Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry Letters
Bioorganic & Medicinal Chemistry Letters 16 (2006) 5356–5359

Genetic introduction of a diketone-containing amino acid into proteins

Huaqiang Zeng, Jianming Xie and Peter G. Schultz*

Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

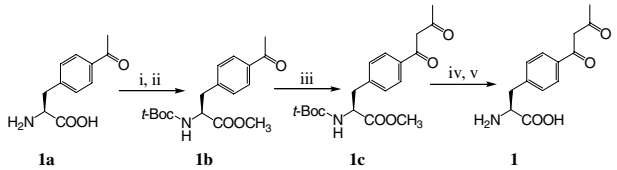
Received 22 June 2006; revised 21 July 2006; accepted 24 July 2006
Available online 24 August 2006

Abstract—An orthogonal tRNA/aminoacyl-tRNA synthetase pair was evolved that makes possible the site-specific incorporation of an unnatural amino acid bearing a β-diketone side chain into proteins in *Escherichia coli* with high translational efficiency and fidelity. Proteins containing this unnatural amino acid can be efficiently and selectively modified with hydroxylamine derivatives of fluorophores and other biophysical probes.

Recently, we showed that by adding a unique amber suppressor tRNA ($mutRNA^{Tyr}_{CUA}$)/aminoacyl-tRNA synthetase (*Mj*TyrRS) to the translational machinery of *Escherichia coli*, unnatural amino acids can be incorporated into proteins in response to nonsense and frameshift codons with high translational fidelity and efficiency.[1–5] This approach has been used to genetically encode over 30 unnatural amino acids including glycosylated, photoreactive, metal binding, and heavy atom-containing amino acids.[6–8] In addition, a series of chemically reactive amino acids with ketone-, azide-, and acetylene-containing side chains have been selectively incorporated into proteins. These mutant proteins can be subsequently modified by nonpeptidic molecules with useful biological and/or physical properties (e.g., polyethylene glycol, biotin, glycomimetics, fluorophores, cross-linking agents, and cytotoxic molecules).[9–13] Because of the unique reactivity of these amino acid side chains, such bioconjugation reactions are highly selective. Herein, we extend this methodology to the genetic incorporation of diketone-containing amino acid **1** in *E. coli*. The β-diketone group can react to form stable complexes with a number of functional groups including alkoxyamines, hydrazides, and primary amines.

Diketone **1** was synthesized from *p*-acetyl-L-phenylalanine containing *tert*-butyloxycarbonyl (*t*-Boc) and methyl ester-protecting groups (Scheme 1). The second carbonyl group was added using potassium *tert*-butoxide in the mixed solvent, 2:3 [v/v] methyl acetate/THF.[14] Removal of *t*-Boc group with TFA, followed by alkaline hydrolysis, afforded the desired diketone-containing unnatural amino acid **1** in an overall yield of 40%.

We previously generated an amber suppressor tRNA/aminoacyl-tRNA synthetase pair ($mutRNA^{Tyr}_{CUA}$-*Mj*TyrRS) derived from the $tRNA^{Tyr}$/TyrRS pair of *Methanococcus jannaschii*, which has been used to selectively and efficiently incorporate a large number of unnatural amino acids in *E. coli* in response to the TAG codon.[9,10,12,15] On the basis of a crystal structure of the *M. jannaschii* TyrRS-tRNA(Tyr)-L-tyrosine complex,[16] six residues (Tyr^{32}, Leu^{65}, Phe^{108}, Gln^{109}, Asp^{158}, and Leu^{162}) in the tyrosine-binding site of *M. jannaschii*



Scheme 1. Reagents and conditions: (i) $SOCl_2$, anhydrous MeOH, 1 h at rt, 90%; (ii) *t*-$(Boc)_2O$, Et_3N, CH_2Cl_2, 2 h at rt, 85%; (iii) potassium *tert*-butoxide, 40% methyl acetate in THF, 1 h at rt, 60%; (iv) TFA, CH_2Cl_2, 5 h at rt, 92%; (v) NaOH, 30% H_2O in MeOH, 6 h at rt, followed by addition of aqueous concentrated HCl, >90%.

