Geneticin.

To integrate the orthogonal aaRS/$tRNA_{CUA}$ pair into the genome, the previously developed pPR1-P_{PGK1} + 3*SUP4*-$tRNA^{Tyr}_{CUA}$ vector (*17*) (Figure 1B) for recombinant overexpression in *S. cerevisiae* was modified. The *p*-acetylphenylalanine (pApa, 1) specific aminoacyl-tRNA synthetase (pApaRS), previously evolved in *S. cerevisiae* (*28*), was inserted between the alcohol dehydrogenase I promoter (P_{ADH1}) and terminator (T_{ADH1}) with a $His_{6\times}$ tag to assay its expression. The cognate *E. coli* $tRNA^{Tyr}_{CUA}$ lacking the 5′ CCA was inserted as three tandem repeats behind the phosphoglycerate kinase 1 promoter (P_{PGK1}). To aid in posttranscriptional processing, the tRNAs were flanked by regions from the yeast suppressor tRNA gene, *SUP4*, as previously described (*17*). Eukaryotic downstream processing adds the 5′ CCA that is required for tRNA function. The 2μ origin and phosphoribosyl anthranilate isomerase (*TRP*) marker of pPR1-P_{PGK1}+3*SUP4*-$tRNA^{Tyr}_{CUA}$ were replaced by the arginosuccinate lyase (*ARG4*) coding region to give the recombinant eukaryotic *ARG4* vector (pREAV-P_{ADH1}-pApaRS) (Figure 1C). Propagation of this cassette is only possible in the event of genomic incorporation since it lacks a eukaryotic origin. Linearization of pREAV-P_{ADH1}-pApaRS in the *ARG4* coding region and subsequent transformation into GS200-$rHSA_{E37X}$ gave the fully prototrophic *P. pastoris* GS200-$rHSA_{E37X}$/pREAV-P_{ADH1}-pApaRS (*HIS4*, *ARG4*, Gen^R, mut^s). As a positive control, the pREAV-P_{ADH1}-pApaRS vector was similarly cloned into GS200-$rHSA_{WT}$ to give a fully prototrophic, mut^s, $rHSA_{WT}$ expressing *P. pastoris* strain.

Amber Suppression in P. pastoris. Clones isolated from GS200-$rHSA_{WT}$/pREAV-P_{ADH1}-pApaRS produced full-length $rHSA_{WT}$ visible by Coomassie staining on an SDS−PAGE gel after 2−3 days when grown under methanol-inducing conditions. In contrast, clones from GS200-$rHSA_{E37X}$/pREAV-P_{ADH1}-pApaRS failed to produce full-length $rHSA_{E37pApa}$ when grown for 6 days with methanol as the primary carbon source and pApa amino acid supplementation (unpublished results). Genomic integration of all constructs was confirmed by genomic PCR (Supporting Information, Figure S1), and transcription of the $tRNA_{CUA}$ was found to be approximately 1.5 times greater than the same cassette in *S. cerevisiae* by Northern blot analysis (Figure 2A). However, no pApaRS was detectable by Western blot for the $His_{6\times}$ tag (Supporting Information, Figure S2). These results indicated that the lack of amber suppression was linked to poor expression of the pApaRS. Therefore, pREAV was further modified to drive expression of pApaRS with the powerful P_{AOX1} promoter, and an enhanced Kozak consensus sequence (ACCATGG) (*29*) was added to the 5′ end of the pApaRS gene. The *ADH1* terminator (T_{ADH1}) was also replaced by the *AOX1* terminator (T_{AOX1}) to give pREAV-P_{AOX1}-pApaRS (Figure 1D). Transformation into GS200-$rHSA_{E37X}$ yielded GS200-$rHSA_{E37X}$/pREAV-P_{AOX1}-pApaRS with the same phenotype as previously noted. Clones from this transformation produced full-length $rHSA_{E37pApa}$ only in the presence of methanol and pApa amino acid, at levels approximately 10−20% of identical clones harboring $rHSA_{WT}$. Protein was visible by SDS−PAGE gel 2−3 days post methanol induction and peaked 6 days after expression with methanol supplementation to 0.5% every 24 h (Figure 2B). Yeast lacking the pREAV cassette, pPIC3.5k cassette, methanol supplementation, or pApa amino acid failed to produce protein detectable by Coomassie staining. The lack of protein expression in the absence of pApa amino acid indicates that no cross-aminoacylation occurs between the pApaRS/$tRNA^{Tyr}_{CUA}$ pair and the endogenous aminoacylation machinery. Site-specific incorporation of pApa into $rHSA_{E37X}$ was confirmed by tryptic digest, LC-MS/MS (Figure 2C). The observed fragment ion shows that the unnatural amino acid was uniquely incorporated at residue 37.

