DOI: 10.1002/cbic.200700605

Enzymatic N-terminal Addition of Noncanonical Amino Acids to Peptides and Proteins

Rebecca E. Connor,[a] Konstantin Piatkov,[b] Alexander Varshavsky,[b] and David A. Tirrell*[a]

The preparation of well-defined protein conjugates is essential for many therapeutic and biochemical technologies. The amino terminus is an especially attractive target for conjugation, and selective N-terminal modification has been achieved by a variety of chemical and enzymatic methods.[1–4] Recently, the *Escherichia coli* leucyl, phenylalanyl-transferase (Aat) has been used to modify proteins with reactive phenylalanine analogues delivered from chemically synthesized aminoacyl–tRNAs.[5] We present here an analytical method for identification of amino acid substrates of Aat and also a fully enzymatic method for N-terminal addition of phenylalanine, leucine, and methionine analogues to peptides and proteins.

The *E. coli* leucyl, phenylalanyl-transferase (Aat), encoded by the *aat* gene, catalyzes the conjugation of leucine (Leu), phenylalanine (Phe), or methionine (Met) from an aminoacylated tRNA to any protein that bears N-terminal arginine or lysine.[6] The transfer of Leu or Phe to the N terminus of a protein in wild-type *E. coli* cells results in a decrease of the protein's in vivo half-life through degradation by ClpAP, an ATP-dependent protease.[7] Aat, ClpAP, and the adapter protein, ClpS, comprise the *E. coli* N-end rule pathway.[8] Aat is tolerant of structural variation in its amino acid substrates and is known to accept Phe analogues through chemically aminoacylated-tRNAs.[5,9]

We used a simple chromatographic assay to demonstrate Aat-mediated transfer of noncanonical amino acids **1**–**10** (Scheme 1) to the acceptor peptide lysylalanyl-7-amino-4-methylcoumarin (KA-AMC, **11**). Representative chromatograms are shown in Figure 1; for the complete set of chromatograms and mass spectra see the Supporting Information.

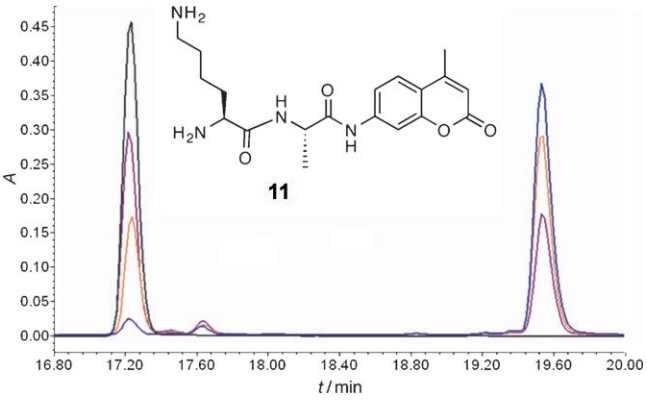
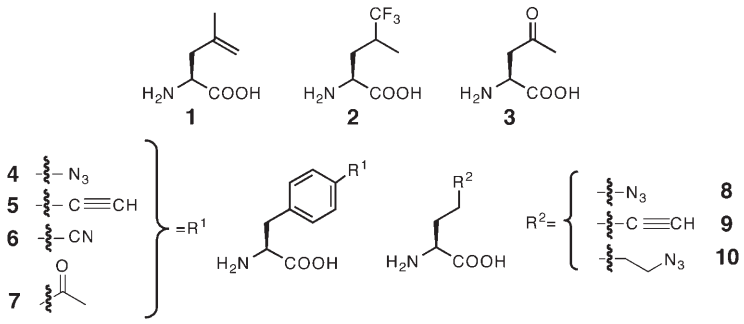


Figure 1. Reversed-phase HPLC analysis of Aat-mediated transfer of **5** to the N terminus of KA-AMC (**11**, 17.2 min retention time) with detection at 324 nm. The chromatograms show the increasing intensity of the signal (at 19.5 min) due to the tripeptide product after 1 h (magenta), 2 h (orange), and 3 h (blue). Reactions were performed with 1 mM **5**, 100 μM KA-AMC, 750 nM Aat, and 400 nM synthetase.



