Addition of the keto functional group to the genetic code of *Escherichia coli*

Lei Wang[†], Zhiwen Zhang[‡], Ansgar Brock[§], and Peter G. Schultz[‡¶]

[†]Department of Chemistry, University of California, Berkeley, CA 94720; [‡]Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037; and [§]Genomics Institute of Novartis Research Foundation, 10675 John Jay Hopkins Drive, San Diego, CA 92121

Edited by William F. DeGrado, University of Pennsylvania School of Medicine, Philadelphia, PA, and approved November 19, 2002 (received for review August 11, 2002)

Although the keto group is the most versatile of the functional groups in organic chemistry, it is absent in the genetically encoded amino acids. To overcome this natural limitation on protein biosynthesis, we have evolved an orthogonal tRNA-synthetase pair that makes possible the efficient incorporation of a keto amino acid, *p*-acetyl-L-phenylalanine, into proteins in *E. coli* with high translational fidelity in response to the amber nonsense codon. To demonstrate the utility of this keto amino acid, we have used it to modify a protein selectively with a small molecule fluorophore and biotin derivative. This additional genetically encoded amino acid should greatly expand our ability to manipulate protein structure and function both *in vitro* and in living cells.

The genetic codes of most known organisms encode the same common 20 amino acids as building blocks for the biosynthesis of proteins. Only in rare cases are selenocysteine (1) or pyrrolysine (2, 3) added. The side chains of the common amino acids comprise a surprisingly limited number of functional groups: nitrogen bases, carboxylic acids and amides, alcohols, and a thiol group, the remainder being simple alkanes or hydrophobic groups. The ability to augment the genetically encoded amino acids with new amino acids, for example, amino acids with metal chelating, fluorescent, redox active, photoactive, or spin-labeled side chains, would significantly enhance our ability to manipulate the structures and functions of proteins and perhaps living organisms themselves. Recently, we reported that, by adding new components to the translational machinery of *Escherichia coli*, one could site-specifically incorporate with high fidelity a number of unnatural amino acids (4–6) into proteins *in vivo*. We now demonstrate that this approach can be extended to add a keto-containing amino acid to the genetic code of *E. coli*, and that the unique reactivity of the keto group can be used to modify proteins selectively *in vitro* with a wide variety of agents.

The keto group is ubiquitous in organic chemistry, and participates in a large number of reactions from addition reactions to aldol condensations. Moreover, the unique reactivity of the keto group allows it to be selectively modified with hydrazide and hydroxylamine derivatives in the presence of the other amino acid side chains (7–9). Although present in cofactors (10), metabolites (11), and as a posttranslational modification to proteins (12), this important functional group is absent from the side chains of the common amino acids. To genetically encode this functional group in *E. coli* in the form of *p*-acetyl-L-phenylalanine, a tRNA-synthetase pair was evolved that is capable of inserting this amino acid site-specifically into proteins in *E. coli* in response to (and only in response to) an amber non-sense codon. Importantly, this tRNA-synthetase pair is orthogonal to its counterparts for the common 20 amino acids; i.e., the orthogonal synthetase (and only this synthetase) aminoacylates the orthogonal tRNA (and only this tRNA) with the unnatural amino acid only, and the resulting acylated tRNA inserts the unnatural amino acid only in response to the amber codon.

Materials and Methods

Preparation of *p*-Acetyl-L-Phenylalanine. Fmoc-*p*-acetyl-L-phenylalanine was purchased from RSP Amino Acid Analogues (Worcester, MA). This compound (1.0 g, 2.3 mmol) was stirred with 4 ml of piperidine [20% in dimethylformamide (DMF)] for 2 h at room temperature. The solvent was evaporated to obtain white powder. The solid was then resuspended in 10 ml of cold water [0.1% trifluoroacetic acid (TFA)], and the supernatant was collected by filtration. Preparative reverse-phase HPLC (Microsorb C18, Rainin Instruments) was used to separate the desired product from the reaction mixture (5–30% CH_3CN in H_2O with 0.1% TFA over 30 min). The eluant (t_R = 12 min) was lyophilized to obtain a white solid (0.45 g, 88%). ^{1}H NMR (400 MHz D_2O): δ 7.85–7.28 (m, 4H), 4.23 (dd, 1H, 5.4 Hz), 3.2 (m, 2H), 2.7 (s, 3H). MS electrospray ionization (ESI): $[M + 1]^+$ calculated for $C_{11}H_{13}NO_3$ 208.09, found 208.47.

