Supplementary Information

A method to site-specifically introduce methyllysine into proteins in *E. coli*

Hui-wang Ai,[†] Jae Wook Lee[†] and Peter G. Schultz*

Department of Chemistry, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037, USA

1. Synthesis of Unnatural Amino Acids

General method

All chemicals were purchased from Sigma-Aldrich (St. Louis, MO) or Acros Organic (Morris Plains, NJ), and used without further purification. ¹H-NMR spectra were recorded on a Bruker 500 MHz instrument with chemical shifts recorded relative to tetramethylsilane. Mass spectra were acquired with an ESI-TOF (high accuracy) system at the Scripps Center for Mass Spectrometry (La Jolla, CA).



Scheme I

2,2-dimethyl-4,12-dioxo-3,13-dioxa-5,11-diazahexadec-15-ene-6-carboxylic acid S2: Allyl chloroformate (171 uL, 1.62 mmol) was added to a solution of Boc-Lys-OH (**S1**, 500 mg, 2.03 mmol) in 1 N NaOH (5 mL). The solution was stirred at RT overnight. The reaction mixture was diluted with H₂O (10 mL), and washed with diethyl ether (3×20 mL). The aqueous layer was isolated with a separatory funnel and acidified with cold 1 N HCl to pH=2, and then extracted with ethyl acetate (3×30 mL). The organic layers were combined and dried over anhydrous Na₂SO₄ and evaporated *in vacuo* to give the oil **S2** (670 mg, quantitative). The compound **S2** was used without further purification in the next step.

S2: ¹H-NMR (500MHz, CDCl₃): δ =1.41-1.49 (m, 9H), 1.50-1.55 (m, 2H), 1.65-1.75 (m, 2H), 1.82-1.90 (m, 2H), 3.18 (m, 2H), 3.73, 4.56 (d, J=2.4Hz, 2H), 4.6 (m, 1H), 5.22 (d, J=10Hz, 1H), 5.30 (d, J=17.5 Hz, 1H), 5.92 (m, 1H). ESI-TOF (high-acc) calcd for C₁₅H₂₆N₂O₆ (M+Na):353.1683, obsd: 353.1684.

 N^{e} -allyloxycarbonyl-L-lysine 1: CH₂Cl₂ (4 mL) and trifluoroacetic acid (4 mL) were added to a stirred suspension of compound S2 (670 mg, 2.03 mmol). The reaction mixture was stirred at RT for 1 hour. Solvent was evaporated *in vacuo*. Cold diethyl ether (50 mL) was slowly added to the remaining oil to yield a solid. The solid was collected by vacuum filtration and washed with cold diethyl ether (10 mL). The product was further purified by dissolution in EtOH (2 mL) and precipitation in cold diethyl ether (15 mL). The resulting solid was collected by vacuum filtration and washed with diethyl ether (10 mL) to give pure white solid 1 (332 mg, 71 % over all steps).

Compound 1 ¹H-NMR (500MHz, D₂O): δ =1.36-1.45 (m, 2H), 1.45-1.62 (m, 2H), 1.8-2.0 (m, 2H), 3.15 (t, J=6.5 Hz, 2H), 3.73 (t, J=6.5 Hz, 1H), 4.56 (d, J=3 Hz, 2H) allylic H, 5.25 (d, J=10.5 Hz, 1H), 5.32 (d, J=17 Hz, 1H), 5.9 (m, 1H). ESI-TOF (high-acc) calcd for C₁₀H₁₈N₂O₄ (M+1): 231.1339, obsd: 231.1342.



Scheme II

 N^{e} -allyloxycarbonyl- N^{e} -methyl-L-lysine 2: 2 was prepared according to modified literature procedures.¹ Basic copper carbonate (376 mg) was added to a solution of H-Lys (Me)-OH (**S3**, 500 mg, 2.0 mmol) in H₂O (3.6 mL). The reaction mixture was heated to reflux for 30 min and the hot solution was filtered through celite. The filter pad was washed with water. The filtrate was made basic by addition of NaHCO₃ (430 mg), followed by addition of allyl chlorofomate (369 mg, 3.0 mmol). The reaction mixture was allowed to stir at RT overnight. 8-Hydroxyquinoline (739 mg, 5.08 mmol) in CH₂Cl₂ (25 mL) was added to the mixture. After stirring for 3 hours, the reaction mixture was filtered through a frit to remove precipitate. The aqueous layer was first washed with CHCl₃ (3× 30 mL) and then neutralized with 0.5 N HCl. Water was removed by lyophilization and the resulting solid was dissolved in MeOH/CH₂Cl₂ (1:4, 50 mL). Insoluble salt was removed by filtration and the filtrate was evaporated. The crude compound was purified by HPLC (water/acetonitrile: 0% to 70 % acetonitrile) to give 2 (533 mg, 70 % over all steps).