Scheme 1. Noncanonical amino acid substrates for Aat.

The required aminoacyl–tRNA substrates were prepared in situ by treatment of the amino acids of interest with their cognate *E. coli* aminoacyl–tRNA synthetases. *p*-Azidophenylalanine (**4**), *p*-ethynylphenylalanine (**5**), and *p*-cyanophenylalanine (**6**) were treated with the A294G mutant of the phenylalanyl–tRNA synthetase (PheRS).[10] A computationally designed A294G/T251G PheRS mutant was used to generate *p*-acetylphenylalanyl–tRNA.[11] 4-Dehydroleucine (**1**) and 5,5,5-trifluoroleucine (**2**) were charged to tRNA by using the wild-type leucyl–tRNA synthetase (LeuRS), while a LeuRS variant that bore a T252Y mutation in the editing domain was used to charge oxonorvaline (**3**).[12] The Met surrogates azidohomoalanine[13] (**8**) and homopropargylglycine[14] (**9**) were charged by the wild-type methionyl–tRNA synthetase (MetRS); azidonorleucine (**10**) required a single (L13G) mutation in MetRS for synthesis of the aminoacyl–tRNA.[15]

All ten noncanonical amino acids were transferred to the N terminus of KA-AMC; typical yields of the tripeptide products are listed in Table 1. Under the conditions used here, analogues **1**, **2**, **4**, and **5** were transferred to the N terminus of **11** in near-quantitative yields within four hours. The remaining analogues were transferred in lower yields. We also examined azaleucine and the phenylalanine analogues 2-pyridylalanine and 3-pyridylalanine, all of which were charged to their respective

[a] *R. E. Connor, Prof. D. A. Tirrell*
Division of Chemistry and Chemical Engineering
California Institute of Technology
1200 E. California Boulevard, Pasadena, CA 91125 (USA)
Fax: (+1) 626-568-8824
E-mail: tirrell@caltech.edu

[b] *K. Piatkov, Prof. A. Varshavsky*
Division of Biology, California Institute of Technology
1200 E. California Boulevard, Pasadena, CA 91125 (USA)

Supporting information for this article is available on the WWW under http://www.chembiochem.org or from the author.

Table 1. Formation of tripeptide products by N-terminal addition of noncanonical amino acids to KA-AMC.

Amino acid	Synthetase	Yield [%][a]
Leu	wt LeuRS	99 ± 2
1	wt LeuRS	97 ± 1
2	wt LeuRS	≥ 99
3	T252Y LeuRS	≤ 10[b]
4	A294G PheRS	92 ± 12
5	A294G PheRS	96 ± 3
6	A294G PheRS	90 ± 17
7	A294G/T251G PheRS	21 ± 1
8	wt MetRS	73 ± 2
9	wt MetRS	84 ± 2
10	L13G MetRS	18 ± 4

[a] Yield after 4 h at 37 °C as determined from chromatographic separation. Uncertainty is standard deviation from two experiments. [b] Yields ranged from 1 to 10% over five experiments. wt: wild type.

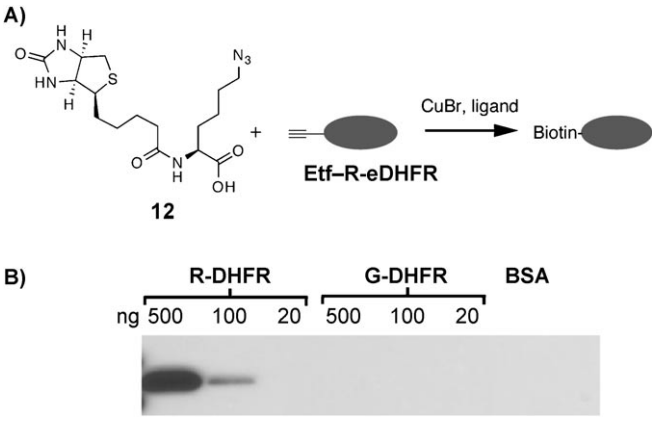


Figure 2. N-terminal modification of e-DHFR. A) After N-terminal addition of Etf, the alkyne moiety was conjugated to azide–biotin probe (**12**). B) Immunoblot analysis of transfer reactions containing decreasing amounts of R-eDHFR and G-eDHFR; BSA: bovine serum albumin negative control.

tRNAs;[10,16] however, we found no evidence in the in vitro assay that any of these analogues can be transferred by Aat to the N terminus of KA-AMC.

