CHEMBIOCHEM

Supporting Information

CHEMBIOCHEM

Supporting Information

for

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

Rebecca E. Connor, Konstantin Piatkov, Alexander Varshavsky, and David A. Tirrell*

Methods

Synthesis of azidonorleucine-biotin (12): Azidonorleucine (12.65 mg, 0.073 mmol) and PFP-biotin (30.1 mg, 0.073 mmol) were dissolved in 1.5 mL of a 9:1:5 mixture of dimethylformamide, dimethylsulfoxide and methanol. Triethylamine (4 equiv) was added and the reaction stirred overnight at room temperature. The reaction was purified on a C18 Sep-Pak column (Waters) using a stepwise gradient from 0.1% trifluoroacetic acid to 100% acetonitrile. The fractions containing the desired product were lyophilized to 6.8 mg of a fluffy white powder (23.4% yield). ESI MS (*m/z*): 398.9 $[M+H]^+$ (expected *m/z*: 398.17)

Synthesis of Azido-PEG5000-Fluorescein (APF, 13): Azidopropylamine was synthesized as previously described.[1] To prepare the azide-functionalized PEG, mSPA-PEG-fluorescein (29 mg, 6 µmol) was dissolved in azidopropylamine (300 µL, 2.9 mmol) and allowed to react at room temperature for 2 h. The reaction mixture was then added dropwise to 20 mL of diethyl ether which caused the PEG reagents to form a yellow precipitate. The precipitate was collected by centrifugation and the ether supernatant decanted. The precipitate was dissolved in 1 mL of methanol and

added dropwise to ether again and the precipitate collected as before. The recovered precipitate was used without further purification. Addition of an azide was confirmed by IR analysis.

Construction of Arg-eDHFR-HA and Gly eDHFR-HA plasmids: DNA fragments encoding the gene for *E. coli* dihydrofolate reductase were amplified by PCR from *E. coli* genomic DNA with the following pairs of primers: Oligo 298 (5'-AGGCTCCGCGGTGGTcgtAAAATGATCAGTCTGATTGCGGC–3') and Oligo 314 (5'-TTTAAGCTTAGGCGTAATCTGGGACATCGTATGGGTAGCCGCTCCCCCGCCGCTCCAGAATCT–3') for Arg-eDHFR-HA; and Oligo 298 and Oligo 301 (5'-AGGCTCCGCGGTGGTggtAAAATGATCAGTCTGATTGCGGC-3') for Gly eDHFR-HA. The resulting fragments were digested with *Sac*II and *Hind*III, and ligated into *Sac*II / *Hind*III-digested pHUE[2] to generate plasmids pKP141 and pKP144, respectively.

Expression and purification of recombinant proteins: Overnight cultures in *E. coli* strain KPS17[3] were subcultured 1:100 into 200 mL Luria broth containing ampicillin and grown to a late exponential phase at 37°C. Protein expression was induced by adding isopropyl-1-thio-β-D-galactopyranoside (IPTG) to a final concentration of 0.1 mM. After 3 h, harvested cells were resuspended in 15 mL of buffer L (50 mM Na_2HPO_4/NaH_2PO_4 at pH 7.4, 500 mM NaCl, 10 mM imidazole), with 1 mM phenylmethylsulfonyl fluoride. His_6-tagged recombinant proteins were purified by nickel-affinity chromatography using batch mode under native conditions, based on the QIA*express* protocol (Qiagen). The cells were lysed by sonication (3 x 1 min bursts at 0°C) and the soluble protein fraction recovered by centrifugation at 4°C (15 min at 15 300 *g*). To the supernatant, 2 mL of a 50% slurry of nickel-nitrilotriacetic acid (Ni-NTA) agarose in buffer L was added; the mixture was then placed on a rotary wheel at 4°C for 1 h. The lysate/Ni-NTA mixture was centrifuged (5 min at ~550*g*) and the remaining Ni-NTA agarose pellet was washed 5 times in 50 mL buffer W (50 mM Na_2HPO_4/NaH_2PO_4 at pH 7.4, 500 mM NaCl, 30 mM imidazole). The His_6-tagged protein was eluted from the Ni-NTA resin in 1 mL fractions with buffer L containing 300 mM imidazole. Chosen fractions were pooled and dialyzed against buffer L, along with deubiquitylating protease Usp2-cc[6] for 12 h at 4°C. After cleavage by the protease, nickel affinity chromatography was used to separate Usp2-cc and the His_6-tagged ubiquitin domain from the modified DHFR according to published methods.[2]

[1] Carboni, B.; Benalil, A.; Vaultier, M. *J. Org. Chem.* **1993**, *58*, 3736-3741.

[2] Catanzariti, A. M.; Soboleva, T. A.; Jans, D. A.; Board, P. G.; Baker, R. T. *Prot. Sci.* **2004**, *13*, 1331-9.

[3] Graciet, E.; Hu, R.G.; Piatkov, K.; Rhee, J.H.; Schwarz, E. M.; Varshavsky, A. *Proc. Natl. Acad. Sci.* **2006**, *103*, 3078-3083.

Table S1. The percent of modified protein Etf-R-DHFR after increasing reaction length as determined by comparison of the molar amounts of arginine found by Edman degradation in analyses of the first and second residues. Arginine found as the first residue arises from unmodified protein, whereas arginine found as the second residue arises from modified protein. The reaction is largely complete after 3 h.