Optimization of Expression. In an effort to optimize expression of $rHSA_{E37pApa}$, a GS200-$rHSA_{E37X}$/pREAV-P_{AOX1}-pApaRS (*HIS4*, *ARG4*, Gen^R) fast methanol utilization (mut^+) mutant was created by insertion of pPIC3.5k-$rHSA_{E37X}$ into the region 5′ of the *AOX1* gene locus (this retains the integrity of the *AOX1* gene). Genomic insertion in this manner can lead to multimerization, yielding tandem copies of both Gen^R and the gene of interest (*7*). The resulting clone displayed resistance to Geneticin up to 1.0 mg/mL, whereas the aforementioned mut^s clone died above 0.25 mg/mL Geneticin, consistent with the incorporation of multiple copies of the cassette (*7, 15*). Analysis of full-length $rHSA_{E37pApa}$ expression from isolated clones in the presence of methanol and pApa amino acid showed that approximately 1.5−2.0 times more protein was produced than with the mut^s counterpart (Supporting Information, Figure S3). To further increase yields of $rHSA_{E37pApa}$, six different promoters (including P_{AOX1}) were compared for their ability to drive pApaRS transcription in the pREAV vector. Transcript mRNA levels, pApaRS protein levels, and overall $rHSA_{E37pApa}$ yields were assayed. Two constitutive promoters derived from yeast GTP binding protein I (*YPT1*) (*30, 31*) and glyceraldehyde-3-phosphate dehydrogenase (*GAP*) (*12, 32*) and three methanol-inducible promoters from alcohol oxidase II (*AOX2*) (*33*), formaldehyde dehydrogenase I (*FLD1*) (*12*), and isocitrate lyase I (*ICL1*) (*12*) were chosen based on their compatibility with methanol induction. A truncated version of P_{AOX2} was used which enhances the promoter by deleting one of the two upstream repressor binding sequences (*33*). The use of the somewhat weaker P_{YPT1} and P_{GAP} promoters (*30*) could be useful in the event that overproduction of the synthetase is toxic to the yeast or sequesters cellular energy away from production of $rHSA_{E37X}$. All promoters were

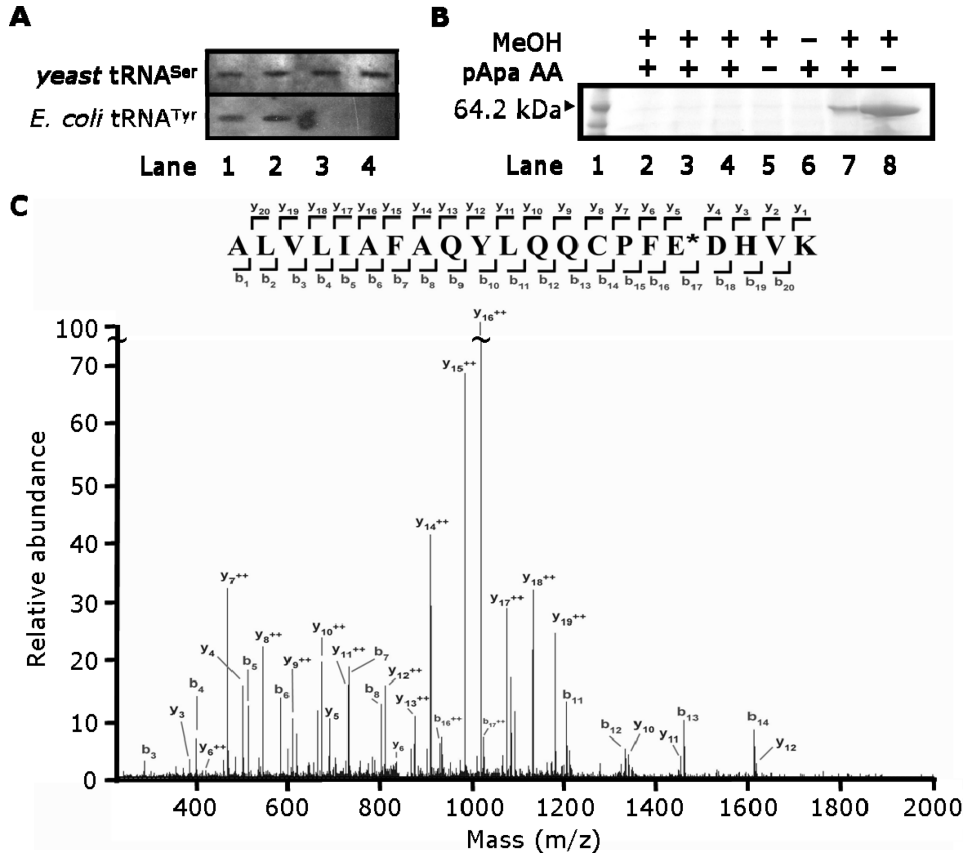

FIGURE 2: Amber suppression with pApa in *P. pastoris*. (A) A Northern blot (bottom gel) was used to assay suppressor tRNA transcription in *S. cerevisiae* + pPR1-P_{PGK1}-3SUP4-tRNA (lane 1) and *P. pastoris* + pREAV-P_{ADH1}-pApaRS (lane 2). For a negative control, lanes 3 and 4 are *S. cerevisiae* and *P. pastoris* strains lacking vectors, respectively. The top gel shows a Northern blot for the endogenous serine tRNA and illustrates equal miRNA preparation in all samples. (B) To assay the fidelity of the system, 25 μL of cleared media from 6 days of growth was analyzed on a denaturing SDS−PAGE gel and stained with Coomassie. Lane 2 is GS200; lane 3 is GS200-$rHSA_{E37X}$; lane 4 is GS200-pREAV-P_{AOX1}-pApaRS; lanes 5−7 are GS200-$rHSA_{E37X}$/pREAV-P_{AOX1}-pApaRS; and lane 8 is GS200-$rHSA_{WT}$/pREAV-P_{ADH1}-pApaRS. Amber suppression only occurs in yeast harboring both vectors and grown with methanol and pApa amino acid (pApa AA). (C) MS/MS fragmentation of a tryptic peptide (top) containing the unnatural amino acid pApa (denoted E*) at residue 37 of mature $rHSA_{E37pApa}$. The substitution is supported without ambiguity by the observed fragment ion series.