Keywords: Bioorganic chemistry; Protein modification; In vivo selection; Aminoacyl-tRNA synthetase; Diketone-containing amino acid.
* Corresponding author. E-mail: schultz@scripps.edu

0960-894X/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2006.07.094

TyrRS were randomly mutated. The mutant *Mj*TyrRS (mut*Mj*TyrRS) library (>10^9 clones) was then passed through three rounds of positive selection in the presence of 1 mM unnatural amino acid **1** (based on the suppression of an amber stop codon in the chloramphenicol acetyl transferase (CAT) gene), alternated with two rounds of negative selection in the absence of **1** (based on the suppression of three amber stop codons introduced into the toxic barnase gene).[9] One clone emerged whose survival in chloramphenicol was dependent on the presence of **1**; this mutant *Mj*TyrRS supported cell growth in 120 μg mL^{-1} chloramphenicol in the presence of **1**, and up to 10 μg mL^{-1} chloramphenicol in the absence of **1**. This result suggests that the mutant synthetase has higher specificity for **1** than for endogenous amino acids. Sequencing revealed the following mutations: $Tyr^{32} \rightarrow Gly^{32}$, $Asp^{158} \rightarrow Gly^{158}$, $Phe^{65} \rightarrow Val^{65}$, $Phe^{108} \rightarrow Thr^{108}$, and $Leu^{162} \rightarrow Ser^{162}$. The Tyr^{32} and Asp^{158} mutations remove the two hydrogen bonds to tyrosine in the wildtype (wt) synthetase, consistent with the decreased activity of the enzyme toward its natural substrate.

To confirm that the observed phenotype is caused by the site-specific incorporation of **1** by the mutRNA^{Tyr}_{CUA}–mut*Mj*TyrRS pair, an amber stop codon was substituted at a permissive site (Lys^7) in the gene encoding the Z-domain protein[17] fused to a C-terminal hexameric His tag. Cells transformed with mutRNA^{Tyr}_{CUA}, mut*Mj*TyrRS, and the mutant Z-domain gene were grown in the presence or absence of 1 mM **1** in minimal medium containing 1% glycerol and 0.3 mM leucine (GMML medium). The expression of the $Lys^7 \rightarrow TAG^7$ Z-domain protein by mut*Mj*TyrRS was induced by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and protein was purified by Ni^{2+} affinity chromatography. Analysis by SDS–PAGE revealed that the expression of mutant protein was strongly dependent on **1**—the yield of purified mutant protein was 1.6 mg/liter of culture in the presence of **1**, and insignificant in the absence of **1** (Fig. 1), consistent with a high fidelity for the incorporation of **1** into proteins by mut*Mj*TyrRS. For comparison, the yield of the wt Z-domain protein using wt *Mj*TyrRS that charges only endogenous tyrosine in response to the amber stop codon was 3.0 mg/liter of culture under the identical conditions. Additional evidence for the site-specific incorporation of **1** was obtained by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS).[18] In addition to the observation of an experimental average mass of 7997 Da ($M_{Theoretical}$ = 7997 Da) for the intact mutant protein, a major peak corresponding to the protein without the first methionine moiety ($M_{Experimental}$ = 7866 Da, $M_{Theoretical}$ = 7866 Da) was also detected (Fig. 2). The other two peaks ($M_{Experimental}$ = 8039 Da and $M_{Experimental}$ = 7908 Da) correspond to the mutant proteins with or without the first methionine in their acetylated forms ($M_{Theoretical}$ = 8039 Da and $M_{Theoretical}$ = 7908 Da), respectively.[19] These results confirm a high fidelity for the incorporation of **1** by mutRNA^{Tyr}_{CUA}/mut*Mj*TyrRS pair.