To produce full-length protein substrates of Aat, a construct that allows the generation of proteins with N-terminal residues other than Met was required. Such a system, based on the ubiquitin (Ub) fusion technique, was developed for the elucidation of the N-end rule pathway in *E. coli*. In this method, a Ub–X–protein fusion was expressed in *E. coli* and the junctional residue X (any desired residue except Pro) was made N-terminal through the removal of the Ub moiety by a coexpressed eukaryotic deubiquitylating enzyme.[7] A recent modification of this technique, in which the fusion protein was purified from *E. coli* and the Ub moiety was then removed in vitro with a purified deubiquitylating enzyme,[17] yielded an *E. coli* dihydrofolate reductase (eDHFR) that bears either N-terminal arginine (R-eDHFR) or glycine (G-eDHFR).

Analogue **5** (Etf) was appended to the N terminus of R-eDHFR to form Etf–R-DHFR and subsequently ligated to azide–biotin probe **12** via copper(I)-catalyzed cycloaddition.[18] As little as 100 ng of biotinylated product could be detected by Western blotting with streptavidin–HRP (Figure 2B). Under identical conditions G-eDHFR was unmodified; this illustrates the specificity of Aat. Edman analysis of the Aat-mediated transfer of **5** showed that R-eDHFR was modified nearly quantitatively after 3 h; this is consistent with the results of the peptide assay (Supporting Information). Ligation of an azide–PEG–fluorescein (APF) conjugate to Etf–R-eDHFR was also accomplished in good yield via copper(I)-catalyzed cycloaddition (Figure 3). Fluorescence detection at 530 nm confirmed that ligation of APF required prior treatment of the protein with Etf, PheRS, and Aat (i.e., a complete transferase reaction mixture). The two-stage modification was achieved in an overall yield of 80 ± 2%, as determined by densitometric analysis of Western blots (Figure 3B).

In conclusion, we have used coupled systems of aminoacyl–tRNA synthetases and Aat to modify the N termini of peptides and proteins with a variety of noncanonical amino acids. Bioorthogonal reactive functional groups, such as alkenes, alkynes, azides, and ketones, can be transferred and used to prepare bioconjugates in high yield. The method introduced here allows new approaches to the engineering of therapeutic proteins through pegylation; to the study of protein interactions through crosslinking; and to the immobilization of proteins for use in sensors, microarrays, and catalytic systems. The HPLC assay for monitoring Aat activity can be applied to any noncanonical amino acid.