amplified by PCR from *P. pastoris* genomic DNA along with their 5′ untranslated regions. After sequence confirmation, each promoter was cloned into the pREAV vector 5′ of pApaRS in place of P_{AOX1} and transformed into the mut^+ GS200-$rHSA_{E37X}$ created previously (Figure 3A). The terminator remained T_{AOX1}. P-$_{promoter}$-pApaRS expression levels were monitored by Northern and Western blots after 6 days of methanol induction and compared to P_{AOX1}-pApaRS (Figure 3B). Due to inherent expression variability with *P. pastoris*, two clones were chosen for Western blot analysis, and the highest producer was analyzed by Northern blot. P_{FLD1} drove pApaRS transcription 4-fold better than P_{AOX1} at the mRNA level and produced 5-fold more pApaRS protein. P_{GAP}, P_{YPT1}, P_{ICL1}, and P_{AOX2} all showed lower pApaRS expression than P_{FLD1}. Consistent with this result, the overall amber suppression was highest with P_{FLD1}-pApa as measured by $rHSA_{E37pApa}$ expression into the media (Figure 4). Maximum yields were >150 mg/L or approximately 43% of $rHSA_{WT}$ yields (352 mg/L) (Supporting Information, Figures S5 and S6).

Oxime Ligation to $rHSA_{E37pApa}$. To demonstrate the utility of this modified rHSA as a carrier for bioactive peptides, an oxime ligation was carried out between the unique keto side chain of $rHSA_{E37pApa}$ and the antiangiogenic peptide ABT-510 (Figure 5). This thrombospondin-1 (TSP-1) properdin type 1 repeat mimetic exhibits potent antitumor activity in humans but suffers from rapid clearance by the kidneys when administered intravenously (*34*−*36*). A nine amino acid peptide mimetic was synthesized with a unique ϵ-(2-(aminooxy)acetyl)-L-lysine UAA in place of the sixth L-norvaline residue. Based on the known structure−activity relationships of TSP-1, modifications at this position are not expected to significantly alter biological activity (*37*). At pH <5 the aminooxy group undergoes a selective oxime ligation with the orthogonal keto group of pApa to covalently link the ABT-510 peptide to residue 37 of $rHSA_{E37pApa}$ (Figure 5A, top). Previous conjugation protocols used an aniline catalyst for efficient ligation (*38, 39*); however, oxime couplings to $rHSA_{E37pApa}$ proceeded in approximately 77% yield without the use of aniline in an overnight reaction using 75 μM $rHSA_{E37pApa}$ and a 30-fold excess of the peptide. The extent of derivatization of $rHSA_{E37pApa}$ with the peptide was confirmed by matrix-assisted laser desorption ionization (MALDI) mass spectrometry (Figure 5B). No conjugation was observed by MALDI mass spectrometry when $rHSA_{WT}$ (glutamic acid at residue 37) was treated with the aminooxy-modified ABT-510 peptide under identical conditions.

Addition of Eight UAAs to the Genetic Repertoire. To illustrate the generality of this newly created recombinant expression system, unnatural aaRSs evolved by the *S. cerevisiae* methodology were inserted into pREAV-P_{FLD1}. The aaRSs specific for *p*-benzoylphenylalanine (pBpa, photo-

