Site-Specific Protein Modification Using a Ketone Handle

Virginia W. Cornish,† Klaus M. Hahn,§ and Peter G. Schultz*,†

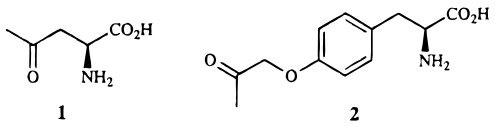
Howard Hughes Medical Institute
Department of Chemistry, University of California
Berkeley, California 94720
Department of Neuropharmacology, Division of Virology
Scripps Research Institute, La Jolla, California 92037

Received April 12, 1996

The reactive thiol side chain of the natural amino acid cysteine has been widely exploited for the selective modification of proteins with a variety of biophysical probes,[1] as a structural probe,[2] to stabilize proteins and control enzyme activity,[3] and for the generation of semisynthetic proteins.[4] The availability of a second, nonproteinogenic amino acid with such selective reactivity would be very useful in situations requiring the incorporation of two different labels into a protein, selective derivatization of unpurified protein, intracellular cross-linking or protein modification, or where a unique cysteine residue cannot be engineered into a protein. Although biosynthetic approaches involving chemically-modified or synthetic aminoacyl-tRNAs can be used to incorporate probes directly into proteins, the biosynthetic machinery places constraints on the nature of the side-chain functionality which can be incorporated.[5] Here we report a new strategy for labeling proteins site-specifically which relies on an unnatural amino acid with an electrophilic ketone side chain (Scheme 1). The unnatural keto amino acid is introduced into a unique site in a protein using the method of unnatural amino acid mutagenesis.[6] The carbonyl group of this amino acid can then be modified with a broad range of molecules containing hydrazide or alkoxyamine groups. To demonstrate this approach, T4 lysozyme containing an unnatural keto amino acid was prepared and then reacted with a fluorescein hydrazide to provide T4 lysozyme labeled site-specifically with fluorescein via a stable hydrazone linkage.

The keto group is an attractive "handle".[7] It is sufficiently stable to be incorporated into a protein in an unprotected form using a crude *Escherichia coli in vitro* extract; and yet it will react rapidly with hydrazides and alkoxyamines in aqueous solution to form hydrazones and oximes, respectively.[8] Furthermore, both hydrazones and oximes are stable under physiological conditions. The reaction of a ketone with a hydrazide is orthogonal to the functional groups present in proteins.[9] This chemistry has been used to label peptides and proteins at their *N*-termini by oxidizing an *N*-terminal serine with periodate to the corresponding aldehyde which is subsequently coupled to an appropriate hydrazide.[8a] In addition, peptide dendrimers and synthetic proteins have been constructed by coupling peptide fragments via hydrazone or oxime linkages.[8b−d]

Initially, we considered two different keto amino acids, **1** and **2**. These two amino acids differ both in the proximity of the carbonyl group to the protein backbone and in their reactivity due to the electron-withdrawing α-phenoxy group and differential steric effects. The longer homologues of **1** were not employed because they are prone to cyclic imine formation.[10] The required *N*-protected amino acid of ketone **1** was prepared from 4,5-dehydroleucine by protection of the α-amine with nitroveratryl chloroformate followed by ozonolysis of the terminal olefin. Ketone **2** was prepared similarly from *O*-(methylallyl)tyrosine, which was synthesized by treating the copper(II) complex of tyrosine with methylallyl bromide in alkaline solution.[11]



1 **2**

The efficiency with which **1** and **2** could be incorporated into a protein by the biosynthetic machinery was tested at two solvent accessible sites (Ser^{44} and Ala^{82}) in the protein T4 lysozyme (T4L) using unnatural amino acid mutagenesis.[6] An amber suppressor tRNA was charged with the keto amino acids **1** and **2** using established methods, and the aminoacylated suppressor tRNAs were combined with mutant T4L genes encoding amber codons at site 44 or site 82 in an *E. coli in vitro* transcription/translation extract.[12] Protein production levels were judged by catalytic activity and by SDS-PAGE and autoradiography of the soluble fractions of transcription/translation reactions charged with ^{35}S-methionine.[13] At both sites 44 and 82 ketone **1** is incorporated with only 5% efficiency (consistent with the size and polarity of the side chain),[5a] while ketone **2** is incorporated with 30% efficiency.[14] Further tests established that ketone **2** is incorporated with 20−30% efficiency at sites in several different proteins, including aspartate transcarbamylase, keto-

* To whom correspondence should be addressed.
† University of California.
§ Scripps Research Institute.

(1) (a) Altenbach, C.; Marti, T.; Khorana, H.; Hubbell, W. *Science* **1990**, *248*, 1088−1092. (b) Brinkley, M. *Bioconj. Chem.* **1992**, *3*, 2−13. (c) Mannuzzu, L.; Moronne, M.; Isacoff, E. *Science* **1996**, *271*, 213−216. (d) Giuliano, K.; Post, P.; Hahn, K.; Taylor, D. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 405−434.