Incubation Time [h]	Percent modification
2	80
3	92
4	93

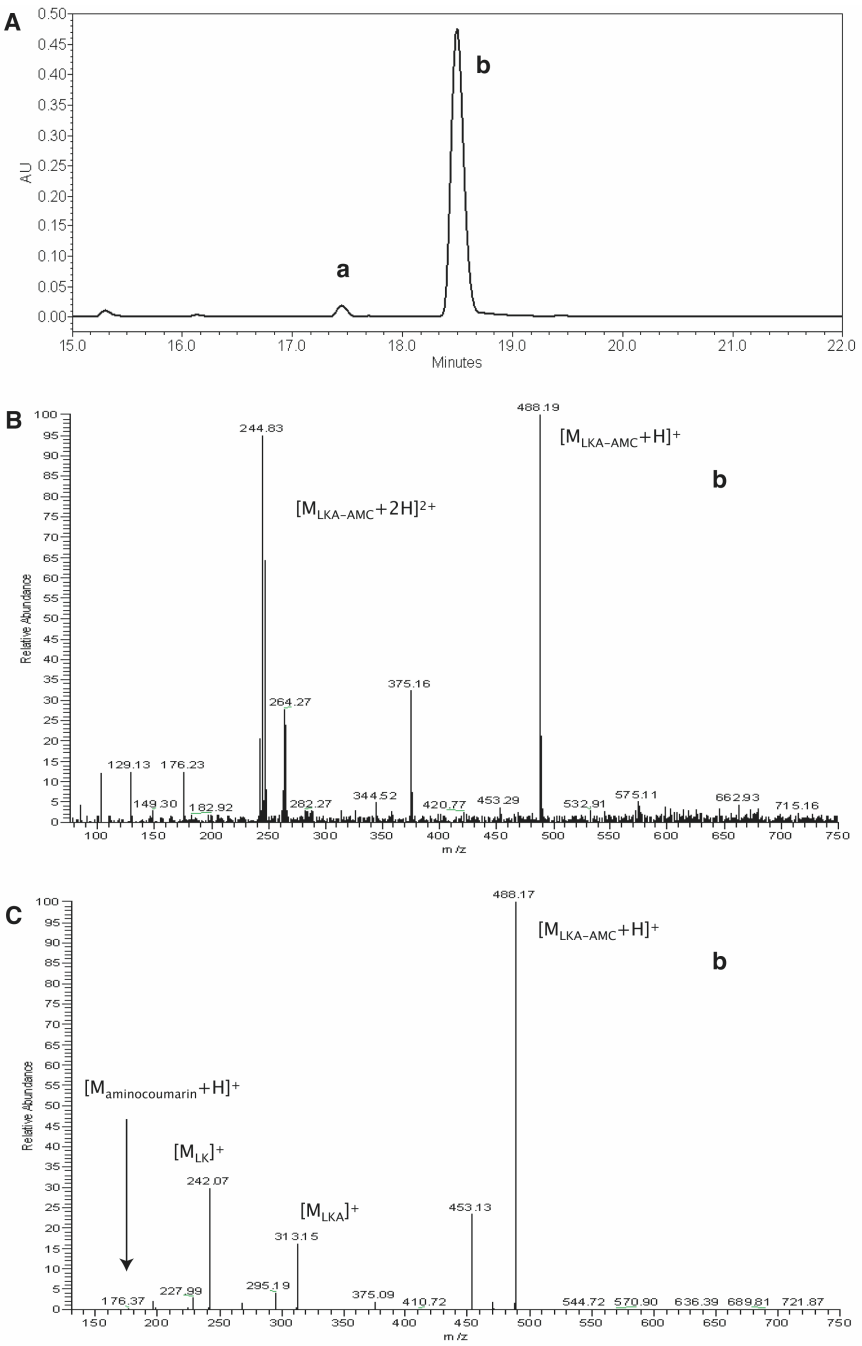

Figure S1. A) Representative reversed-phase HPLC analysis (UV detection at 324 nm) of the transfer of leucine to KA-AMC after incubation. A small amount of the starting material (a), and the tripeptide product, LKA-AMC, (b) are visible. B) The ESI mass spectrum of (b) confirms its identity as LKA-AMC, expected $[M_{LKA\text{-}AMC}+H]^+$=488.28 *m/z* and expected $[M_{LKA\text{-}AMC}+2H]^{2+}$=244.6 *m/z*. C) MS/MS analysis of the singly charged ion of the product further confirms the N-terminal addition of leucine to KA-AMC with fragmentation into dipeptide LK, expected $[M_{LK}]^+$=242.19 *m/z*, tripeptide LKA, $[M_{LKA}]^+$=313.22 *m/z*, and the aminocoumarin, $[M_{aminocoumarin}+H]^+$=176.06 *m/z*.

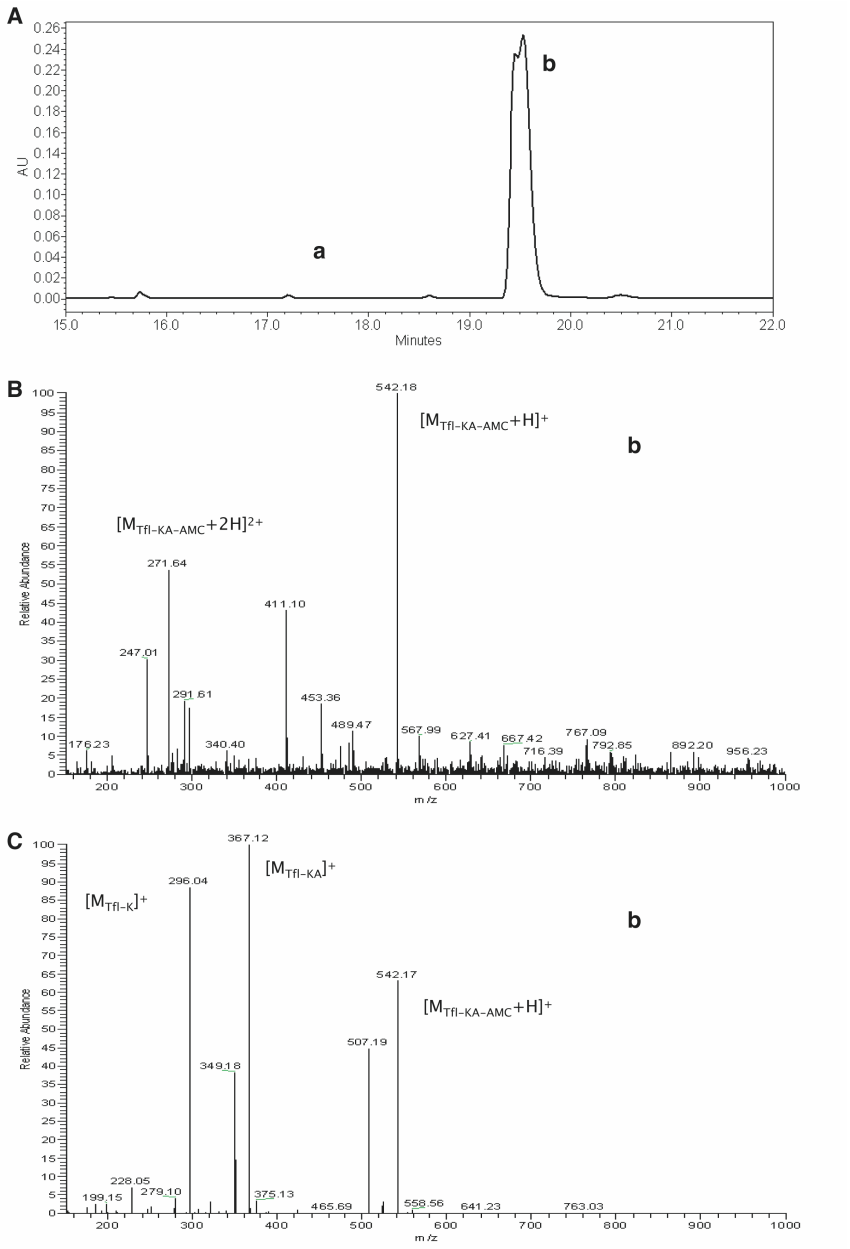

Figure S2. A) Representative reversed-phase HPLC analysis of the transfer reaction of trifluoroleucine **2** (Tfl) to KA-AMC using UV detection at 324 nm. The starting material is not detectable at 17.5 min and the isomeric product (b) appears at 19.6 minutes. B) Analysis of peak (b) by ESI mass spectrometry confirms that it contains the tripeptide product, Tfl-KA-AMC, expected $[M_{\text{Tfl-KA-AMC}}+H]^+$ =542.25 *m/z* and $[M_{\text{Tfl-KA-AMC}}+2H]^{2+}$ =271.67 *m/z.* C) MS/MS analysis of the singly charged ion of the product further confirms the N-terminal addition of Tfl to KA-AMC with fragmentation into dipeptide Tfl-K, expected $[M_{\text{Tfl-K}}]^+$=296.16 *m/z* and the tripeptide Tfl-KA, expected $[M_{\text{Tfl-KA}}]^+$=367.2 *m/z.*