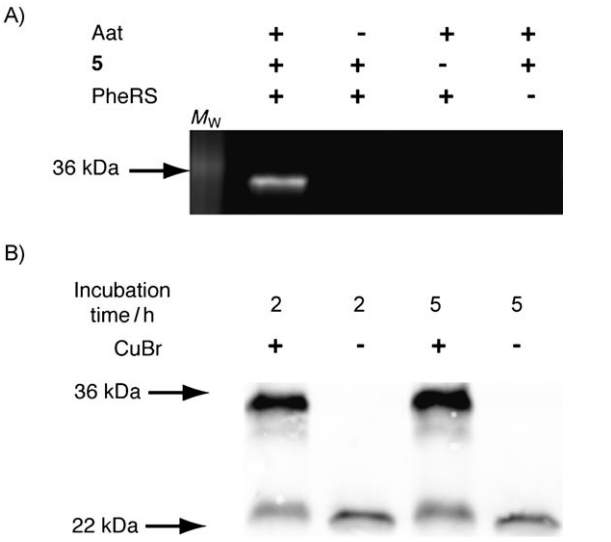


Figure 3. PEGylation of Etf–eDHFR with azide–PEG-5000–fluorescein (APF). A) Fluorescence image (detection at 530 nm) of gel separation, showing products of complete and negative-control transfer reactions after treatment with APF. The pegylated protein migrates with an apparent molecular mass of about 36 kDa. B) Western blot analysis of eDHFR–hemagglutinin (HA) subsequent to transfer of Etf and treatment with APF (with or without CuBr, incubated for 2 or 5 h at 4 °C). DHFR–HA bears a C-terminal HA tag, and was detected with Cy5–anti-mouse immunoglobulin and mouse anti-hemagglutinin. The pegylated product migrates with an apparent molecular mass of about 36 kDa; unmodified eDHFR–HA with an apparent molecular mass of about 22 kDa.

Experimental Section

Materials: Chemical reagents were purchased from Sigma–Aldrich (St. Louis, MO, USA) and used without further purification, unless otherwise noted. The substrate dipeptide, Lys-Ala-4-aminomethylcoumarin, and amino acids **4** and **6** were purchased from Bachem (Bubendorf, Switzerland) and used as received. PFP–biotin was purchased from Pierce (Rockford, IL, USA) and mSPA–PEG-5000–fluorescein was obtained from Nektar (San Carlos, CA, USA). Amino acid **2** was purchased from Fluorochem (Derbyshire, UK); **1**, **3**, and **9** were prepared by alkylation of diethyl acetamidomalonate with the appropriate tosylate as described previously.[14] Compounds **8** and **10** were synthesized by diazo transfer to the equivalent amine-bearing amino acids;[19] **5** and **7** were prepared as previously reported.[10,11] The *E. coli* phenylalanyl–, leucyl–, and methionyl–tRNA synthetases were produced as described.[10,20] The leucyl, phenylalanyl transferase was purified as previously described.[8a]

Reaction conditions for chromatographic analysis: Each 50 μL reaction mixture contained amino acid (1 mM), cognate synthetase (400 nM), total *E. coli* tRNA (30 μg, 11.5 pmol), Aat (800 nM), and KA-AMC (100 μM) in the transferase reaction buffer (50 mM Tris, pH 8.0; 50 mM β-mercaptoethanol; 5 mM ATP; 10 mM $MgCl_2$; 150 mM KCl). The reaction mixtures were incubated at 37 °C and stopped either by filtration through a Microcon YM-10 filter or by precipitation of the proteins with acetone. After centrifugation through the YM-10 filter and 2–3 washes with water (100 μL), the filtrate was separated by HPLC and used for subsequent MALDI analysis. If precipitation was used to remove the added synthetase and transferase, 4 volumes of acetone were added and the reaction mixture was incubated at −20 °C for at least 1 h. The resulting precipitate was collected by centrifugation and the supernatant, which contained both the substrate and product, was transferred to a new tube, concentrated, and analyzed by HPLC.

HPLC separation of tripeptide products from KA-AMC: All liquid chromatography, except that on the *p*-azidophenylalanine reaction mixture, was performed by using a Waters HPLC system with a Microsorb C18 column (Varian, Inc). The buffers used for separation of the tripeptide products were trifluoroacetic acid (0.1%, eluent A) and acetonitrile (100%, eluent B). The gradient consisted of 0–5 min, 0% B; 5–10 min, 0–30% B; 10–25 min, 30–60% B; 25–30 min, 60–100% B; 30–40 min, 100% B. Dual detection at 214 and 324 nm was used to identify the substrate; product peaks and fractions containing the product were collected for analysis by electrospray ionization mass spectrometry. For oxonorvaline, a modified gradient was required to separate the substrate from the product tripeptide. The gradient was 0–5 min, 0% B; 5–10 min, 0–24% B; 10–25 min, 24–48% B; 25–40 min, 48–100% B; 40–45 min, 100% B. The *p*-azidophenylalanine reaction mixture was separated by using a Varian HPLC system with a Microsorb C18 column. The buffers for separation were trifluoroacetic acid (0.1%) and acetonitrile (80%), trifluoroacetic acid (0.06%). The gradient consisted of 0–5 min, 0% B; 5–7 min, 0–30% B; 7–27 min, 30–100% B; 27–37 min, 100% B. Detection at 324 nm was used to identify the substrate and product peaks.