(2) (a) Creighton, T. *Methods Enzymol.* **1986**, *131*, 83−106. (b) O'Shea, E.; Rutkowski, R.; Kim, P. *Science* **1989**, *243*, 538−542. (c) Lynch, B.; Koshland, D. E., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10402−10406.

(3) (a) Sauer, R. T. *et al. Biochemistry* **1986**, *25*, 5992−5998. (b) Wells, J. A.; Powers, D. B. *J. Biol. Chem.* **1986**, *261*, 6564−6570. (c) Wetzel, R.; Perry, L. J.; Baase, W. A.; Becktel, W. J. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 401−405. (d) Matsumura, M.; Matthews, B. *Science* **1989**, *243*, 792−794. (e) Corey, D.; Schultz, P. *J. Biol. Chem.* **1989**, *264*, 3666−3669.

(4) (a) Corey, D. R.; Schultz, P. G. *Science* **1987**, *233*, 1401−1403. (b) Kaiser, E. T. *Acc. Chem. Res.* **1989**, *22*, 47−54. (c) Schnolzer, M.; Kent, S. B. H. *Science* **1992**, *256*, 221−225.

(5) (a) Cornish, V. W.; Benson, D. R.; Altenbach, C. A.; Hideg, K.; Hubbell, W. L.; Schultz, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 2910−2914. (b) For example, three glycosyl amino acids (α-*O*-rhamnosyl serine, N^{ϵ}-[β-*C*-carbonyl(*N*-acetylglucosamine)]lysine, and N^{β}-[β-*C*-carbonyl(*N*-acetylglucosamine)]-2,3-diaminopropanoic acid) were synthesized as the necessary aminoacyl-pdCpAs and tested for incorporation at sites 44 and 82 in T4L using unnatural amino acid mutagenesis. All three were incorporated with <1% efficiency: Cornish, V. W.; Michael, K.; Hoffmann, M.; Kessler, H.; Schultz, P. G. Unpublished results.

(6) (a) Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. *Science* **1989**, *244*, 182−188. (b) Cornish, V. W.; Mendel, D.; Schultz, P. G. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 621−633. (c) Mendel, D.; Cornish, V. W.; Schultz, P. G. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 435−550. (d) Bain, J. D.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R.; Diala, E. S. *J. Am. Chem. Soc.* **1989**, *111*, 8013−8014.

(7) Jencks, W. *J. Am. Chem. Soc.* **1959**, *81*, 475−481.

(8) (a) Geoghegan, K.; Smith, J. *Bioconj. Chem.* **1992**, *3*, 138−146. (b) Shao, J.; Tam, J. *J. Am. Chem. Soc.* **1995**, *117*, 3893−3899. (c) Xu, G.; Lin, B.; Tanaka, K.; Dunn, D.; Wood, D.; Gesteland, R.; White, R.; Weiss, R.; Tamanoi, F. *Cell* **1990**, *63*, 835−841. (d) Rose, K. *J. Am. Chem. Soc.* **1994**, *116*, 30−33. (e) Canne, L.; Ferre-D'Amare, A.; Burley, S.; Kent, S. *J. Am. Chem. Soc.* **1995**, *117*, 2998−3007.

(9) Imine formation with the ϵ-amino group of lysine is readily reversible, and the equilibrium lies toward free ketone and amine.

(10) Mauger, A.; Witkop, B. *Chem. Rev.* **1966**, *66*, 47−86.

(11) Bodansky, M.; Bodansky, A. *The Practice of Peptide Synthesis*; Springer-Verlag: Berlin, 1984; pp 62−63.

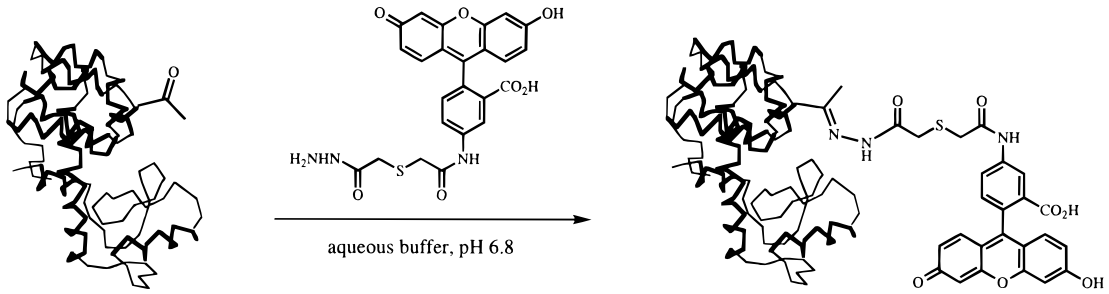
(12) (a) Robertson, S. A.; Ellman, J. A.; Schultz, P. G. *J. Am. Chem. Soc.* **1991**, *113*, 2722−2729. (b) Ellman, J. A.; Mendel, D.; Anthony-Cahill, S.; Noren, C. J.; Schultz, P. G. *Methods Enzymol.* **1992**, *202*, 301−336.

(13) Ellman, J. A.; Mendel, D.; Schultz, P. G. *Science* **1992**, *255*, 197−200.