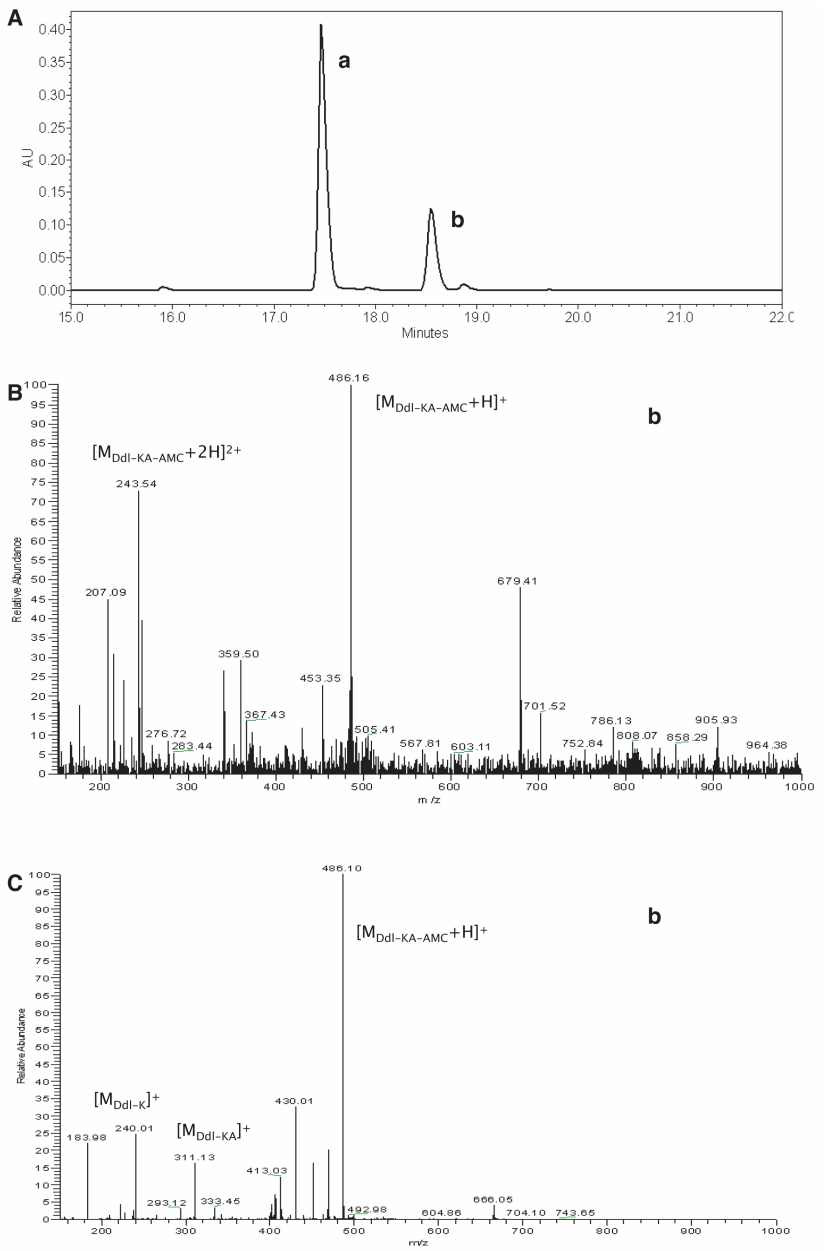

Figure S3. A) Representative reversed-phase HPLC analysis of the transfer reaction of dehydroleucine **1** (Ddl) to KA-AMC using UV detection at 324 nm. Both the starting material (a) and the tripeptide product (b) are visible. B) Analysis of peak (b) by ESI mass spectrometry confirms that it contains the tripeptide product, Ddl-KA-AMC, expecte $[M_{Ddl\text{-}KA\text{-}AMC}+H]^{+}$ = 486.26 *m/z* and expected$[M_{Ddl\text{-}KA\text{-}AMC}+2H]^{2+}$ =243.63 *m/z.* C) MS/MS analysis of the singly charged ion of the product further confirms the N-terminal addition of Ddl to KA-AMC with fragmentation into dipeptide Ddl-K, $[M_{Ddl\text{-}K}]^{+}$=240.17 *m/z* and the tripeptide Ddl-KA, $[M_{Ddl\text{-}KA}]^{+}$= 311.21 *m/z.*