Synthesis of *p*-Acetyl-(±)-L-Phenylalanine (13). N-bromosuccinimide (NBS) was recrystallized before usage. NBS (18.5 g, 105 mmol) was added to a stirred solution of 4-methyl acetophone (13.4 g, 100 mmol) in 400 ml of carbon tetrachloride, followed by the addition of 2′,2′-azobisiosbutyronitrile (AIBN; 0.43 g, 2.5 mmol). The reaction mixture was then heated to reflux for 4 h. After completion of reaction (TLC: 8:1/hexanes:EtOAc), the solution was washed with water (1 × 100 ml), 1 M aqueous HCl (3 × 100 ml), 0.5% aqueous $NaHCO_3$ (3 × 100 ml) and brine (1 × 100 ml). The organic layer was collected and dried over anhydrous $MgSO_4$, and solvent was evaporated to obtain a yellow solid which was recrystallized with hexanes to afford the desired 1-(4-bromoethyl-phenyl)thanone as a solid (16.8 g, 78%). Dry ethanol (50 ml) was added dropwise to pentane-washed sodium pieces (2.3 g, 0.1 mol) under an argon atmosphere over 15 min, and the solution was stirred for another 15 min. Solid diethyl acetamidomalonate (2.7 g, 10 mmol) was then added over 30 min with stirring, followed by the dropwise addition of 1-(4-bromoethyl-phenyl)thanone (2.1 g, 10 mmol) in dry ethanol over 90 min. After the mixture was heated to reflux overnight and cooled, diethyl ether (150 ml) and water (100 ml) were added to the solution. The organic layer was separated and washed successively with 0.5% $NaHCO_3$ (3 × 100 ml) and brine (1 × 100 ml). After drying over anhydrous $MgSO_4$, solvent was removed *in vacuo* to afford a brown gummy solid. Hexanes-dichloromethane (4:1) was added to the residue, and the insoluble material was filtered out and washed exhaustively with 10:1 dichloromethane-benzene to afford 2-acetylamino-2-(4-acetyl-benzyl)-malonic acid diethyl ester as a yellow solid (3.3 g, 95% crude yield). This compound was stirred with 4 M HCl in dioxane overnight. The mixture was then evaporated to dryness and recrystallized with water to afford *p*-acetyl-(±)-phenylalanine (13.2 g, 64% overall yield) as a white solid. ^{1}H NMR (400 MHz,

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: TyrRS, tyrosyl-tRNA synthetase; t_R, retention time.

¶To whom correspondence should be addressed. E-mail: schultz@scripps.edu.

 www.pnas.org/cgi/doi/10.1073/pnas.0234824100

D_2O): δ 7.85–7.28 (m, 4H), 4.27 (dd, 1H, 5.4 Hz), 3.30 (m, 2H), 2.68 (s, 3H). ^{13}C NMR (400 MHz, D_2O): δ 195.8, 174.3, 145.9, 133.1, 128.9, 127.8, 60.2, 38.3, 26.5. MS-ESI: $[M + 1]^+$ calculated for $C_{11}H_{13}NO_3$ 208.09, found 208.07.

Mutant Synthetase Evolution. In the positive selection, plasmid pYC-J17 was used to express the $mutRNA^{Tyr}_{CUA}$ gene and the chloramphenicol acetyl transferase (CAT) gene with a TAG stop codon at Asp-112 (4). Supercoiled DNA encoding the tyrosyl-tRNA synthetase (TyrRS) library was transformed into *E. coli* DH10B competent cells containing pYC-J17. Cells were then plated on minimal media plates containing 1% glycerol and 0.3 mM leucine (GMML) with 17 μg/ml tetracycline, 25 μg/ml kanamycin, 60 μg/ml of chloramphenicol, and 1 mM *p*-acetyl-L-phenylalanine. After incubation at 37°C for 40 h, colonies were pooled, and plasmids were isolated. Plasmids encoding mutant synthetases (pBK plasmids) were separated from pYC-J17 by using gel electrophoresis and transformed into *E. coli* DH10B competent cells containing pLWJ17B3 for negative selection. Plasmid pLWJ17B3 expresses the $mutRNA^{Tyr}_{CUA}$ under the control of the *lpp* promoter and *rrnC* terminator, and the barnase gene with three amber codons at Gln-2, Asp-44, and Gly-65 under the control of arabinose promoter. Transformed cells were grown on LB plates containing 0.2% arabinose, 50 μg/ml kanamycin, and 35 μg/ml chloramphenicol. After 8 h, cells were removed from the plate, and pBK plasmids were purified for further rounds of selection. In the second and third round of positive selection, the concentration of chloramphenicol was increased to 80 and 100 μg/ml, respectively. After three positive selections alternating with two negative selections, eleven mutant TyrRS were identified that afforded an IC_{50} value of 9 μg/ml chloramphenicol in the absence of *p*-acetyl-L-phenylalanine and 120 μg/ml chloramphenicol in the presence of *p*-acetyl-L-phenylalanine in an *in vivo* CAT assay (14). The protein sequences of these mutant TyrRS converged on three independent clones LW1, LW5, and LW6, although the codon usage of each mutant TyrRS differs.

Protein Expression and Purification. Plasmid pLEIZ was used to express the Z-domain gene with an amber codon at the 7th position and a COOH-terminal His-6 tag under the control of a bacteriophage *T5* promoter and t_0 terminator, and the $mutRNA^{Tyr}_{CUA}$ gene under the control of the *lpp* promoter and *rrnC* terminator. The mutant synthetase gene isolated from clone LW1 (LW1RS) was encoded in plasmid pBK-LW1RS under the control of the constitutive *E. coli* GlnRS promoter and terminator. *E. coli* DH10B cells cotransformed with pLEIZ and pBK-LW1RS were grown in minimal medium containing 1% glycerol and 0.3 mM leucine (GMML medium) with 25 μg/ml kanamycin, 34 μg/ml of chloramphenicol, and 1.0 mM *p*-acetyl-(±)-phenylalanine. When cells reached an OD_{600} of 0.5, isopropyl-β-D-thiogalactopyranoside (1 mM) was added to induce protein expression. After 5 h, cells were pelleted, and the protein was purified by Ni^{2+} affinity chromatography under denaturing conditions according to the manufacturer's protocol (Qiagen, Chatsworth, CA). Proteins were then desalted with a PD-10 column (Amersham Pharmacia) and eluted in water. The yield of protein was measured by Bradford assay (BCA kit, Bio-Rad). Aliquots of protein were used for SDS/PAGE and mass spectrometry.