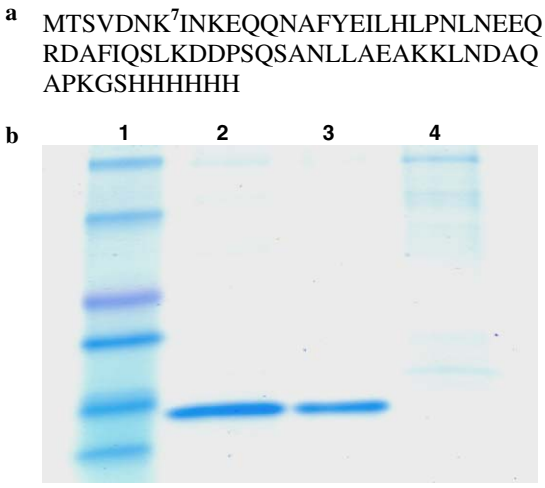


Figure 1. (a) The sequence of the Z-domain protein. The codon for Lys^7 was mutated to TAG in order to introduce amino acied **1**. (b) Gelcode Blue-stained SDS–PAGE analysis of $Lys^7 \rightarrow TAG^7$ Z-domain protein expressed under different conditions. Lane 1: molecular mass marker; lane 2: expression with wt *Mj*TyrRS; lane 3: expression with mut*Mj*TyrRS in the presence of **1**; lane 4: expression with mut*Mj*TyrRS in the absence of **1**. wt = wildtype.

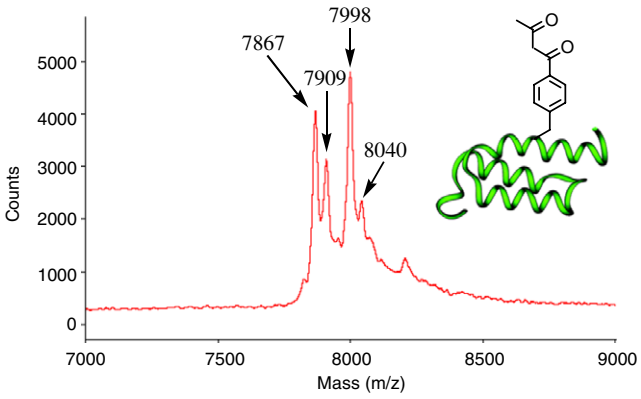


Figure 2. MALDI-TOF analysis of the mutant Z-domain protein containing **1**.[18] The observed masses of intact mutant protein in its protonated and acetylated forms are 7998 and 8040 Da, respectively. The observed masses of mutant protein without the first methionine in its protonated and acetylated forms are 7867 and 7909 Da, respectively. All of these experimental data are in excellent agreement with theoretically calculated values (see text for more detail).

The possibility of using the diketone moiety as a chemical handle for the selective modification of protein with exogenous agents was then tested by labeling diketone-containing protein with a biotin hydroxylamine derivative (Fig. 3a, M_W = 331.39, purchased from Molecular Probes).[9] The purified mutant and wt Z-domain proteins (~0.1 mM) were treated with 1.0 mM biotin hydroxylamine in PBS buffer (20 mM sodium phosphate, 150 mM NaCl) at pH 4.0 at 25 °C for 12 h. After dialysis against distilled water to remove excess biotin hydroxylamine, the proteins were analyzed by MALDI-TOF MS. Experimental average masses of 8315 Da ($M_{Theoretical}$ = 8310 Da for biotin-labeled intact mutant protein), 8182 Da ($M_{Theoretical}$ = 8179 Da for biotin-labeled