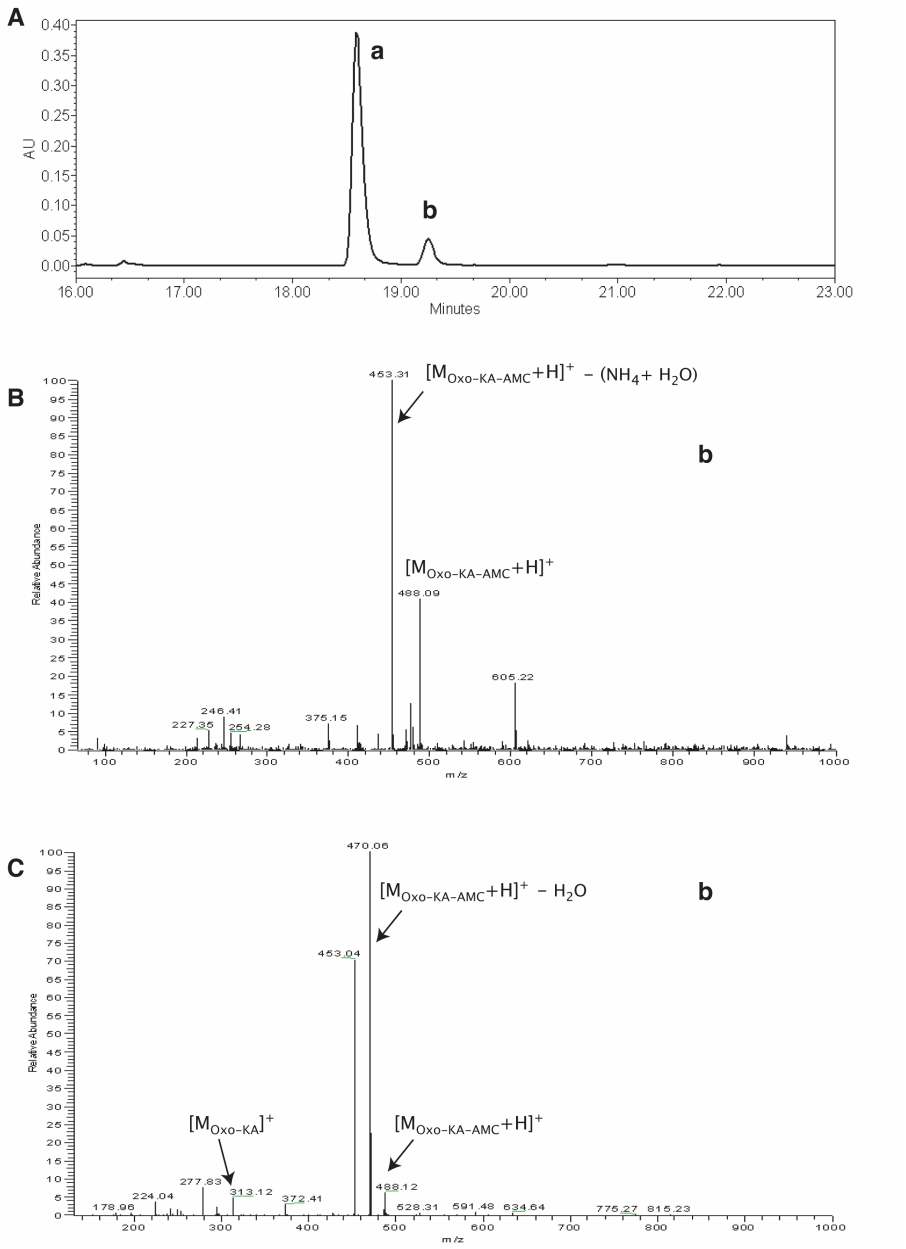

Figure S4. A) Representative reversed-phase HPLC analysis of the transfer reaction of oxo-norvaline **3** (Oxo) to the dipeptide. Both the starting material (a) and the tripeptide product (b) are visible. B) The ESI mass spectra of peak (b) confirms that it contains the tripeptide-aminocoumarin product, expected $[M_{Zxo\text{-}KA\text{-}AMC}+H]^+$ = 488.24 *m/z.* C) Fragmentation of the tri-peptide-AMC product ion results in production of the tripeptide ion, Oxo-KA, expected $[M_{Oxo\text{-}KA}]^+$=312.19 *m/z*, as well as loss of water, *m/z* = 470, and ammonium, *m/z* = 453 from the parent ion at 488.12 *m/z*.

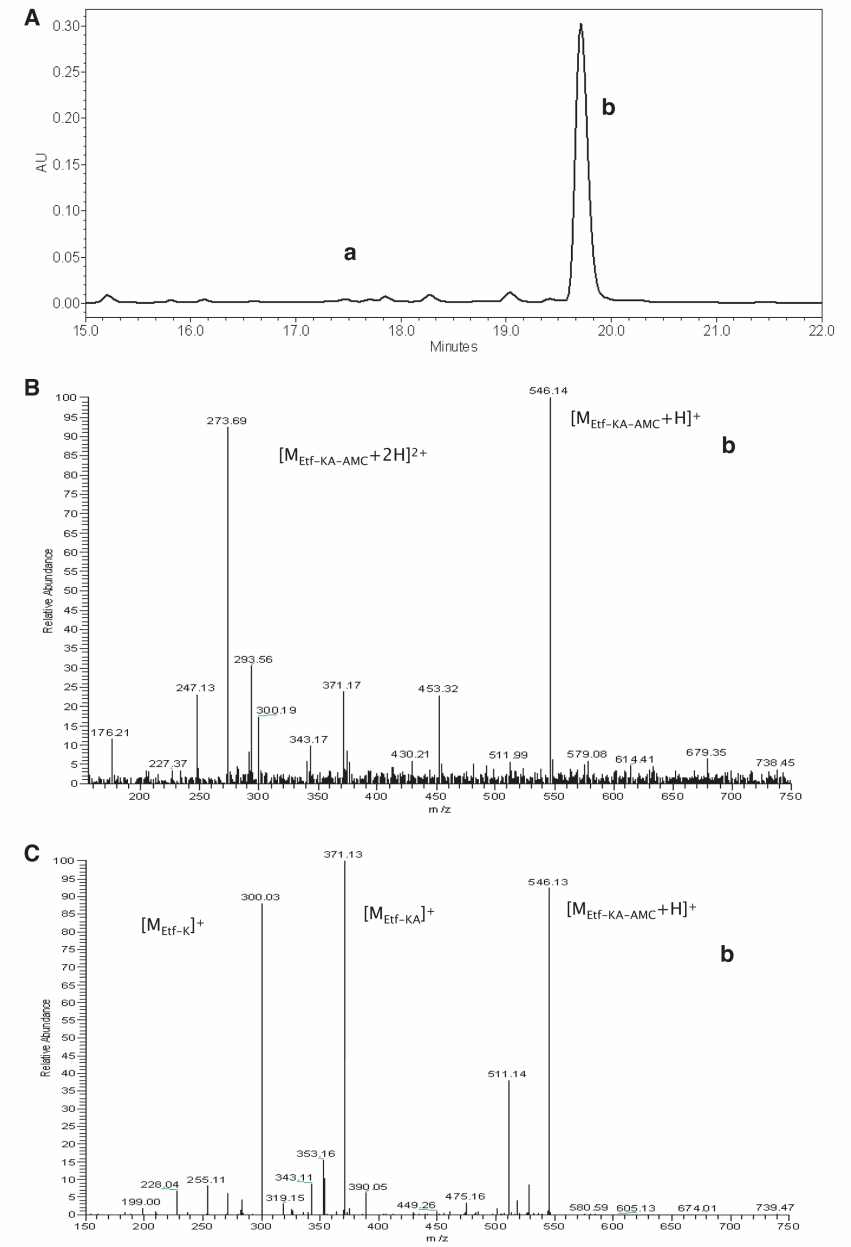

Figure S5. A) Representative reversed-phase HPLC analysis of the transfer reaction of *p*-ethynylphenylalanine **5** (Etf) to KA-AMC using UV detection at 324 nm. Only the tripeptide product (b) are visible. B) Analysis of peak (b) by ESI mass spectrometry confirms that it contains the tripeptide product, Etf-KA-AMC, expected $[M_{\text{Etf-KA-AMC}}+H]^+$ =546.26 *m/z* and $[M_{\text{Etf-KA-AMC}}+2H]^{2+}$ = 273.63 *m/z.* C) MS/MS analysis of the singly charged ion of the product further confirms the N-terminal addition of Etf to KA-AMC with fragmentation into dipeptide Etf-K, expected $[M_{\text{Etf-K}}]^+$=300.17 *m/z* and the tripeptide Etf-KA, $[M_{\text{Etf-KA}}]^+$= 371.45 *m/z.*