***In Vitro* Protein Modification with Fluorescein Hydrazide and Biotin Hydrazide.** The purified WT and mutant Z domain proteins were exchanged into PBS solution (PBS buffer, 100 mM potassium phosphate, pH 6.5, 0.5 M sodium chloride) by dialysis. Fluorescein hydrazide 1 (Molecular Probes) or biotin hydrazide 2 (Molecular Probes) was dissolved in DMF and added into 0.07 μmol of each protein in silanized Eppendorf tubes to a final concentration of 1 mM. PBS buffer (pH 6.5) was added to bring the final volume to 0.5 ml. The reaction mixture was kept at 25°C for 18 h. Unreacted dye or biotin was removed from the protein by using a PD-10 column (Amersham Pharmacia), and proteins were eluted with PBS buffer. To determine the labeling efficiency, the eluted protein samples were then analyzed by reverse-phase HPLC (ZORBAX SB-C18, 4.6 mm × 250 mm, flow rate = 1.0 ml/min, 10 → 40% CH_3CN in aqueous 50 mM triethylamine acetate buffer, pH 7.0, over 70 min, Agilent, Palo Alto, CA). The retention time (t_R) for mutant Z domain without labeling was 39.3 min; the t_R for fluorescein hydrazide-labeled mutant Z domain was 40.7 min; the t_R for biotin hydrazide-labeled mutant Z domain was 40.9 min.

Fluorescence Spectrum Measurement. All fluorescence emission spectra were recorded by using a FluoroMax-2 spectrofluorometer (Instruments SA, Edison, NJ) with excitation at 490 nm, both excitation and emission bandpass of 4 nm, a photomultiplier tube voltage of 950 V, and a scan rate of 1 nm/sec. Ten nanomoles of each labeled protein were used. The reported spectra represent an average of three scans.

Results and Discussion

A Keto Amino Acid. The keto group provides a unique chemical reactivity not present in the common 20 amino acids because of its ability to participate in addition reactions involving either the carbonyl group or the acidic Cα position. This group also provides an alternative to the natural amino acid cysteine for the selective modification of proteins with a large variety of chemical reagents. The reactive thiol group of cysteine has been used extensively to attach various biophysical probes to proteins (15–22). Unfortunately, the labeling of single cysteine residues is often complicated by the presence of more than one reactive residue in a protein, as well as exchange reactions in the presence of free thiol when a disulfide linkage is used. Therefore, the availability of a nonproteinogenic amino acid with orthogonal reactivity makes possible selective modification of protein in cases where a single cysteine cannot be selectively labeled or where two different labels are needed. The keto group reacts readily with hydrazides, hydroxylamines, and semicarbazides under mild conditions in aqueous solution, and forms hydrazone, oxime, and semicarbazone linkages, respectively, which are stable under physiological conditions (23, 24).

Several methods have been developed to incorporate selectively the carbonyl group into peptides and small proteins. Initially, an aldehyde was introduced at the N termini of peptides by oxidizing N-terminal serine or threonine with periodate. The aldehyde group was coupled to biotin and fluorescent reporters (8) or protein fragments containing a COOH-terminal hydrazide through a hydrazone linkage (25). The carbonyl group introduced by this method is restricted to the N terminus, and the protein must be stable to oxidation. Solid phase peptide synthesis (SPPS) was later used for the preparation of peptide segments containing either a hydrazide or hydroxylamine, which subsequently react with a branched aldehyde core matrix to form peptide dendrimers (24, 26) or with a keto-containing peptide segment to form synthetic proteins (27). SPPS allows the keto group to be incorporated throughout the protein, but suffers the inherent difficulties associated with the synthesis of large peptides or proteins. This size limitation can be overcome in some cases by expressed protein ligation, in which a synthetic peptide is chemically ligated to the COOH terminus of recombinant proteins (28). A ketone group containing peptide was prepared by SPPS and ligated to the Src homology 3 domain of the Abelson protein tyrosine kinase (29).

An *in vitro* biosynthetic method has also been used to incorporate the keto group into proteins (7). In this method, the unnatural amino acid containing the keto group is chemically

acylated to an amber suppressor tRNA. When the acylated tRNA and the mutant gene are combined in an *in vitro* extract capable of supporting protein biosynthesis, the unnatural amino acid is selectively incorporated in response to a UAG codon. This method requires the suppressor tRNA to be chemically aminoacylated with the unnatural amino acid *in vitro*, and the acylated tRNA is consumed as a stoichiometric reagent during translation and cannot be regenerated, resulting in low protein yields. By evolving an orthogonal tRNA-synthetase pair with specificity for *p*-acetyl-L-phenylalanine, it should be possible to incorporate a keto amino acid into proteins in response to the UAG codon directly in living *E. coli* cells. There should be no size limitation on the target protein as long as it can be expressed in *E. coli*, and it should be possible to express large amounts of the mutant protein. Moreover, as long as the labeling reagent is cell-permeable and nontoxic, it may be possible to selectively introduce the label in whole cells.