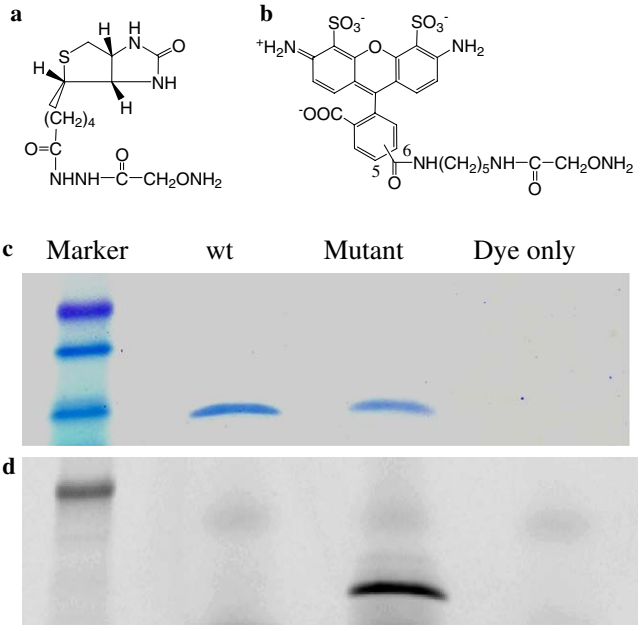


Figure 3. In vitro labeling of the mutant Z-domain protein containing **1** with hydroxylamine derivatives of (a) biotin and (b) Alexa Fluor 488. (c) Gelcode Blue-stained SDS/PAGE analysis of wt and mutant Z-domain proteins treated with Alexa Fluor 488 hydroxylamine. (d) Fluorescence imaging of wt and mutant Z-domain proteins treated with Alexa Fluor 488 hydroxylamine. wt = wildtype.

mutant protein without the first methionine residue), and 8225 Da ($M_{\text{Theoretical}}$ = 8221 Da for biotin-labeled mutant protein without the first methionine residue in its acetylated form) were found, confirming that biotin hydroxylamine reacted with the mutant Z-domain proteins in a molar ratio of 1:1. As expected, no labeling products were detected for wt Z-domain protein, indicating that the labeling reaction occurred only between the hydroxylamine and the diketone group, but not any existing functional groups in the wt protein. Based on the integration ratio of the peaks in the MALDI-TOF spectrum, the labeling efficiency was estimated to be >85% (Fig. 4c).[20]

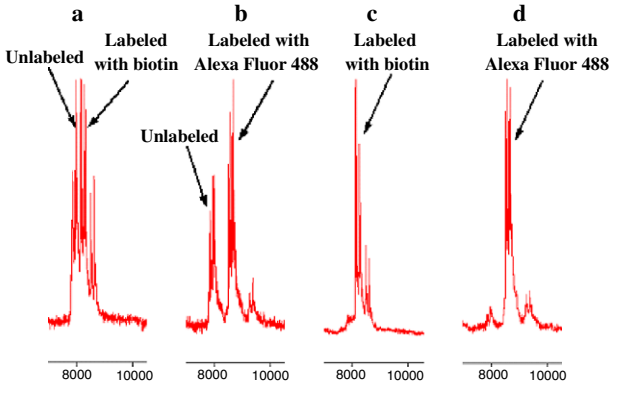


Figure 4. MALDI-TOF analysis[18] of doping experiments with comparable amounts of unlabeled diketone-containing Z-domain protein and its corresponding labeled versions with either (a) biotin or (b) Alexa Fluor 488, and of the diketone-containing Z-domain protein labeled with (c) biotin and (d) Alexa Fluor 488.[20]

To demonstrate the generality of this approach, the labeling of Z-domain protein with a fluorophore hydroxylamine derivative was also carried out. The purified mutant and wt Z-domain proteins (~0.05 mM) were treated with 0.5 mM Alexa Fluor 488 hydroxylamine derivative (Fig. 3b, Molecular Probes) in PBS buffer at pH 4.0 at 25 °C for 12 h (Fig. 3b). After conjugation, proteins were dialyzed against distilled water to remove excess fluorophore and then analyzed by SDS–PAGE. The gel was first imaged with a Phosphoimager[21] (Fig. 3d) and then stained with Gelcode Blue (Fig. 3c). The band for mutant Z-domain shows a fluorescent signal, whereas no fluorescence at the same position can be detected for either the wt Z-domain or dye molecules alone, indicating that Alexa Fluor 488 hydroxylamine derivative was selectively coupled to the diketone group in the presence of the other amino acid side chains in the protein. The labeled protein was confirmed by MALDI-TOF MS ($M_{\text{Experimental}}$ = 8678 Da and $M_{\text{Theoretical}}$ = 8668 Da for the dye-labeled intact protein) to be the product formed between one molecule of Alexa Fluor 488 hydroxylamine and one molecule of mutant Z-domain. The labeling efficiency was >85% as estimated by MALDI-TOF analysis (Fig. 4d).[20]