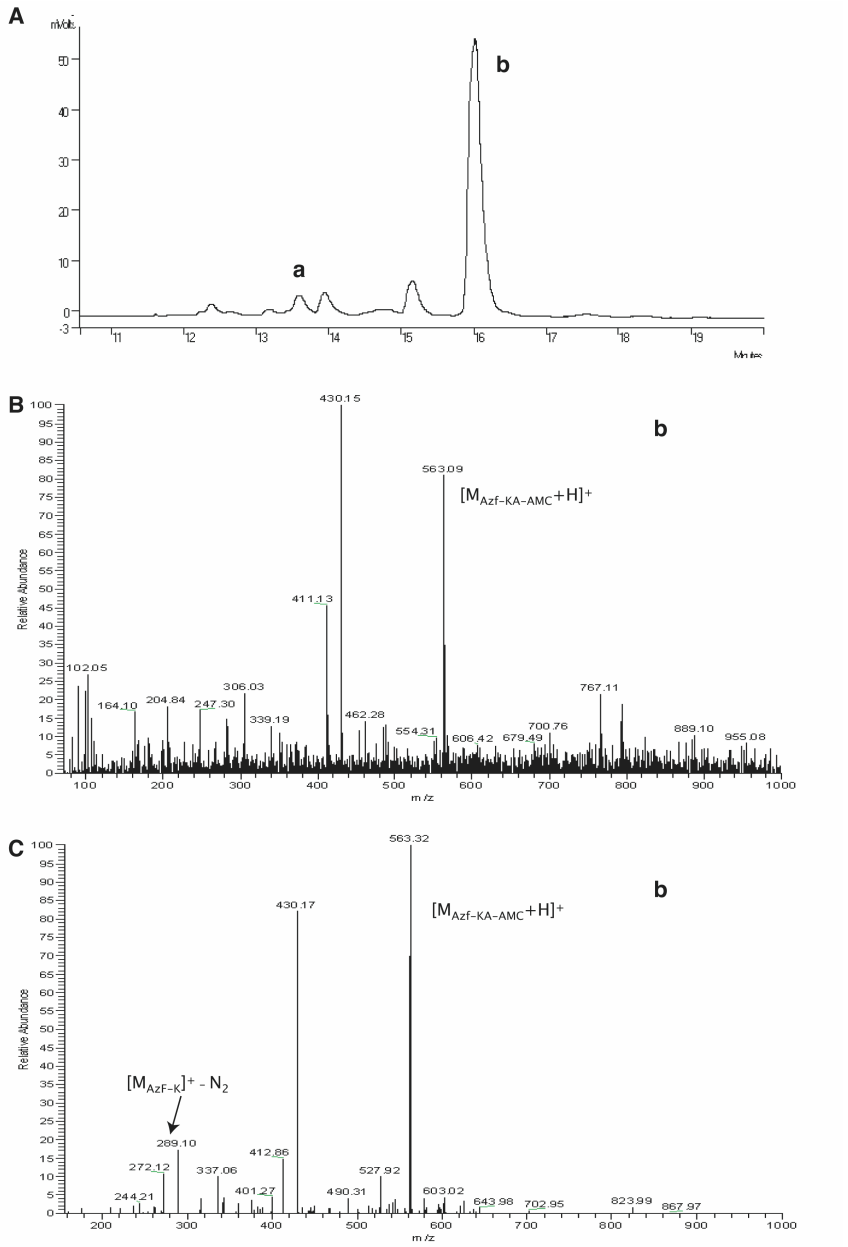

Figure S6. A) Representative reversed-phase HPLC analysis of the addition reaction of *p*-azidophenylalanine **4** (Azf) with detection at 324 nm. The retention time of the substrate peak (a) is indicated at 13.5 min. The identity of the product peak (b), Azf-KA-AMC, was verified by ESI MS. B) Analysis of peak (b) by ESI mass spectrometry confirms that it contains the tri-peptide product, Azf-KA-AMC, expected $[M_{\text{Azf-KA-AMC}}+H]^+$ =563.27 *m/z.* C) MS/MS analysis of the singly charged ion of the tripeptide product leads to the loss of dinitrogen from daughter ion Azf-K, expected $[M_{\text{amino-FK}}+H]^+$=289.18 *m/z.*

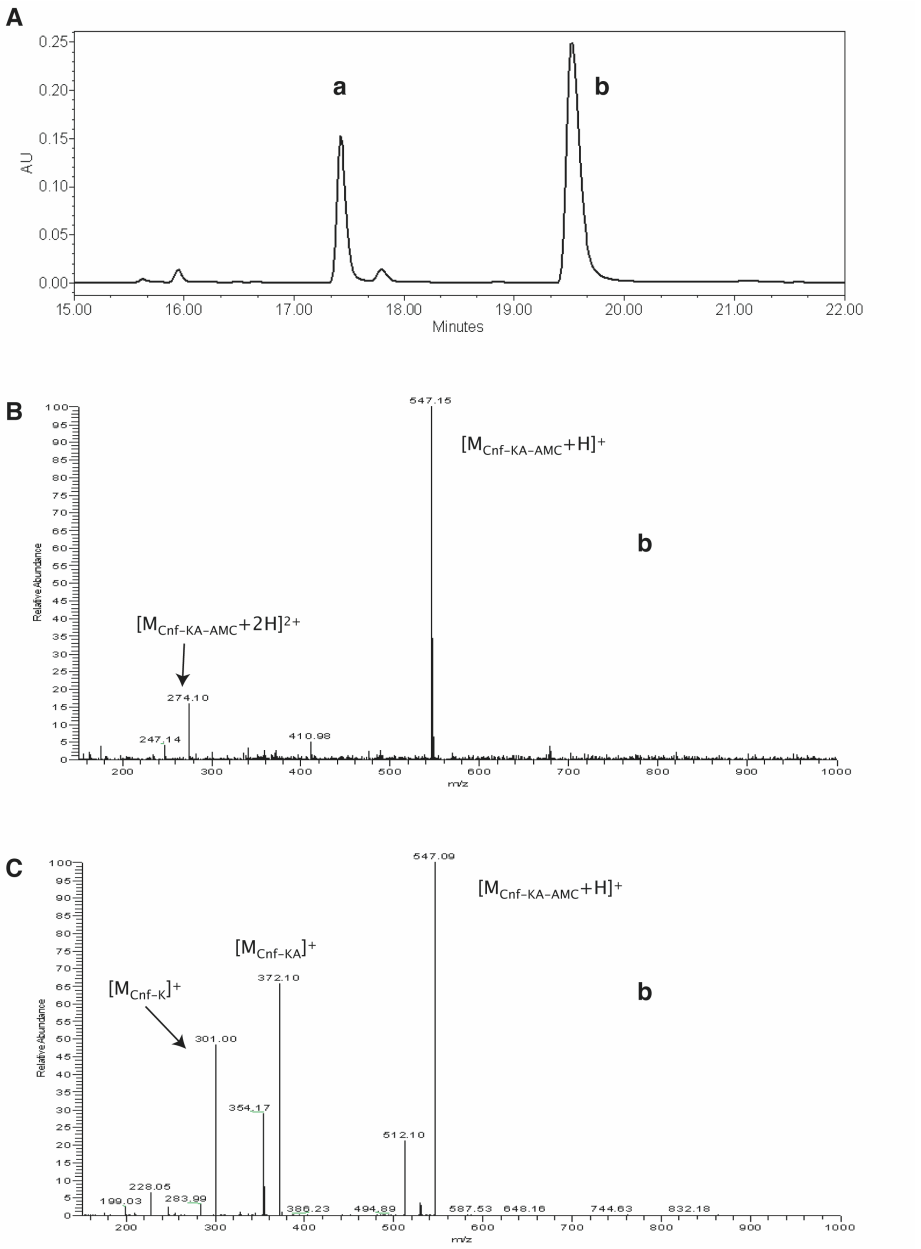

Figure S7. A) Representative reversed-phase HPLC analysis of the transfer reaction of *p*-cyanophenylalanine **6** (Cnf) to KA-AMC using UV detection at 324 nm. Both the starting material (a) and the tripeptide product (b) are visible. B) Analysis of peak (b) by ESI mass spectrometry confirms that it contains the tripeptide product, Cnf-KA-AMC, expected $[M_{\text{Cnf -KA-AMC}}+H]^+$ = 547.26 *m/z* and $[M_{\text{Cnf -KA-AMC}}+2H]^{2+}$ =274.23 *m/z.* C) MS/MS analysis of the singly charged ion of the product further confirms the N-terminal addition of Cnf to KA-AMC with fragmentation into dipeptide Cnf-K, expected $[M_{\text{Cnf -K}}]^+$=301.17 *m/z* and the tripeptide Cnf-KA, expected $[M_{\text{Cnf -KA}}]^+$= 372.2 *m/z.*