N-terminal protein modification: eDHFR (7.5 μg) was modified with Etf (**5**) in a reaction volume of 75 μL in modified Aat buffer (1 mM ATP, 50 mM Tris pH 8.0, 50 mM β-mercaptoethanol, 10 mM creatine phosphate, 45 μg *E. coli* total tRNA, 20 $\mu g\,mL^{-1}$ creatine phosphokinase) with mutant PheRS (800 nM) and Aat (1.5 μg). The reaction mixture was incubated at 37 °C for 4 h, and then buffer-exchanged twice into phosphate buffered saline (PBS, pH 7.5) by using gel-filtration columns (Bio-Spin 6, Bio-Rad). For Edman degradation of modified samples, the reaction mixture was treated with 2 volumes of ice-cold acetone to precipitate the proteins. The precipitated proteins were then separated on a gel (12%) and transferred to a PVDF membrane. The protein band corresponding to Etf–R-DHFR was excised and analyzed by Edman degradation at the Caltech Peptide and Protein Molecular Analysis Laboratory. *p*-Ethynylphenylalanine-modified R-eDHFR was used directly after buffer exchange into PBS. Aliquots (10 μL) of the filtrate (0.5 μg eDHFR) were diluted into PBS (100 μL) containing CuBr (400 μM), *tris*-(benzyltriazolylmethyl)amine (TBTA, 400 μM), and either azidonorleucine–biotin (25 μM) or azide–PEG-5000–fluorescein (500 nM) for [3+2]cycloaddition. The reaction mixtures were either treated with 1:8 trichloroacetic acid/acetone (900 μL) for analysis by immunoblotting or used directly for analysis by gel electrophoresis and fluorescence imaging. Biotinylation was detected with streptavidin–HRP (R&D Systems). Fluorescence images were obtained with a Molecular Imager FX from Bio-Rad with an excitation wavelength of 488 nm for detection of fluorescein. For quantitative pegylation, the reaction mixture (25 μL) obtained from modification of eDHFR with **5** (1.25 μg Etf–R-eDHFR) was diluted into PBS (100 μL) with azide–PEG-5000–fluorescein (100 μM), CuBr (400 μM) suspended in water, and TBTA (200 μM). The reaction mixture was incubated for 5 h at 4 °C, and an aliquot was taken after 2 h. Sodium azide (1 mM) was added to quench the reaction. The reaction mixture (25 μL) was then separated directly by SDS-PAGE and transferred to a nitrocellulose membrane for Western analysis with mouse anti-hemagglutinin and Cy5–anti-mouse immunoglobulin. The Cy5 fluorescence on the blot was detected with a Molecular Imager FX and the extent of pegylation determined by densitometry by using ImageJ software.[21]

Acknowledgements

This work was supported by NIH grant GM62523 (to D.A.T.), NIH grant DK39520 (to A.V.), the ARO-sponsored Institute for Collaborative Biotechnologies, and the Beckman Institute at Caltech. We thank Christina Smolke for use of the Molecular Imager and Mona Shahgholi for help with mass spectrometry. Pin Wang and A. James Link kindly provided the purified synthetases. Edman degradation was performed by the Peptide and Protein Molecular Analysis Laboratory of the Beckman Institute at Caltech.

Keywords: amino acids · peptides · proteins · transferases

[1] D. P. Baker, E. Y. Lin, K. Lin, M. Pellegrini, R. C. Petter, L. L. Chen, R. M. Arduini, M. Brickelmaier, D. Y. Wen, D. M. Hess, L. Q. Chen, D. Grant, A. Whitty, A. Gill, D. J. Lindner, R. B. Pepinsky, *Bioconjugate Chem.* **2006**, *17*, 179–188.

[2] a) J. M. Gilmore, R. A. Scheck, A. P. Esser-Kahn, N. S. Joshi, M. B. Francis, *Angew. Chem.* **2006**, *118*, 5433–5437; *Angew. Chem. Int. Ed.* **2006**, *45*, 5307–5311; b) H. B. F. Dixon, *J. Protein Chem.* **1984**, *3*, 99–108; c) P. G. Wu, L. Brand, *Methods. Enzymol.* **1997**, *278*, 321–330.