Evolution of Mutant Synthetases with Specificities for *p*-Acetyl-L-Phenylalanine. The *Methanococcus jannaschii* TyrRS and a mutant tyrosine amber suppressor tRNA ($mutRNA^{Tyr}_{CUA}$) were used as the starting point for the generation of the orthogonal tRNA-synthetase pairs. Previously, this pair was shown to be orthogonal in *E. coli* (14, 30). To change the amino acid specificity of the TyrRS so that it charges *p*-acetyl-L-phenylalanine and not any of the common 20 amino acids, a library of *M. jannaschii* TyrRS mutants was generated and screened. The crystal structure of the homologous *Bacillus stearothermophilus* TyrRS (31) was used to identify those residues that are within 6.5 Å of the *para* position of the aryl ring of bound tyrosine. Five corresponding residues (Tyr-32, Glu-107, Asp-158, Ile-159, and Leu-162) in the active site of *M. jannaschii* TyrRS were randomly mutated by PCR to generate a library 1.6×10^9 in size (4). This TyrRS mutant library was first passed through a positive selection in the presence of 1 mM *p*-acetyl-L-phenylalanine, which is based on the suppression of an amber stop codon at nonessential residue (Asp-112) in CAT gene encoded on plasmid pYC-J17 (4) in *E. coli*. Cells surviving in chloramphenicol must encode a mutant synthetase that aminoacylates the $mutRNA^{Tyr}_{CUA}$ with either a common amino acid(s) or *p*-acetyl-L-phenylalanine. DNA encoding the mutant synthetases was then isolated and transformed into a negative-selection strain expressing the gene of a toxic protein, barnase, containing three amber codons at permissive sites (encoded on plasmid pLWJ17B3). Cells encoding a mutant synthetase that charges the $mutRNA^{Tyr}_{CUA}$ with natural amino acids will produce barnase and die. Because no *p*-acetyl-L-phenylalanine was added to the growth medium in the negative selection, survivors must encode a synthetase with specificity for the unnatural amino acid. After three rounds of positive selection at increasing concentrations of chloramphenicol, alternating with two rounds of negative selection, a number of clones emerged whose survival in chloramphenicol depended on the addition of *p*-acetyl-L-phenylalanine. These TyrRSs were characterized by using an *in vivo* assay based on the suppression of the Asp-112TAG codon in the CAT gene (14). Eleven TyrRS mutants were identified. Cells expressing the selected synthetase and the $mutRNA^{Tyr}_{CUA}$ survived in the absence of *p*-acetyl-L-phenylalanine on 9 μg/ml chloramphenicol on minimal media plates containing 1% glycerol and 0.3 mM leucine (GMML plate); in the presence of this unnatural amino acid, cells survived in 120 μg/ml chloramphenicol on GMML plates. This result suggests that the selected mutant synthetase has higher activity for *p*-acetyl-L-phenylalanine than for natural amino acids. Sequencing the DNA of these mutants revealed that they converge on three independent mutants on the protein level (LW1, LW5, and LW6), although they have different codon usage for amino acids. The active site mutations of the mutant synthetases are listed in Table 1. Based on the crystal structure of the homologous TyrRS from *B. stearothermophilus*, the conserved side chain of *M. jannaschii* Tyr-32 and Asp-158 likely form hydrogen bonds with the hydroxyl group of the substrate tyrosine. In the mutant synthetases, Tyr-32 is mutated to either Leu or Ala, and Asp-158 is mutated to Gly-158. These mutations should disfavor the binding of tyrosine and may, at the same time, create extra room to accommodate the methyl group of *p*-acetyl-L-phenylalanine. Determination of the x-ray crystal structures of the mutants should clarify the exact roles of these mutations.

Table 1. Amino acid residues in the WT *M. jannaschii* (Mj) TyrRS and the evolved mutant synthetases with specificities for *p*-acetyl-L-phenylalanine

Amino acid residue	32	158	159	162	167
WT Mj TyrRS	Tyr	Asp	Ile	Leu	Ala
LW1	Leu	Gly	Cys	Arg	Ala
LW5	Leu	Gly	Thr	Arg	Ala
LW6	Ala	Gly	Gly	Leu	Ile

Characterization of Mutant Protein Containing *p*-Acetyl-L-Phenylalanine. To test the ability of the evolved synthetase and the $mutRNA^{Tyr}_{CUA}$ to incorporate selectively *p*-acetyl-L-phenylalanine into proteins, an amber stop codon was substituted at a permissive site (Lys-7) in the gene for the Z domain of staphylococcal protein A (32) with a COOH-terminal His-6 tag. Z domain has a molecular weight of ≈7.9 kDa, so its mass can be measured with very high accuracy by using ion cyclotron resonance (ICR) MS. Cells transformed with the $mutRNA^{Tyr}_{CUA}$, LW1RS, and Z domain gene (Lys-7TAG) were grown in the presence of 1 mM *p*-acetyl-(±)-phenylalanine. The addition of the unnatural amino acid did not affect the growth rate of cells. The mutant protein was purified by Ni^{2+} affinity chromatography with an overall isolated yield of 3.6 mg/liter in minimal media. For comparison, the yield of Z domain was 9.2 mg/liter in minimal media when the mutant TyrRS was replaced with the WT TyrRS. No Z domain was obtained in the absence of either *p*-acetyl-(±)-phenylalanine, the $mutRNA^{Tyr}_{CUA}$, or LW1RS (Fig. 1), indicating a very high fidelity in the incorporation of the unnatural amino acid at this site. We have also been successful