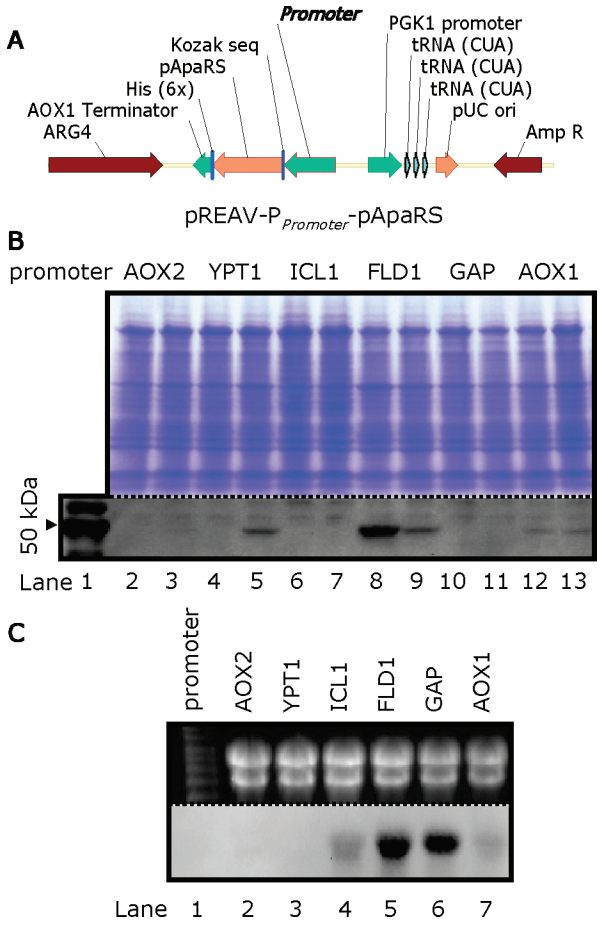

FIGURE 3: Comparison of pApaRS promoters for optimized amber suppression. (A) Linear map of pREAV-$P_{promoter}$-pApaRS illustrating the promoter region (green, red outline) being varied. Promoters were PCR amplified from genomic DNA (Supporting Information, Figure S4). (B) Two clones from each transformation of GS200-$rHSA_{E37X}$ with pREAV-$P_{promoter}$-pApaRS were grown with methanol as the primary carbon source for 6 days, lysed, and separated on an SDS−PAGE gel (top gel). The gel was stained with Coomassie to verify equal loading. Lysates were analyzed via Western blot for pApaRS-$His_{6\times}$ (bottom gel). (C) The clones which produced the most protein in (B) were analyzed by Northern blot for pApaRS mRNA transcription (bottom gel). Bands for the 18s and 28s rRNA were stained with ethidium bromide (top gel) to confirm RNA integrity and equal loading.

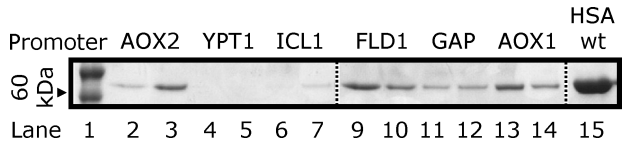

FIGURE 4: Amber suppression levels with P_{AOX2}, P_{YPT1}, P_{ICL1}, P_{FLD1}, P_{GAP}, or P_{AOX1} driven aaRS assayed by $rHSA_{E37pApa}$ in the media. The two clones from each promoter system were independently grown for 6 days with methanol as the primary carbon source and pApa amino acid. Twenty-five microliters of the cleared media was run on a denaturing SDS−PAGE gel and stained with Coomassie. $rHSA_{WT}$ (lane 15) was calculated to be 351.6 mg/L by band density with BSA control. By density, P_{FLD1} (lanes 9 and 10 averaged) expressed 43% as much protein, or 151.2 mg/L. Details on protein quantification can be found in Supporting Information, Figures S5 and S6.

cross-linker 3) (*2*), *p*-azidophenylalanine (pAzapa, photo-cross-linker, chemically reactive 4) (*40*), *p*-(propargyloxy)-phenylalanine (pPpa, chemically reactive 5) (*40*), *p*-methoxyphenylalanine (pMpa, structure/function probe 6) (*2*), and *p*-iodophenylalanine (pIpa, heavy atom 7) (*2*) were all inserted behind P_{FLD1} in the optimized pREAV-P_{FLD1} vector (Figure 6A,B). For comparison, wild-type *E. coli* tyrosyl-RS (wt, 2) was also inserted into the new expression vector. After transformation into GS200-$rHSA_{E37X}$ (*HIS4*, *arg4*, Gen^R, mut^+), selected clones were compared to the strain harboring pREAV-P_{FLD1}-pApaRS in their ability to suppress the amber mutation at position 37 in $rHSA_{E37X}$. Suppression yields were similar for the pApa and pAzapa mutants (40−45% the yield of $rHSA_{WT}$); all other mutants with the exception of pIpa expressed >20% the yield of $rHSA_{WT}$ (Figure 6C and unpublished results). No protein expression was observed in the absence of the cognate amino acid, demonstrating the high orthogonality of this new system.

Recently, a second orthogonal *E. coli* leucyl-derived RS/$tRNA_{CUA}$ pair (aaRS denoted as LeuRS) was generated to incorporate additional unnatural amino acids into proteins in *S. cerevisiae* (*41, 42*). To accommodate unnatural LeuRSs derived from this orthogonal pair in the new *P. pastoris* expression system, the tRNA region of pREAV-P_{FLD1} plasmid was modified. The existing $tRNA^{Tyr}_{CUA}$ cassette downstream of P_{PGK1} was excised and replaced by a coding region corresponding to three tandem repeats of $tRNA^{Leu5}_{CUA}$ lacking the 5′ CCA and separated by *SUP4* segments, as previously described, to create $pREAV_{Leu}$-P_{FLD1}. LeuRS mutants specific for 4,5-dimethoxy-2-nitrobenzylserine (DMNB-S, photo-caged serine 8) (*41*) and 3-amino-3-(5-(dimethylamino)naphthalene-1-sulfonamde)propanoic acid (dansylalanine, dansyl fluorophore 9) (*42*) were inserted behind P_{FLD1} to create $pREAV_{Leu}$-P_{FLD1}-LeuRS (Figure 6D-F). After transformation into the mut^+ GS200-$rHSA_{E37X}$, selected clones were used to express the corresponding mutant $rHSA_{E37X}$ (Figure 6F). The LeuRS mutant specific for DMNB-S was recently shown to accept the cysteine analogue of DMNB-S (DMNB-C), which was used in these expression experiments due to easier synthetic accessibility (unpublished results). Although small amounts of full-length protein (a background of 35% of $rHSA_{E37DMNB-S}$ for DMNB-S and 6% of $rHSA_{E37dansyl}$ for dansyl) were produced in the absence of the cognate amino acid, LC-MS/MS of a tryptic digest confirmed high fidelity of the system in the presence of the corresponding unnatural amino acid (Supporting Information, Figure S7). Indeed, nonspecific readthrough of a nonsense codon is often suppressed by the presence of an aminoacylated suppressor tRNA. Suppression yields were approximately 37% the yield of $rHSA_{WT}$ for $rHSA_{E37DMNB-C}$ and 23% the yield of $rHSA_{WT}$ for $rHSA_{E37dansyl}$ after 3 days of expression.