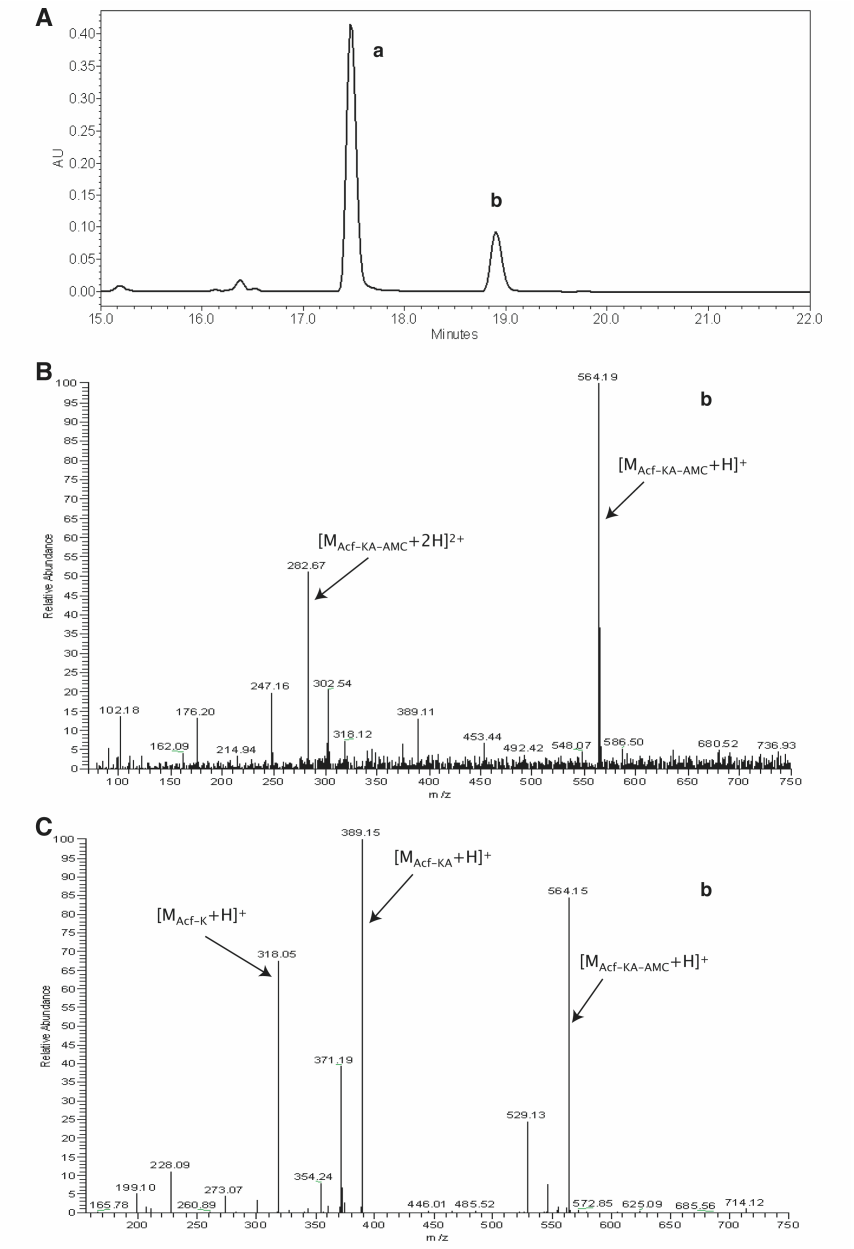

Figure S8. A) Representative reversed-phase HPLC analysis of the transfer reaction of *p*-acetylphenylalanine **7** (Acf) to KA-AMC using UV detection at 324 nm. Both the starting material (a) and the tripeptide product (b) are visible. B) Analysis of peak (b) by ESI mass spectrometry confirms that it contains the tripeptide product, Acf-KA-AMC, expected $[M_{\text{Acf-KA-AMC}}+H]^+$ =564.27 *m/z* and $[M_{\text{Acf-KA-AMC}}+2H]^{2+}$ =282.64 *m/z*. C) MS/MS analysis of the singly charged ion of the product further confirms the N-terminal addition of Acf to KA-AMC with fragmentation into dipeptide Acf-K, $[M_{\text{Ddl-K}}]^+$=318.18 *m/z* and the tripeptide Acf-KA, $[M_{\text{Acf-KA}}]^+$= 389.22 *m/z.*