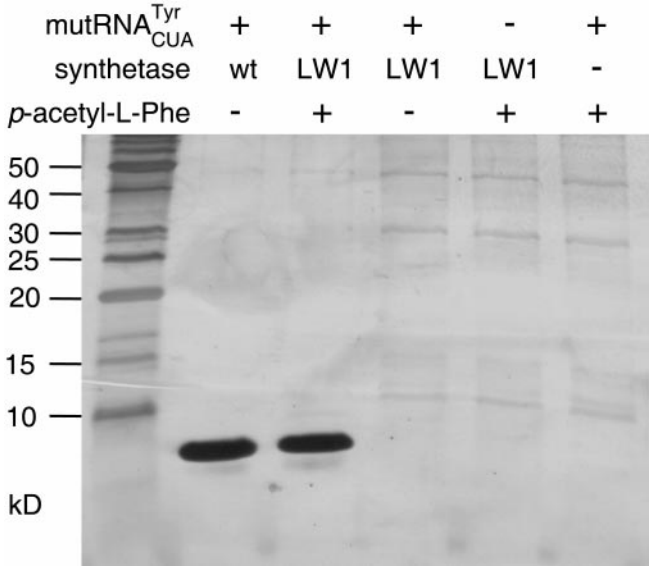

Fig. 1. SDS/PAGE analysis of Z domain accumulated under different expression conditions. The left lane is a molecular mass marker.

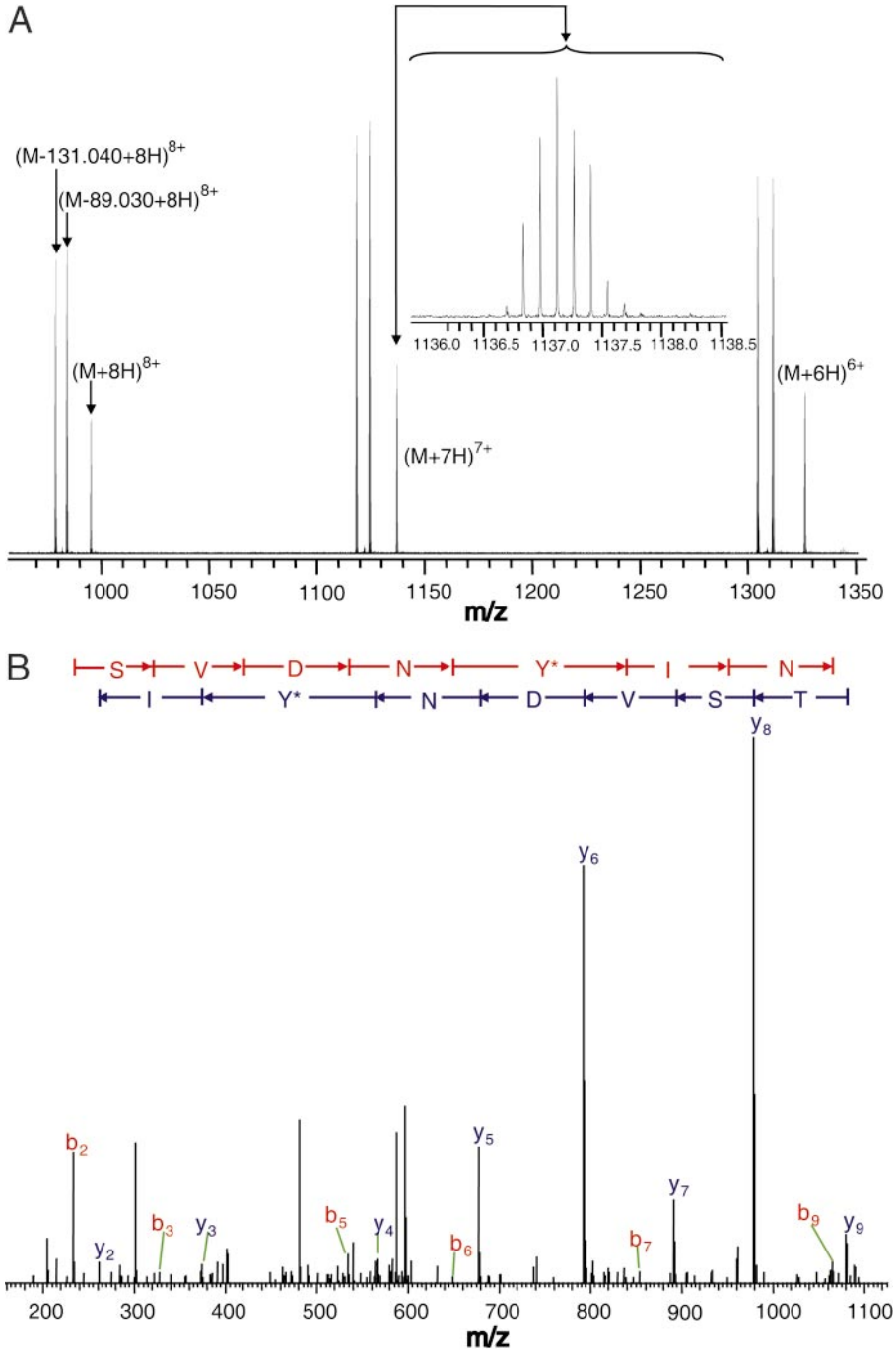

Fig. 2. (*A*) High-resolution FT-ICR mass spectrum of the intact mutant protein Z domain containing *p*-acetyl-L-phenylalanine. A series of peaks corresponding to different charge states of the protein can be observed. In each series, there are three peaks corresponding to the protein without the first methionine, its acetylated form, and the intact protein as labeled for the 8^+ charge state. (*Inset*) The expansion of the molecular peak of the Z domain protein from the 7^+ isotopic cluster. (*B*) Tandem mass spectrum of the NH_2-terminal peptide MTSVDNY*INK. The partial sequence of TSVDNY*IN of the peptide containing *p*-acetyl-L-phenylalanine (Y*) can be assigned from the annotated b (red) and y (blue) ion series.

in incorporating *p*-acetyl-L-phenylalanine into other proteins such as Cdc42.