CONCLUSION

Previous attempts to optimize the expression of proteins containing unnatural amino acids in *S. cerevisiae* resulted in maximal yields of 8−15 mg/L in model systems, more than an order of magnitude less than demonstrated in the *P. pastoris* system developed here. Work in the Wang laboratory has recently shown that knockdown of the nonsense-mediated mRNA decay (NMD) pathway in yeast can increase protein expression up to 2-fold (*43*). Coupled with the use of a promoter derived from SNR52 to drive $tRNA_{CUA}$ transcription, they were able to achieve 300-fold higher yields of mutant protein than previously produced in *S. cerevisiae*, approximately 15 mg/L (*43*). Thus, knockout of the *UPF1* gene of the NMD pathway and use of the SNR52-$tRNA_{CUA}$

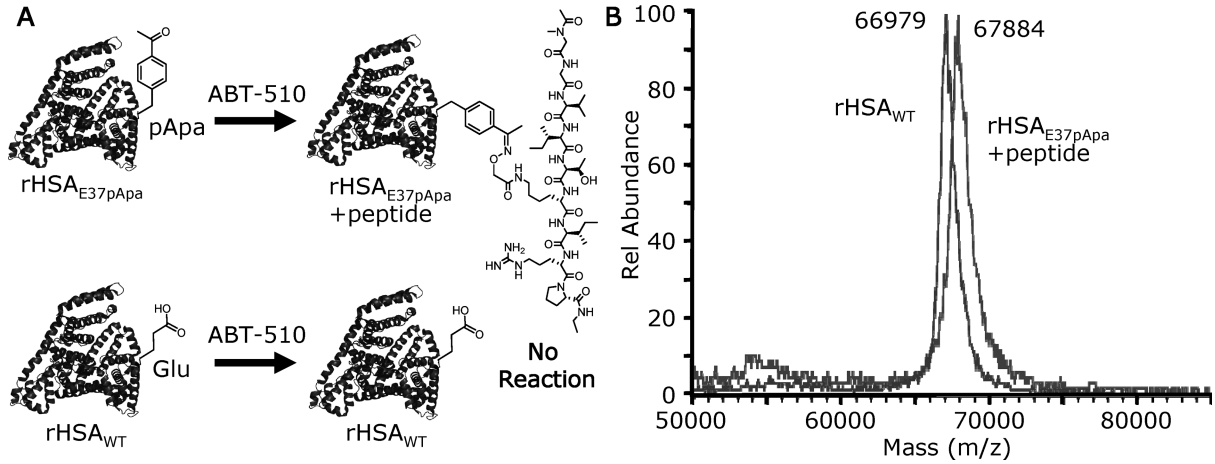

FIGURE 5: Oxime ligation of ABT-510 peptide to rHSA$_{E37pApa}$. (A) Schematic representation of ligation. The ABT-510 peptide harbors an ϵ-(2-(aminooxy)acetyl)-L-lysine as the sixth residue. Incubation of 75 μM rHSA$_{E37pApa}$ (top) with 2.25 mM peptide overnight at 37 °C results in the formation of an oxime linkage (top right). No reaction occurs with rHSA$_{WT}$ (bottom) under identical conditions. (B) MALDI mass spectrometry shows the extent of conjugation. Incubation of peptide with keto-containing rHSA$_{E37pApa}$ results in a 905 Da mass shift compared with incubation with rHSA$_{WT}$, indicating approximately 77% of rHSA$_{E37pApa}$ is linked to ABT-510.