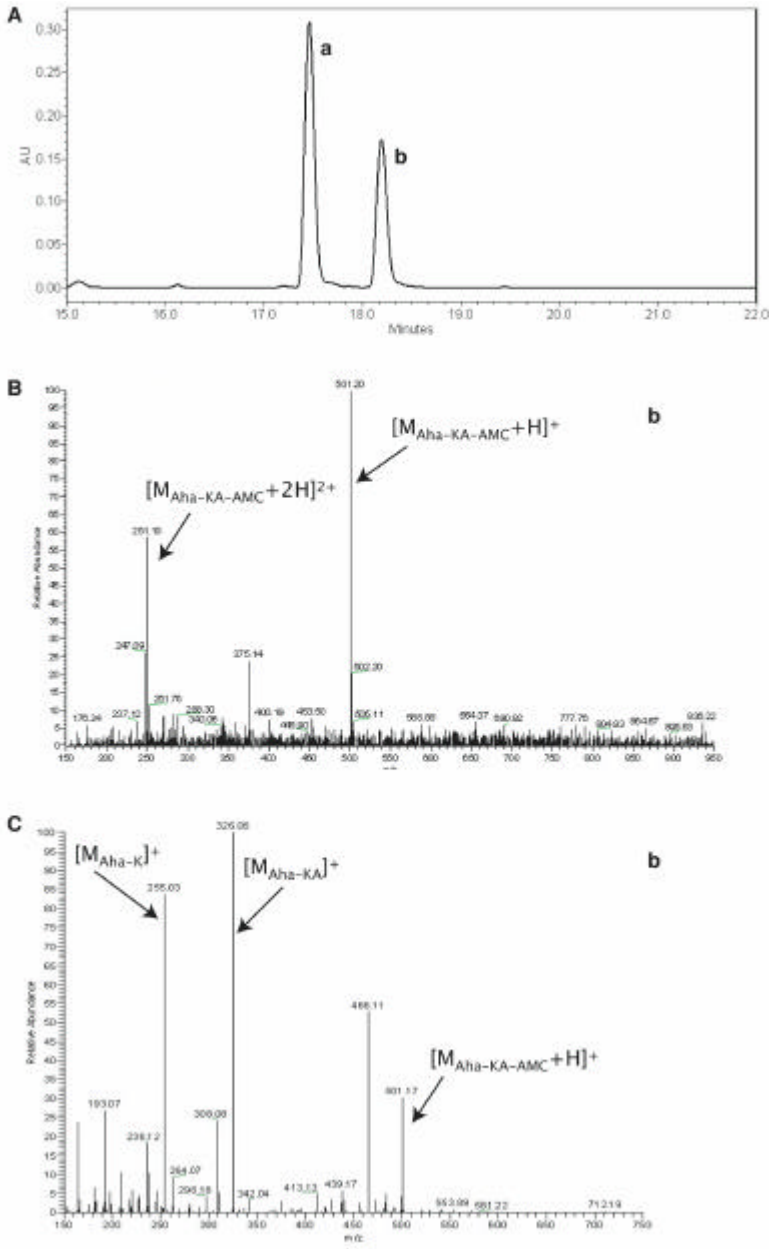

Figure S9. A) Representative reversed-phase HPLC analysis of the transfer reaction of azidohomoalanine **8** (Aha) to KA-AMC using UV detection at 324 nm. Both the starting material (a) and the tripeptide product (b) are visible. B) Analysis of peak (b) by ESI mass spectrometry confirms that it contains the tripeptide product, Aha-KA-AMC, expected $[M_{Aha\text{-}KA\text{-}AMC}+H]^+$ = 501.25 *m/z* and $[M_{Aha\text{-}KA\text{-}AMC}+2H]^{2+}$ =251.13 *m/z.* C) MS/MS analysis of the singly charged ion of the product further confirms the N-terminal addition of Aha to KA-AMC with fragmentation into dipeptide Aha-K, $[M_{Aha\text{-}K}]^+$=255.30 *m/z* and the tripeptide Aha-KA, $[M_{Aha\text{-}KA}]^+$= 326.19 *m/z.*

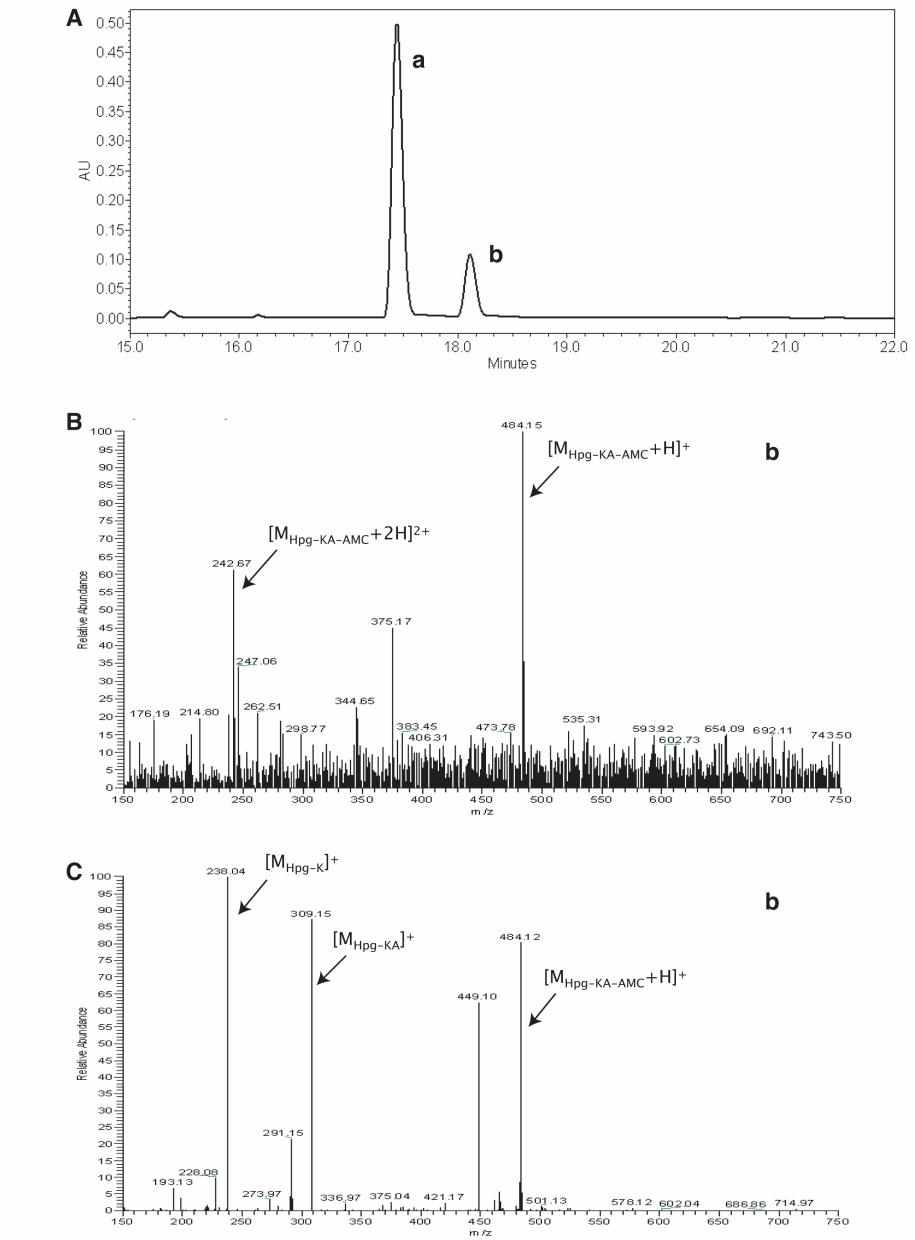

Figure S10. A) Representative reversed-phase HPLC analysis of the transfer reaction of homopropargylglycine **9** (Hpg) to KA-AMC using UV detection at 324 nm. Both the starting material (a) and the tripeptide product (b) are visible. B) Analysis of peak (b) by ESI mass spectrometry confirms that it contains the tripeptide product, Hpg-KA-AMC, expected $[M_{\text{Hpg-KA-AMC}}+H]^+$=484.25 *m/z* and $[M_{\text{Hpg-KA-AMC}}+2H]^{2+}$ =242.63 *m/z.* C) MS/MS analysis of the singly charged ion of the product further confirms the N-terminal addition of Hpg to KA-AMC with fragmentation into dipeptide Hpg-K, expected $[M_{\text{Hpg-K}}]^+$=238.16 *m/z* and the tripeptide Hpg-KA, expected $[M_{\text{Hpg-KA}}]^+$= 309.19 *m/z.*