Both the WT Z domain protein expressed by mutRNA$^{Tyr}_{CUA}$/WT TyrRS and the mutant Z domain protein expressed by the mutRNA$^{Tyr}_{CUA}$/LW1RS were analyzed by electrospray ionization Fourier transform (FT)-ICR MS. For the WT Z domain protein, three peaks were observed with masses corresponding to the intact protein, the protein without the first methionine, and the acetylated form of the protein without the first methionine (confirmed by tandem mass spectrometric analysis of the N-terminal tryptic-digested peptide fragment). For the mutant Z domain protein (Fig. 2*A*), the experimental monoisotopic mass of the intact protein was 7,949.893 Da, which is within 2.2 ppm of the theoretical mass of 7,949.874 Da. Two other peaks correspond to the protein without the first methionine ($M_{Experimental}$ = 7,818.838 Da, $M_{Theoretical}$ = 7,818.833 Da) and its acetylated form ($M_{Experimental}$ = 7,860.843 Da, $M_{Theoretical}$ = 7,860.844 Da), respectively. No peaks corresponding to mutant proteins with any other amino acid at the amber codon

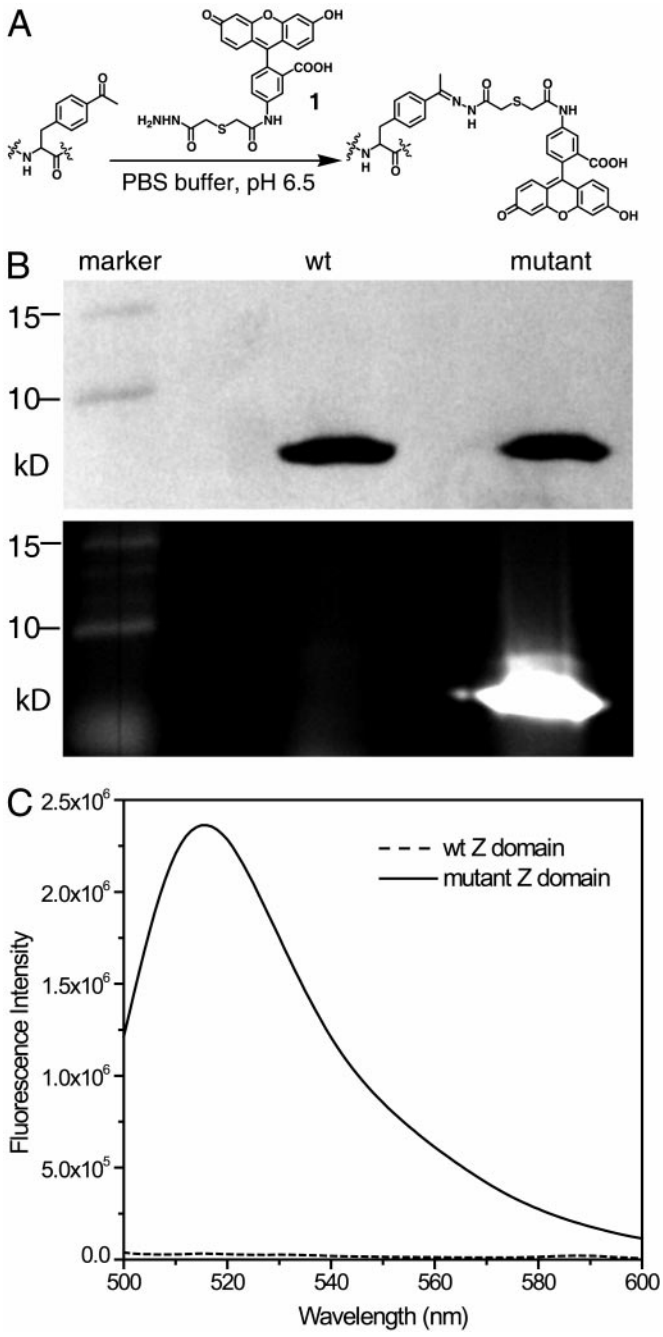

Fig. 3. *In vitro* labeling of mutant Z domain containing *p*-acetyl-L-phenylalanine with fluorescein hydrazide 1. (*A*) Labeling reaction of *p*-acetyl-L-phenylalanine by fluorescein hydrazide 1. (*B*) Silver-stained SDS/PAGE (*Upper*) analysis and fluorescence imaging (*Lower*) of WT and mutant Z domain labeled with fluorescein hydrazide 1. (*C*) Fluorescence spectra for WT and mutant Z domain labeled with fluorescein hydrazide 1.

position were observed in the spectra. The signal-to-noise ratio of >1,500 observed in the intact protein mass spectrum translates to a fidelity for the incorporation of *p*-acetyl-L-phenylalanine of >99.8%. Liquid chromatography tandem mass spectrometry of the tryptic digest was carried out to confirm the sequence of the NH_2-terminal peptide. The precursor ion at 606.23 Da, which corresponds to the doubly charged molecular ion of the NH_2-terminal tryptic peptide MTSVDNY*INK, was isolated and fragmented with an ion trap mass spectrometer.