[3] a) K. F. Geoghegan, J. G. Stroh, *Bioconjugate Chem.* **1992**, *3*, 138–146; b) H. F. Gaertner, R. E. Offord, *Bioconjugate Chem.* **1996**, *7*, 38–44; c) S. Kushnir, Y. Marsac, R. Breitling, I. Granovsky, V. Brok-Volchanskaya, R. S. Goody, C. F. W. Becker, K. Alexandrov, *Bioconjugate Chem.* **2006**, *17*, 610–617; d) X. Li, L. Zhang, S. E. Hall, J. P. Tam, *Tetrahedron Lett.* **2000**, *41*, 4069–4073.

[4] a) T. Tanaka, N. Kamiya, T. Nagamune, *FEBS Lett.* **2005**, *579*, 2092–2096; b) M. Lewinska, C. Seitz, A. Skerra, F. P. Schmidtchen, *Bioconjugate Chem.* **2004**, *15*, 231–234; c) S. Mamaev, J. Olejnik, E. K. Olejnik, K. Rothschild, *Anal. Biochem.* **2004**, *326*, 25–32; d) V. Ramachandiran, C. Willms, G. Kramer, B. Hardesty, *J. Biol. Chem.* **2000**, *275*, 1781–1786.

[5] a) M. Taki, A. Kuno, S. Matoba, Y. Kobayashi, J. Futami, H. Murakami, H. Suga, K. Taira, T. Hasegawa, M. Sisido, *ChemBioChem* **2006**, *7*, 1676–1679; b) M. Taki, M. Sisido, *Biopolymers* **2007**, *88*, 263–271.
[6] M. J. Leibowitz, R. L. Soffer, *J. Biol. Chem.* **1970**, *245*, 2066–2073.
[7] J. W. Tobias, T. E. Shrader, G. Rocap, A. Varshavsky, *Science* **1991**, *254*, 1374–1377.
[8] a) E. Graciet, R. G. Hu, K. Piatkov, J. H. Rhee, E. M. Schwarz, A. Varshavsky, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3078–3083; b) A. Erbse, R. Schmidt, T. Bornemann, J. Schneider-Mergener, A. Mogk, R. Zahn, D. A. Dougan, B. Bukau, *Nature* **2006**, *439*, 753–756.
[9] M. J. Leibowitz, R. L. Soffer, *J. Biol. Chem.* **1971**, *246*, 5207–5212.
[10] K. Kirshenbaum, I. S. Carrico, D. A. Tirrell, *ChemBioChem* **2002**, *3*, 235–237.
[11] D. Datta, P. Wang, I. S. Carrico, S. L. Mayo, D. A. Tirrell, *J. Am. Chem. Soc.* **2002**, *124*, 5652–5653.
[12] Y. Tang, PhD thesis, California Institute of Technology (USA) **2002**.
[13] K. L. Kiick, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 19–24.
[14] J. C. M. van Hest, K. L. Kiick, D. A. Tirrell, *J. Am. Chem. Soc.* **2000**, *122*, 1282–1288.
[15] A. J. Link, M. K. S. Vink, N. J. Agard, J. A. Prescher, C. R. Bertozzi, D. A. Tirrell, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10180–10185.
[16] B. Stieglitz, J. M. Calvo, *J. Bacteriol.* **1971**, *108*, 95–104.
[17] A. M. Catanzariti, T. A. Soboleva, D. A. Jans, P. G. Board, R. T. Baker, *Protein Sci.* **2004**, *13*, 1331–1339.
[18] C. W. Tornøe, C. Christensen, M. Meldal, *J. Org. Chem.* **2002**, *67*, 3057–3064.
[19] A. J. Link, M. K. S. Vink, D. A. Tirrell, *J. Am. Chem. Soc.* **2004**, *126*, 10598–10602.
[20] Y. Tang, D. A. Tirrell, *Biochemistry* **2002**, *41*, 10635–10645.
[21] W. S. Rasband, ImageJ, US National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/, **1997**–**2001**.

Received: October 10, 2007
Published online on January 18, 2008