The present study demonstrates the site-specific incorporation of the β-diketone moiety into proteins in vivo with high fidelity and good efficiency by means of an amber suppressor tRNA/aminoacyl-tRNA synthetase pair. The β-diketone can serve as a unique chemical handle for the efficient and selective modification of target proteins with a host of exogenous alkoxyamine derivatives in vitro. Additionally, the β-diketone group may be able to participate in carbon–carbon bond forming Michael reactions.[22,23] Moreover, an enamine adduct stabilized by a six-membered intramolecular hydrogen bond can be formed between the β-diketone and a primary amine.[24] Consistent with this observation, we found that, while an imine adduct formed between an aryl monoketone and butyl amine readily undergoes hydrolysis at pH 4.0–10.4, the enamine derived from the β-diketone is highly resilient toward hydrolysis in the same pH range.[25] Consequently, by placing a diketone-containing unnatural amino acid immediately adjacent to a lysine residue situated in a favorable hydrophobic environment, one may be able to form intermolecular or intramolecular protein crosslinks.

Acknowledgment

This work was supported by a Department of Energy Grant DE-FG03-00ER46051 to P.G.S.

References and notes

1. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498.
2. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z. W.; Schultz, P. G. *Science* **2003**, *301*, 964.

3. Anderson, J. C.; Wu, N.; Santoro, S. W.; Lakshman, V.; King, D. S.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7566.
4. Zhang, Z. W.; Alfonta, L.; Tian, F.; Bursulaya, B.; Uryu, S.; King, D. S.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 8882.
5. Zhang, Z. W.; Gildersleeve, J.; Yang, Y. Y.; Xu, R.; Loo, J. A.; Uryu, S.; Wong, C. H.; Schultz, P. G. *Science* **2004**, *303*, 371.
6. Wang, L.; Schultz, P. G. *Chem. Commun.* **2002**, 1.
7. Wang, L.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2005**, *44*, 34.
8. Xie, J.; Schultz, P. G. *Curr. Opin. Chem. Biol.* **2005**, *9*, 548.
9. Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 56.
10. Zhang, Z.; Smith, B. A. C.; Wang, L.; Brock, A.; Cho, C.; Schultz, P. G. *Biochemistry* **2003**, *42*, 6735.
11. Liu, H.; Wang, L.; Brock, A.; Wong, C.-H.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 1702.
12. Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 11782.
13. Tsao, M. L.; Tian, F.; Schultz, P. G. *ChemBioChem* **2005**, *6*, 2147.
14. To a solution of **1b** (3.85 g, 12.0 mmol) in methyl acetate (24 mL) was added dropwise 36 mL of 1 M potassium *tert*-butoxide in THF at room temperature, and the solution was stirred for 45 min. After the removal of THF, CH_2Cl_2 (300 mL) was added, which was washed twice with distilled water, once with brine and dried over anhydrous Na_2SO_4. The solvent was removed and the resulting solid was subjected to the flash chromatography on a silical gel (30:70 [v/v] ethyl acetate/hexane) to afford pure product **1c** (2.60 g, 60%) as a white solid. ^1H NMR (399 MHz, $CDCl_3$, 298 K): δ 7.80 (d, J = 8.4 Hz, 2H), 7.21 (d, J = 8.0 Hz, 2H), 6.26 (s, 2H), 4.99 (m, 1H), 4.61 (m, 1H), 3.72 (s, 3H), 3.38–2.90 (m, 2H), 2.20 (s, 3H), 1.42 (s, 9H). Exact mass *m*/*z* without Boc group calcd for $C_{14}H_{17}NO_4$ 263.1, found (LC/MS) 263.1.
15. Alfonta, L.; Zhang, Z.; Uryu, S.; Loo, J. A.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 14662.
16. Kobayashi, T.; Nureki, O.; Ishitani, R.; Yaremchuk, A.; Tukalo, M.; Cusack, S.; Sakamoto, K.; Yokoyama, S. *Nat. Struct. Biol.* **2003**, *10*, 425.
17. Nilsson, B.; Moks, T.; Jansson, B.; Abrahmsen, L.; Elmblad, A.; Holmgren, E.; Henrichson, C.; Jones, T. A.; Uhlen, M. *Protein Eng.* **1987**, *1*, 107.