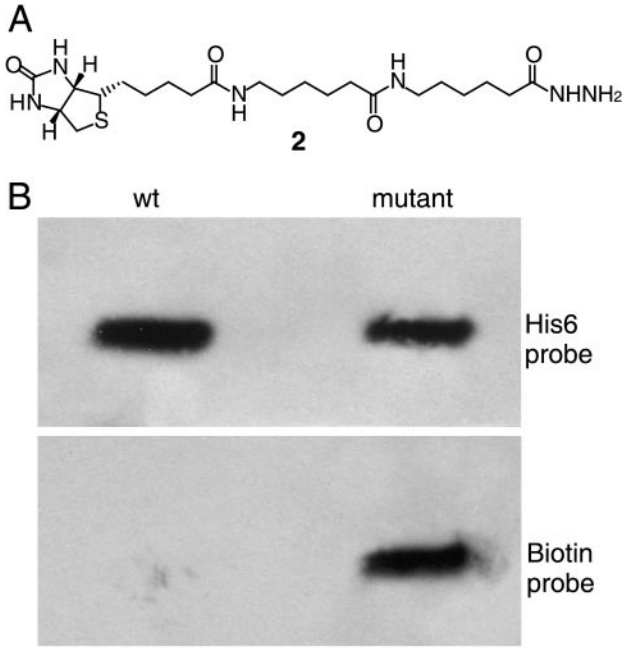

Fig. 4. *In vitro* labeling of mutant Z domain containing *p*-acetyl-L-phenylalanine with biotin hydrazide 2. (*A*) The structure of the biotin hydrazide derivative used, 6-({6-[(biotinoyl)amino]hexanoyl}amino)hexanoic acid hydrazide (Molecular Probes). (*B*) Western blot analysis of WT and mutant Z domain labeled by biotin hydrazide 2.

The fragment ion masses could be unambiguously assigned as shown in Fig. 2*B*, confirming the site-specific incorporation of *p*-acetyl-L-phenylalanine. These results clearly demonstrate that the evolved synthetase together with the mutRNA$^{Tyr}_{CUA}$ incorporate *p*-acetyl-L-phenylalanine and not any natural amino acid into the position encoded by the amber codon and at no other positions.

Site-Specific Protein Modification with Fluorescein Hydrazide. Next, we determined whether the keto group of *p*-acetyl-L-phenylalanine could serve as a chemical handle for the site-specific modification of proteins *in vitro*. The purified mutant *p*-acetyl-L-phenylalanine Z domain protein (mutant Z domain) and WT Z domain protein were treated with 1 mM fluorescein hydrazide 1 (Fig. 3*A*) at 25°C for 18 h in phosphate buffer. After the reaction, proteins were separated from excess fluorescein hydrazide by size exclusion chromatography and analyzed with SDS/PAGE. The gel was first imaged with a fluoroimaging system and then silver-stained (Fig. 3*B*). The band for mutant Z domain shows a fluorescent signal, whereas no fluorescence can be detected from the WT Z domain band. Aliquots of these two proteins were used to measure the fluorescence spectrum with 490 nm excitation (Fig. 3*C*). Only the Z domain protein containing *p*-acetyl-L-phenylalanine shows a fluorescence spectrum similar to that of fluorescein. No fluorescence signal was detected for WT Z domain, indicating that the labeling reaction occurred only between the hydrazide and the ketone, and not any existing functional groups in the WT protein. The labeled product was analyzed with quadrupole time-of-flight (QTOF) MS. An experimental monoisotopic mass of 8,425.160 Da ($M_{Theoretical}$ = 8,424.958 Da) was obtained, confirming that the fluorescein hydrazide reacted with the mutant Z domain protein in a molar ratio of 1:1. To determine the labeling extent, the reaction mixture was separated by HPLC. The ratio of the peak area of the labeled Z domain over that of the unlabeled Z domain was 90 ± 5%.

Site-Specific Protein Modification with Biotin Hydrazide. To demonstrate the generality of this approach, we also labeled Z domain with the biotin hydrazide derivative 2 (Fig. 4*A*). The purified mutant and WT Z domain were treated with 1 mM biotin hydrazide 2 in phosphate buffer at 25°C for 18 h. After dialysis against phosphate buffer to remove excess biotin hydrazide, the proteins were subject to SDS/PAGE. Separated proteins were transferred to nitrocellulose membrane and probed with a biotin-specific avidin–horseradish peroxidase conjugate (Fig. 4*B*). As expected, only the mutant Z domain containing *p*-acetyl-L-phenylalanine was detected, indicating that it was labeled with biotin hydrazide. No signal was observed for WT Z domain. The labeling efficiency was 80 ± 10% as determined by HPLC analysis, as described in the fluorescein-labeling experiment. The labeled protein was confirmed by QTOF MS ($M_{Experimental}$ = 8,416.236, $M_{Theoretical}$ = 8,416.146 Da) to be the product formed between one molecule of biotin hydrazide and one molecule of mutant Z domain. These experiments demonstrate the excellent specificity of the ketone handle for the *in vitro* modification of proteins.