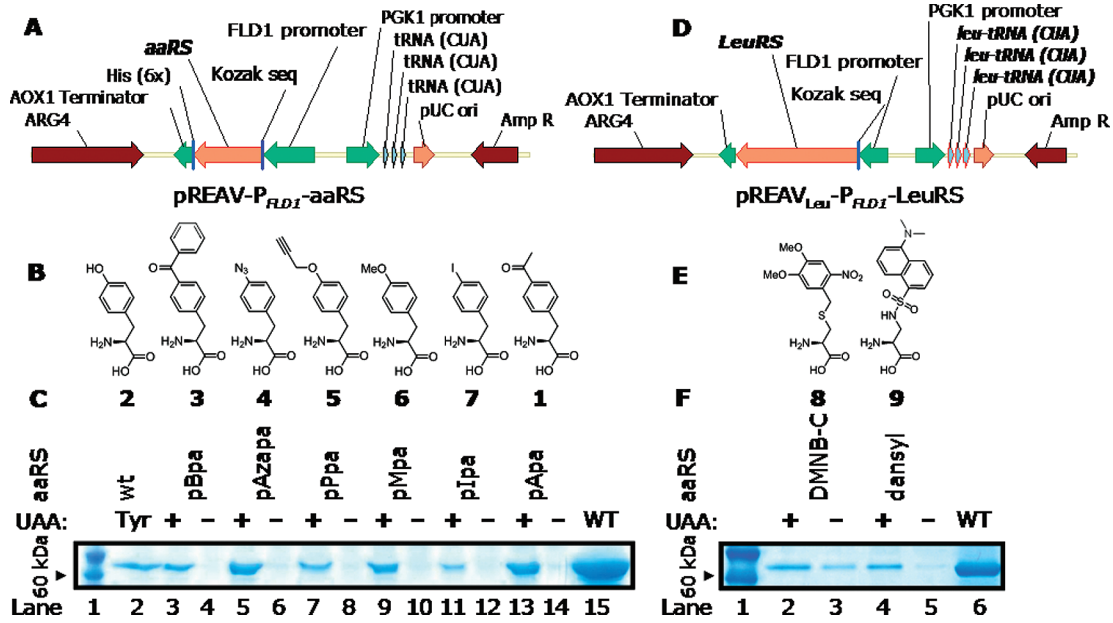

FIGURE 6: Addition of eight unnatural amino acids to the genetic repertoire. (A) Schematic of the optimized pREAV-P$_{FLD1}$ vector with *E. coli* tyrosyl-RS gene (orange) and tyrosyl suppressor tRNA cassette (tRNA (CUA), light blue). (B) Structures of six unnatural amino acids (**1**, **3**−**7**) and tyrosine (**2**) with specific *E. coli* tyrosyl-RS. (C) Expression of rHSA$_{E37X}$ (where X is defined as the unnatural amino acid) in the presence (+) and absence (−) of unnatural amino acids **1** and **3**−**7** with their corresponding aaRS. Twenty-five microliters of unpurified cleared media was run on a SDS−PAGE gel and stained with Coomassie. Lane 2 is rHSA$_{E37Y}$ expression with the wild-type (wt) tyrosyl-RS. Lane 15 is expression of rHSA$_{WT}$. (D) Schematic of the optimized pREAV$_{Leu}$-P$_{FLD1}$ vector with *E. coli* leucyl-RS gene (LeuRS, orange) and leucyl suppressor tRNA cassette (Leu-tRNA (CUA), light blue, red outline). (E) Structure of the DMNB-C and dansyl unnatural amino acids (**8**, **9**) with specific *E. coli* leucyl aaRSs. (F) Expression of rHSA$_{E37X}$ in the presence (+) and absence (−) of unnatural amino acids, **8** and **9**, with their corresponding LeuRS. Twenty-five microliters of unpurified cleared media from each protein expression was analyzed on an SDS−PAGE gel and stained with Coomassie. Lane 4 is expression of rHSA$_{WT}$, also after 3 days.

promoter system may further increase yields in *P. pastoris*. Additionally, work in the Kobayashi laboratory has demonstrated that yields of rHSA$_{WT}$ from *P. pastoris* are more than an order of magnitude better (>10 g/L) when expressed in fed-batch fermentation rather than in standard shake flasks (*10*).

In summary, we have extended the methodology for the biosynthetic incorporation of unnatural amino acids into methylotrophic yeast. Two aaRS/tRNA$_{CUA}$ pairs were shown to be orthogonal in *P. pastoris* and used to express mutant proteins with eight different unnatural amino acids in response to an amber codon at residue 37 of rHSA$_{E37X}$. Mutant proteins were expressed in high levels in shake flasks and with excellent fidelities. These results suggest that this expression system will be amenable to many other unnatural amino acids with synthetases currently being evolved in *S. cerevisiae* and is not limited to the unnatural amino acids or aaRS/tRNA$_{CUA}$ pairs discussed here. The high yields and fidelities of this new system should make it possible to obtain useful amounts of therapeutic proteins with unique biological and pharmacological properties. For example, chemistries such as oxime ligation or the copper-catalyzed 1,3-cycloaddition reaction ("click chemistry") can be exploited to site-specifically PEGylate or

cross-link proteins, metal ion binding amino acids can be incorporated to bind radioisotopes, and peptide, toxin, or siRNA conjugates can be made to carrier proteins such as HSA or targeting proteins such as antibodies. In addition, the aforementioned $rHSA_{E37pApa}$-ABT-510 conjugates are being tested in *in vitro* antiangiogenesis assays. The use of $rHSA_{E37pApa}$ as an endogenous, nonimmunogenic carrier is also being applied to other rapidly cleared peptides including glucagen-like peptide 1 mimetics (GLP-1) and parathyroid horomone (PTH) peptides.

ACKNOWLEDGMENT

We thank J. Cregg for the *P. pastoris* GS200 strain and pBLARG vector and S. Chen for insight into amber suppression in yeast.