¹H-NMR (500 MHz, D₂O): δ =1.30-1.45 (m, 2H), 1.45-1.65 (m, 2H), 1.65-1.90 (m, 2H), 3.17 (m, 2H), 3.45 (s, 3H), 3.6 (m, 1H), 4.60 (d, J=3 Hz, 2H) allylic H, 5.22 (d, J=10.5 Hz, 1H), 5.32 (d, J=17 Hz, 1H), 5.9 (m, 1H). ESI-TOF (high-acc) calcd for C₁₁H₂₀N₂O₄ (M+1): 245.1496, obsd: 245.1496.

2. Protein expression in E. coli.

The plasmid pMyo-Lys99TAG has been described previously.² Wild-type *M. barkeri* PylRS was inserted between the Nde1 and Pst1 sites of the pBK vector.² pBK-*Mb*PylRS-Tyr384Phe was then created by QuikChange mutagenesis (Stratagene, La Jolla, CA). Both pMyo and pBK plasmids were used to transform DH10B *E. coli* cells. Single colonies were used to inoculate 25 mL 2YT media containing 0.2% L-arabinose, 20 μ g/ml tetracycline and 50 μ g/ml kanamycin in the presence or absence of 1 mM UAAs at 30 °C for 14 hours. His-tagged myoglobin proteins were purified with Ni-NTA columns (Qiagen, Germantown, MD) under native conditions by following manufacturer's instructions.



Scheme III Vector map of pDule-mbPyl

The plasmid pDule-PylRS_{Y384F} encoding PylRS-Tyr384Phe and its cognate tRNA, was constructed by replacing the myoglobin gene in pMyo with the PyIRS-Tyr384Phe gene under control of a GlnRS promoter. pDule-PylRS_{Y384F} and the plasmid pET-22b harboring histone H2B-Lys27TAG gene were used to cotransform E. coli BL21(DE3) cells. A single colony was used to inoculate 5 mL overnight 2YT culture, which was then diluted into 100 mL fresh 2YT (containing 13 µg/ml tetracycline, 100 µg/ml ampicillin, 0.02% L-arabinose and 1 mM UAA). When OD reached 0.7-0.9, 1 mM IPTG was used to induce protein expression at 30 °C for 14 hours. His-tagged proteins were purified from inclusion bodies and then refolded by dialysis. Briefly, after cells were disrupted by sonication, centrifugation was carried out to separate inclusion bodies. The inclusion bodies were then solubilized in phosphate buffered saline (PBS, pH 7.8) containing 8 M guanidinium chloride. His-tagged myoglobin proteins were purified with Ni-NTA columns (Qiagen, Germantown, MD) under denaturing conditions by following manufacturer's instructions. The resulting protein was dialyzed using a Slide-A-Lyzer Cassette (3.5k MWCO, Pierce, Rockford, IL) against a buffer containing 20 mM sodium phosphate (pH 7.8) and 300 mM NaCl.

3. Allylcarbamoyl elimination

Proteins were prepared at 0.5 mg/ml in PBS. Chloro-pentamethylcyclopentadienylcyclooctadiene-ruthenium(II) (Sigma, St. Louis, MO) and thiophenol (Sigma, St. Louis, MO) were dissolved in DMSO to 20 mM and 200 mM, respectively. 86 μ l protein solution was then mixed with 4 μ l of [Cp*Ru(cod)Cl] stock solution and 10 μ l of thiophenol stock solution. Reaction mixtures were incubated at 37 °C or room temperature for 3 hours, and then directly analyzed by ESI-MS, or by western blot.

4. Protein mass spectrometry and western blot analysis

Proteins in PBS or proteins after allylcarbamoyl elimination were analyzed with an Agilent 1100 Series LC/MS instrument (quadrupole MSD equipped with a short C-8 column to remove salt impurities). Observed spectra were deconvoluted to derive masses of intact proteins using the Agilent LC/MSD Deconvolution package provided along with the instrument. This ESI-MS system detects protein masses within the expected $\pm 0.02\%$ mass error.

For western blots, PVDF membranes with blotted proteins were first blocked with 1% BSA for 1 hour, and incubated with anti-methyllysine antibody (Abcam, Cambridge, MA) in 1/500 dilution at 4 °C for 14 hours. HRP-conjugated anti-rabbit secondary antibody (Sigma, St. Louis, MO) and ECL western blotting substrate (Pierce, Rockford, IL) were used to visualize the hybridization using a Gel-Doc system.

- 1. D. P. Nguyen, M. M. Garcia Alai, P. B. Kapadnis, H. Neumann and J. W. Chin, *J. Am. Chem. Soc.*, 2009, **131**, 14194-14195.
- 2. P. Chen, D. Groff, J. Guo, W. Ou, S. Cellitti, B. Geierstanger and P. Schultz, *Angew. Chem. Intl. Ed.*, 2009, **48**, 4052-4055.



Fig. S1 (a) ESI-MS of myoglobin expressed in the presence of 1 mM 1. (b) ESI-MS of myoglobin in panel **a** following treatment with the ruthenium catalyst. (c) ESI-MS of myoglobin expressed in the presence of 1 mM 2. (d) ESI-MS of myoglobin in panel **c** following treatment with the ruthenium catalyst. Inserts show the deconvoluted spectra.