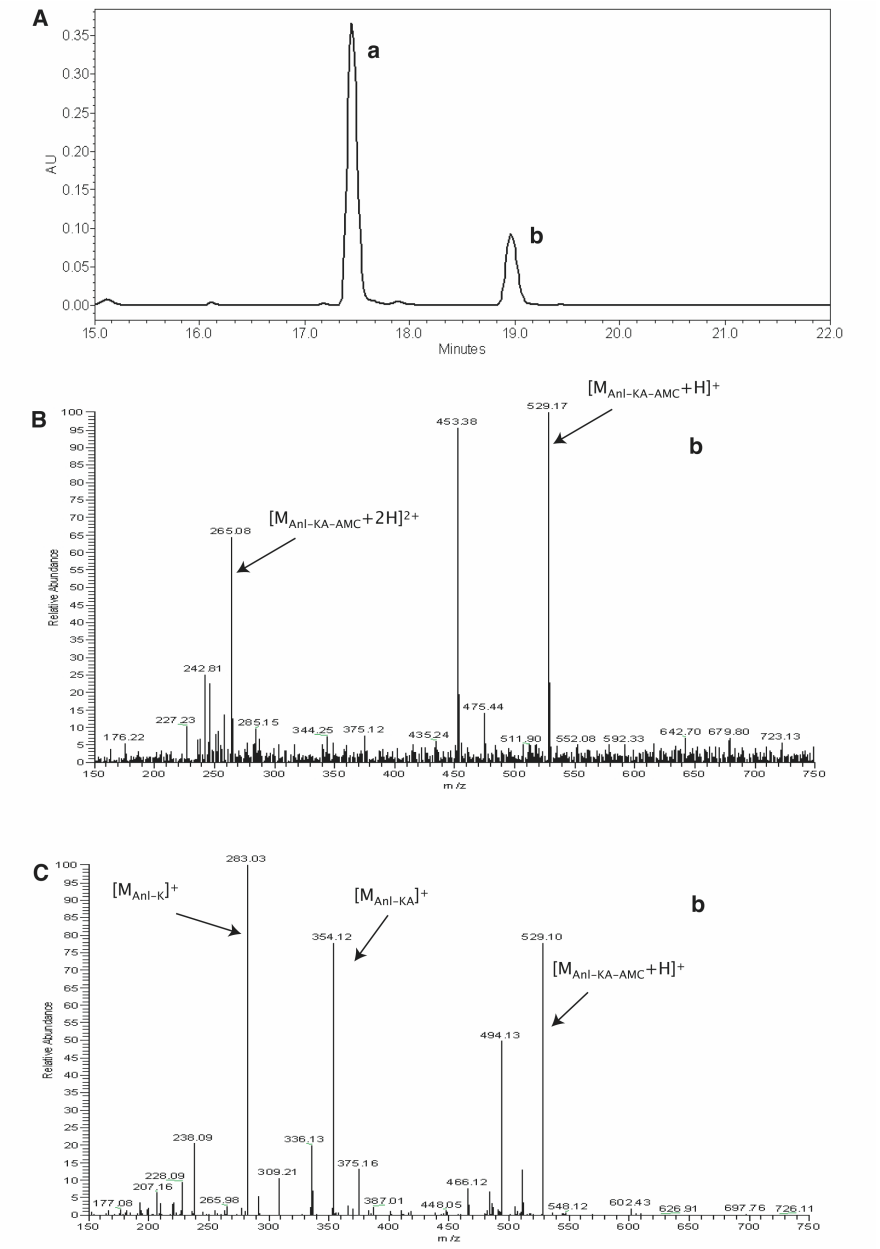

Figure S11. A) Representative Reversed-phase HPLC analysis of the transfer reaction of azidonorleucine **10** (Anl) to KA-AMC using UV detection at 324 nm. Both the starting material (a) and the tripeptide product (b) are visible. B) Analysis of peak (b) by ESI mass spectrometry confirms that it contains the tripeptide product, Anl-KA-AMC, expected $[M_{\text{Anl-KA-AMC}}+\text{H}]^+$ = 529.28 *m/z* and $[M_{\text{Anl-KA-AMC}}+2\text{H}]^{2+}$ = 265.14 *m/z.* C) MS/MS analysis of the singly charged ion of the product further confirms the N-terminal addition of Anl to KA-AMC with fragmentation into dipeptide Anl-K, expected $[M_{\text{Anl-K}}]^+$=283.19 *m/z* and the tripeptide Anl-KA, expected $[M_{\text{Anl-KA}}]^+$= 354.23 *m/z.*

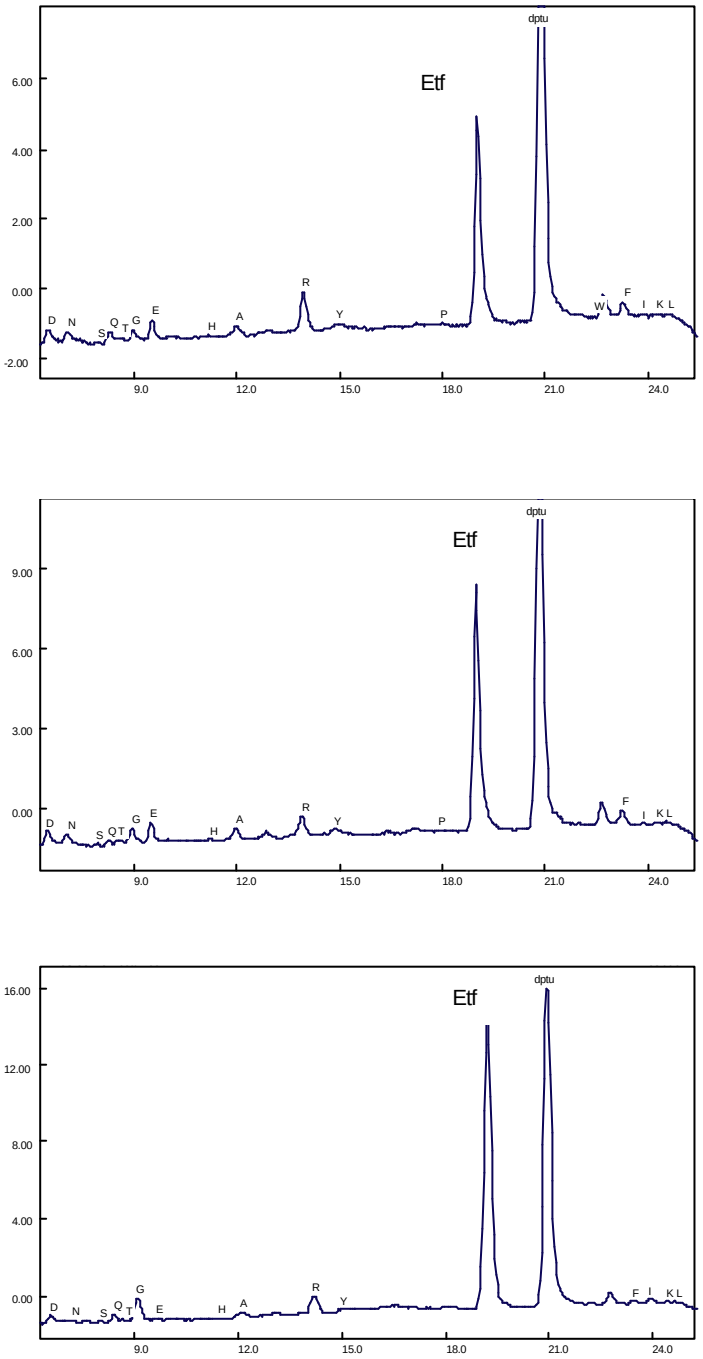

Figure S12. Analysis of the first residue of Etf-R-eDHFR by Edman degradation after A) 2 h, B) 3 h, and C) 4 h. The peak labeled “R” arises from unmodified protein while the “Etf” peak arises from modified protein.