SUPPORTING INFORMATION AVAILABLE

List of primers used in PCR (Table T1); DNA gel confirming plasmid genomic incorporation (Figure S1); Western blot proving underexpression of pApaRS under P_{ADH1} control (Figure S2); protein gel comparing mut^+ and mut^s phenotypes (Figure S3); genomic PCR of promoters (Figure S4); bar graph of relative expression yields under varying promoters (Figure S5); calculation of protein yield with BSA standard (Figure S6); LC-MS/MS data demonstrating incorporation of DMNB-C (Figure S7). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Xie, J., and Schultz, P. G. (2006) A chemical toolkit for proteins—an expanded genetic code. *Nat. Rev. Mol. Cell. Biol. 7*, 775–782.
2. Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z., and Schultz, P. G. (2003) An expanded eukaryotic genetic code. *Science 301*, 964–967.
3. Wang, L., Xie, J., and Schultz, P. G. (2006) Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct. 35*, 225–249.
4. Liu, W., Brock, A., Chen, S., and Schultz, P. G. (2007) Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods 4*, 239–244.
5. Wang, L., and Schultz, P. G. (2001) A general approach for the generation of orthogonal tRNAs. *Chem. Biol. 8*, 883–890.
6. Chin, J. W., Cropp, T. A., Chu, S., Meggers, E., and Schultz, P. G. (2003) Progress toward an expanded eukaryotic genetic code. *Chem. Biol. 10*, 511–519.
7. Cereghino, J. L., and Cregg, J. M. (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol. Rev. 24*, 45–66.
8. Shekhar, C. (2008) *Pichia* power: India's biotech industry puts unconventional yeast to work. *Chem. Biol. 15*, 201–202.
9. Clare, J. J., Rayment, F. B., Ballantine, S. P., Sreekrishna, K., and Romanos, M. A. (1991) High-level expression of tetanus toxin fragment C in *Pichia pastoris* strains containing multiple tandem integrations of the gene. *Biotechnology (New York) 9*, 455–460.
10. Ohya, T., Ohyama, M., and Kobayashi, K. (2005) Optimization of human serum albumin production in methylotrophic yeast *Pichia pastoris* by repeated fed-batch fermentation. *Biotechnol. Bioeng. 90*, 876–887.
11. Sreekrishna, K., Nelles, L., Potenz, R., Cruze, J., Mazzaferro, P., Fish, W., Fuke, M., Holden, K., Phelps, D., Wood, P., et al. (1989) High-level expression, purification, and characterization of recombinant human tumor necrosis factor synthesized in the methylotrophic yeast *Pichia pastoris*. *Biochemistry 28*, 4117–4125.
12. Cos, O., Ramon, R., Montesinos, J. L., and Valero, F. (2006) Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: a review. *Microb. Cell Fact. 5*, 17.
13. Cregg, J. M., Vedvick, T. S., and Raschke, W. C. (1993) Recent advances in the expression of foreign genes in *Pichia pastoris*. *Biotechnology (New York) 11*, 905–910.
14. Li, H., Sethuraman, N., Stadheim, T. A., Zha, D., Prinz, B., Ballew, N., Bobrowicz, P., Choi, B. K., Cook, W. J., Cukan, M., Houston-Cummings, N. R., Davidson, R., Gong, B., Hamilton, S. R., Hoopes, J. P., Jiang, Y., Kim, N., Mansfield, R., Nett, J. H., Rios, S., Strawbridge, R., Wildt, S., and Gerngross, T. U. (2006) Optimization of humanized IgGs in glycoengineered *Pichia pastoris*. *Nat. Biotechnol. 24*, 210–215.
15. Invitrogen Life, T. (2005)Multi-Copy *Pichia* Expression Kit, Version F, p 85, Invitrogen Life Technologies, Carlsbad, CA 92008.
16. Zheng, L., Baumann, U., and Reymond, J. L. (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res. 32*, e115.
17. Chen, S., Schultz, P. G., and Brock, A. (2007) An improved system for the generation and analysis of mutant proteins containing unnatural amino acids in *Saccharomyces cerevisiae*. *J. Mol. Biol. 371*, 112–122.
18. Matthiesen, R., Bauw, G., and Welinder, K. G. (2004) Use of performic acid oxidation to expand the mass distribution of tryptic peptides. *Anal. Chem. 76*, 6848–6852.
19. Licklider, L. J., Thoreen, C. C., Peng, J., and Gygi, S. P. (2002) Automation of nanoscale microcapillary liquid chromatography-tandem mass spectrometry with a vented column. *Anal. Chem. 74*, 3076–3083.
20. Burnette, W. N. (1981) "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate−polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal. Biochem. 112*, 195–203.
21. Higgins, D. R., and Cregg, J. M. (1998) Introduction to *Pichia pastoris*. *Methods Mol. Biol. 103*, 1–15.
22. Kim, J. G., Baggio, L. L., Bridon, D. P., Castaigne, J. P., Robitaille, M. F., Jette, L., Benquet, C., and Drucker, D. J. (2003) Development and characterization of a glucagon-like peptide 1-albumin conjugate: the ability to activate the glucagon-like peptide 1 receptor in vivo. *Diabetes 52*, 751–759.
23. Huang, Y. S., Chen, Z., Chen, Y. Q., Ma, G. C., Shan, J. F., Liu, W., and Zhou, L. F. (2008) Preparation and characterization of a novel exendin-4 human serum albumin fusion protein expressed in *Pichia pastoris*. *J. Pept. Sci. 14*, 588–595.
24. Chuang, V. T., Kragh-Hansen, U., and Otagiri, M. (2002) Pharmaceutical strategies utilizing recombinant human serum albumin. *Pharm. Res. 19*, 569–577.
25. Kobayashi, K. (2006) Summary of recombinant human serum albumin development. *Biologicals 34*, 55–59.
26. Buckholz, R. G., and Gleeson, M. A. (1991) Yeast systems for the commercial production of heterologous proteins. *Biotechnology 9*, 1067–1072.
27. Kupcsulik, B., and Sevella, B. (2005) Optimization of specific product formation rate by statistical and formal kinetic model descriptions of an HSA producing *Pichia pastoris* Mut(S) strain. *Chem. Biochem. Eng. Q. 19*, 99–108.
28. Zhang, Z., Smith, B. A., Wang, L., Brock, A., Cho, C., and Schultz, P. G. (2003) A new strategy for the site-specific modification of proteins in vivo. *Biochemistry 42*, 6735–6746.
29. Kozak, M. (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. U.S.A. 87*, 8301–8305.
30. Sears, I. B., O'Connor, J., Rossanese, O. W., and Glick, B. S. (1998) A versatile set of vectors for constitutive and regulated gene expression in *Pichia pastoris*. *Yeast 14*, 783–790.
31. Segev, N., Mulholland, J., and Botstein, D. (1988) The yeast GTP-binding YPT1 protein and a mammalian counterpart are associated with the secretion machinery. *Cell 52*, 915–924.
32. Waterham, H. R., Digan, M. E., Koutz, P. J., Lair, S. V., and Cregg, J. M. (1997) Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene 186*, 37–44.
33. Ohi, H., Miura, M., Hiramatsu, R., and Ohmura, T. (1994) The positive and negative cis-acting elements for methanol regulation in the *Pichia pastoris* AOX2 gene. *Mol. Gen. Genet. 243*, 489–499.
34. Hoekstra, R., de Vos, F. Y., Eskens, F. A., Gietema, J. A., van der Gaast, A., Groen, H. J., Knight, R. A., Carr, R. A., Humerickhouse, R. A., Verweij, J., and de Vries, E. G. (2005) Phase I safety, pharmacokinetic, and pharmacodynamic study of the thrombospondin-1-mimetic angiogenesis inhibitor ABT-510 in patients with advanced cancer. *J. Clin. Oncol. 23*, 5188–5197.