(14) The incorporation efficiency is defined as the ratio of the amount of T4L produced by suppressing a gene containing an amber codon *in vitro* with a given aminoacyl-$tRNA_{CUA}$ relative to the amount produced from the wt gene *in vitro*. As a control, the amount of T4L produced when the aminoacyl-tRNA is replaced by an unacylated tRNA is shown to be ≤1%.

S0002-7863(96)01216-4 CCC: $12.00

Scheme 1



steroid isomerase, and thymidylate synthase.[15] Based on these results, ketone **2** was chosen for use in the subsequent chemical derivatization experiments.

In order to determine the efficiency with which proteins containing ketone **2** can be modified, a T4L mutant containing **2** at site 82 was isolated from a large-scale *in vitro* reaction and labeled with a commercially available fluorescein hydrazide (Scheme 1). As a control, wild-type (wt) T4L was isolated from a large-scale *in vitro* reaction and taken through identical labeling conditions.[16] T4L Ala^{82} → **2** and wt T4L were prepared from 10 mL and 3 mL *in vitro* reactions, respectively, and partially purified by ion-exchange chromatography using tandem DEAE-CM cartridges to yield ca. 10 μg of each protein. Portions of these crude protein solutions were exchanged into 100 mM potassium phosphate, pH 6.8, 0.5 M sodium chloride and concentrated. Fluorescein hydrazide was added to the protein solutions such that the labeling reactions contained 1 μM T4L, 1 mM fluorescein hydrazide, 20% dimethyl sulfoxide, and 80 mM potassium phosphate, pH 6.8, 0.4 M sodium chloride. The mixture was allowed to react for 36 h at room temperature and then was desalted and purified to homogeneity by ion-exchange chromatography. A comparison of the fluorescence spectra of labeled T4L Ala^{82} → **2** and "labeled" wt T4L shows that only the protein containing the ketone at site 82 has been labeled with fluorescein (Figure 1). Based on a comparison to the fluorescence emission spectrum of T4L labeled nonspecifically with the NHS ester of fluorescein, the extent of labeling is judged to be approximately 50%.[17] This strategy should facilitate the versatile alteration of protein structure, including controlled post-translational modification with polysaccharides and terpenes and the introduction of biophysical probes. Currently the keto "handle" is being used for the site-specific introduction of novel fluorophores[18] into proteins to produce biosensors which can report specific protein activities within individual, living cells.[1d]

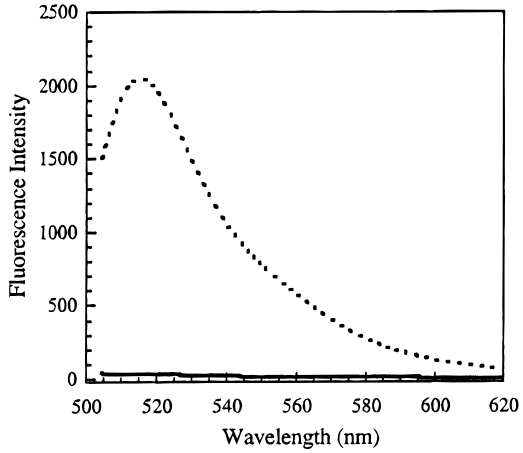


Figure 1. Fluorescence emission spectra (uncorrected) of wild-type T4 lysozyme (solid line) and of the T4 lysozyme mutant Ala^{82} → ketone **2** (dotted line) that have been treated with fluorescein hydrazide; excitation at 490 nm.

Acknowledgment. We are grateful for financial support for this work from the National Institutes of Health, grants AI34929 (K.M.H.), AI09484 (K.M.H.), and RO1GM49220 (P.G.S.). P.G.S. is a Howard Hughes Medical Institute (HHMI) Investigator. V.W.C. was supported by predoctoral fellowships from the National Science Foundation, the American Chemical Society, Division of Organic Chemistry, and HHMI. This is publication 9986-NP, Division of Virology, Department of Neuropharmacology, Scripps Clinic and Research Foundation.

Supporting Information Available: Experimental procedures for the syntheses of ketones **1** and **2**, for protein purification, for the labeling reactions, and for the fluorescence experiments (7 pages). See any current masthead page for ordering and Internet access instructions.

JA961216X

(15) Barrett, J. E.; Dayem, L.; Froland, W.; Cornish, V. W.; Schultz, P. G. Unpublished results.

(16) For both the T4L mutant Ala^{82} → ketone **2** and wt T4L, the protein purification, labeling, and fluorescence experiments were also done at pH 4.5, giving the same results.

(17) *In vivo* produced T4L was labeled non-specifically with the NHS ester of fluorescein. The extent of labeling was judged to be approximately 1:1 dye to protein based on the ratio of A_{490} to A_{280}, and the concentration of fluorescein was determined from the A_{490}. A caveat to this quantification method is that the fluorescence of fluorescein is influenced by the exact protein environment.

(18) (a) Hahn, K.; Waggoner, A.; Taylor, D. *J. Biol. Chem.* **1990**, *265*, 20335−20345. (b) Hahn, K.; DeBiasio, R.; Taylor, D. *Nature* **1992**, *359*, 736−739.