Conclusion

In summary, we have site-specifically incorporated an additional functional group, the keto group, into proteins *in vivo*. This functional group can be selectively and efficiently labeled with fluorescein and biotin *in vitro* by a specific chemical reaction between the keto group and hydrazide derivatives. By using this approach, it should be possible to label proteins selectively with a wide variety of other hydrazide or hydroxylamine derivatives including sugars, spin labels, metal chelators, crosslinking agents, polyethers, fatty acids, and toxins. Such modifications can be used to probe protein structure and function to generate proteins with enhanced catalytic or therapeutic properties, or for the development of bioassays using either immobilized or soluble proteins. The ability to site-specifically incorporate a unique chemical handle into proteins directly in a living cell may make possible the *in vivo* modification of proteins with small molecule fluorophores for the *in vivo* imaging of protein localization, protein movement, and conformational changes in proteins at molecular resolution. Finally, it will be of interest to determine through either directed or random mutagenesis whether keto amino acids can enhance protein function directly, for example, by forming Schiff base intermediates that participate in catalysis or intramolecular or intermolecular protein crosslinks.

This work was supported by Department of Energy Grant DE-FG03-00ER45812, National Institutes of Health Grant GM62159, and The Skaggs Research Institute (TSRI). This is TSRI manuscript 15233-CH.

1. Bock, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B. & Zinoni, F. (1991) *Mol. Microbiol.* **5,** 515–520.
2. Srinivasan, G., James, C. M. & Krzycki, J. A. (2002) *Science* **296,** 1459–1462.
3. Hao, B., Gong, W., Ferguson, T. K., James, C. M., Krzycki, J. A. & Chan, M. K. (2002) *Science* **296,** 1462–1466.
4. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. (2001) *Science* **292,** 498–500.
5. Wang, L., Brock, A. & Schultz, P. G. (2002) *J. Am. Chem. Soc.* **124,** 1836–1837.
6. Zhang, Z., Wang, L., Brock, A. & Schultz, P. G. (2002) *Angew. Chem. Int. Ed. Engl.* **41,** 2840–2842.
7. Cornish, V. W., Hahn, K. M. & Schultz, P. G. (1996) *J. Am. Chem. Soc.* **118,** 8150–8151.
8. Geoghegan, K. F. & Stroh, J. G. (1992) *Bioconjugate Chem.* **3,** 138–146.
9. Mahal, L. K., Yarema, K. J. & Bertozzi, C. R. (1997) *Science* **276,** 1125–1128.
10. Begley, T. P., Kinsland, C., Taylor, S., Tandon, M., Nicewonger, R., Wu, M., Chiu, H., Kelleher, N., Campobasso, N. & Zhang, Y. (1997) *Top. Curr. Chem.* **195,** 93–142.
11. Diaz, E., Ferrandez, A., Prieto, M. A. & Garcia, J. L. (2001) *Microbiol. Mol. Biol. Rev.* **65,** 523–569.
12. Okeley, N. M. & van der Donk, W. A. (2000) *Chem. Biol.* **7,** R159–R171.
13. Cleland, G. H. (1969) *J. Org. Chem.* **34,** 744–747.
14. Wang, L. & Schultz, P. G. (2001) *Chem. Biol.* **8,** 883–890.
15. Creighton, T. E. (1986) *Methods Enzymol.* **131,** 83–106.
16. Altenbach, C., Marti, T., Khorana, H. G. & Hubbell, W. L. (1990) *Science* **248,** 1088–1092.
17. Brinkley, M. (1992) *Bioconjugate Chem.* **3,** 2–13.
18. Giuliano, K. A., Post, P. L., Hahn, K. M. & Taylor, D. L. (1995) *Annu. Rev. Biophys. Biomol. Struct.* **24,** 405–434.
19. Mannuzzu, L. M., Moronne, M. M. & Isacoff, E. Y. (1996) *Science* **271,** 213–216.
20. Griffin, B. A., Adams, S. R. & Tsien, R. Y. (1998) *Science* **281,** 269–272.
21. Llopis, J., Adams, S. R., McCaffery, J. M., Teter, K., Kulomaa, M. S., Machen, T. E., Moore, H. P., Tsien, R. Y. & Griffin, B. A. (2000) *Methods Enzymol.* **327,** 546–564.
22. Gaietta, G., Deerinck, T. J., Adams, S. R., Bouwer, J., Tour, O., Laird, D. W., Sosinsky, G. E., Tsien, R. Y. & Ellisman, M. H. (2002) *Science* **296,** 503–507.
23. Jencks, W. P. (1959) *J. Am. Chem. Soc.* **81,** 475–481.
24. Shao, J. & Tam, J. P. (1995) *J. Am. Chem. Soc.* **117,** 3893–3899.
25. Gaertner, H. F., Offord, R. E., Cotton, R., Timms, D., Camble, R. & Rose, K. (1994) *J. Biol. Chem.* **269,** 7224–7230.
26. Rose, K. (1994) *J. Am. Chem. Soc.* **116,** 30–33.
27. Canne, L. E., Ferre-D'Amare, A. R., Burley, S. K. & Kent, S. B. H. (1995) *J. Am. Chem. Soc.* **117,** 2998–3007.
28. Muir, T. W., Sondhi, D. & Cole, P. A. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 6705–6710.
29. Ayers, B., Blaschke, U. K., Camarero, J. A., Cotton, G. J., Holford, M. & Muir, T. W. (1999) *Biopolymers* **51,** 343–354.
30. Wang, L., Magliery, T. J., Liu, D. R. & Schultz, P. G. (2000) *J. Am. Chem. Soc.* **122,** 5010–5011.
31. Brick, P., Bhat, T. N. & Blow, D. M. (1989) *J. Mol. Biol.* **208,** 83–98.
32. Nilsson, B., Moks, T., Jansson, B., Abrahmsen, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T. A. & Uhlen, M. (1987) *Protein Eng.* **1,** 107–113.