35. Yang, Q., Tian, Y., Liu, S., Zeine, R., Chlenski, A., Salwen, H. R., Henkin, J., and Cohn, S. L. (2007) Thrombospondin-1 peptide ABT-510 combined with valproic acid is an effective antiangiogenesis strategy in neuroblastoma. *Cancer Res. 67*, 1716–1724.
36. Reiher, F. K., Volpert, O. V., Jimenez, B., Crawford, S. E., Dinney, C. P., Henkin, J., Haviv, F., Bouck, N. P., and Campbell, S. C. (2002) Inhibition of tumor growth by systemic treatment with thrombospondin-1 peptide mimetics. *Int. J. Cancer 98*, 682–689.
37. Haviv, F., Bradley, M. F., Kalvin, D. M., Schneider, A. J., Davidson, D. J., Majest, S. M., McKay, L. M., Haskell, C. J., Bell, R. L., Nguyen, B., Marsh, K. C., Surber, B. W., Uchic, J. T., Ferrero, J., Wang, Y. C., Leal, J., Record, R. D., Hodde, J., Badylak, S. F., Lesniewski, R. R., and Henkin, J. (2005) Thrombospondin-1 mimetic peptide inhibitors of angiogenesis and tumor growth: design, synthesis, and optimization of pharmacokinetics and biological activities. *J. Med. Chem. 48*, 2838–2846.
38. Dirksen, A., Hackeng, T. M., and Dawson, P. E. (2006) Nucleophilic catalysis of oxime ligation. *Angew. Chem., Int. Ed. Engl. 45*, 7581–7584.
39. Dirksen, A., Dirksen, S., Hackeng, T. M., and Dawson, P. E. (2006) Nucleophilic catalysis of hydrazone formation and transimination: implications for dynamic covalent chemistry. *J. Am. Chem. Soc. 128*, 15602–15603.
40. Deiters, A., Cropp, T. A., Mukherji, M., Chin, J. W., Anderson, J. C., and Schultz, P. G. (2003) Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*. *J. Am. Chem. Soc. 125*, 11782–11783.
41. Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M., and Schultz, P. G. (2007) Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat. Chem. Biol. 3*, 769–772.
42. Summerer, D., Chen, S., Wu, N., Deiters, A., Chin, J. W., and Schultz, P. G. (2006) A genetically encoded fluorescent amino acid. *Proc. Natl. Acad. Sci. U.S.A. 103*, 9785–9789.
43. Wang, Q., and Wang, L. (2008) New methods enabling efficient incorporation of unnatural amino acids in yeast. *J. Am. Chem. Soc. 130*, 6066–6067.

BI802178K