18. MALDI-TOF MS experiments were conducted on Applied Biosystems DE-STR mass spectrometer at the Scripps Center for Mass Spectrometry. The Z-domain protein solution (0.5 μL, ~0.05 mg/mL) was mixed with the matrix (0.5 μL, prepared by dissolving α-cyano-4-hydroxycinnamic acid in a mixture of water and acetonitrile or methanol) and deposited on a metal plate that is used in the mass spectrometer. Data were acquired in reflector/negative ion mode by using an accelerating voltage of 25,000 V, grid voltage 90%, guide wire 0.3%, and an extraction delay time of 250 ns.
19. Although the enzyme that catalyzes acetylation of Z-domain protein is not known, it is very likely that protein acetylation of the Z-domain protein is mediated by known or unknown bacterial *N*-acetyltransferases at either the N-terminus or one of five internal lysine residues found in Z-domain protein. For the recent description of bacterial *N*-acetyltransferases, see: (a) Brooke, E. W.; Davies, S. G.; Mulvaney, A. W.; Pompeo, F.; Sim, E.; Vickers, R. J. *Bioorg. Med. Chem.* **2003**, *11*, 1227; (b) Vetting, M. W.; S de Carvalho, L. P.; Yu, M.; Hegde, S. S.; Magnet, S.; Roderick, S. L.; Blanchard, J. S. *Arch. Biochem. Biophys.* **2005**, *433*, 212.
20. Doping experiments with comparable amounts of diketone-containing Z-domain protein and its corresponding labeled versions with either biotin or Alexa Fluor 488 demonstrated that these proteins in either unlabeled or labeled forms have comparable ionization efficiencies under mass detection conditions. In calculating the labeling efficiency, the small peak areas from 7867 to 8040 Da were taken as the upper-limit value for unlabeled diketone-containing proteins. The integration of the corresponding peak areas from 8182 to 8357 Da for biotin-labeled proteins (Fig. 4c) and from 8540 to 8715 Da (Fig. 4d) for dye-labeled proteins in their respective MALDI-TOF spectra suggests labeling efficiencies of greater than 85% for conjugation reactions involving biotin and Alexa Fluor 488 molecules. A correction factor of 0.8 obtained by integration of their corresponding peak areas in Figure 4b has been taken into account when calculating the labeling efficiency involving Alexa Fluor 488 molecule.
21. The fluorescence image was recorded on a Molecular Dynamics Storm 860 Phosphoimager (Molecular Dynamics, Sunnyvale, CA) by chemi-fluorescence/blue-excited scans at 450 nm, both excitation and emission bandpass of 4 nm, a photomultiplier tube voltage of 750 V, and a scan rate of 1 nm/s. 0.1–0.3 nanomoles of each protein were used.
22. Carlqvist, P.; Svedendahl, M.; Branneby, C.; Hult, K.; Brinck, T.; Berglund, P. *ChemBioChem* **2005**, *6*, 331.
23. Svedendahl, M.; Hult, K.; Berglund, P. *J. Am. Chem. Soc.* **2005**, *127*, 17988.
24. For the formation of a stable enamine adduct derived from the β-diketone and the amino group of a lysine residue activated by burial in a hydrophobic pocket, see: Barbas, C. F.; Heine, A.; Zhong, G. F.; Hoffmann, T.; Gramatikova, S.; Bjornestedt, R.; List, B.; Anderson, J.; Stura, E. A.; Wilson, I. A.; Lerner, R. A. *Science* **1997**, *278*, 2085.
25. The imine and enamine adducts derived by conjugation of an aryl monoketone and the β-diketone to butyl amine, respectively, were treated with PBS buffer with constant stirring at pH values of 4.0–10.4 for up to one week and the percentage of unhydrolyzed imine and enamine adducts was then assayed using LC/MS. It was found that the enamine adduct essentially remains intact at physiological pH and above. In sharp contrast, the imine adduct is completely hydrolyzed at pH 4.0–10.4 after overnight stirring.