Supplementary Information

Expanding the Genetic Code for Site-Specific Labelling on Tobacco Mosaic Virus Coat Protein and Building Biotin-functionalized Virus-Like Particles

F. C. Wu, [a,c,d,‡] H. Zhang,[b,‡] Q. Zhou,[a,c] M. Wu,[a] Z. Ballard,[d] Y. Tian,[a] J. Y. Wang,[b,*] Z. W. Niu [a,*] and Y. Huang [a]

[a] Technical Institute of Physics and Chemistry, CAS, Beijing 100190, China.

[b] National Key Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China.

[c] University of Chinese Academy of Sciences, Beijing 100080, China.

[d] Max Planck Institute for the Science of Light, Erlangen, D-91058, Germany.

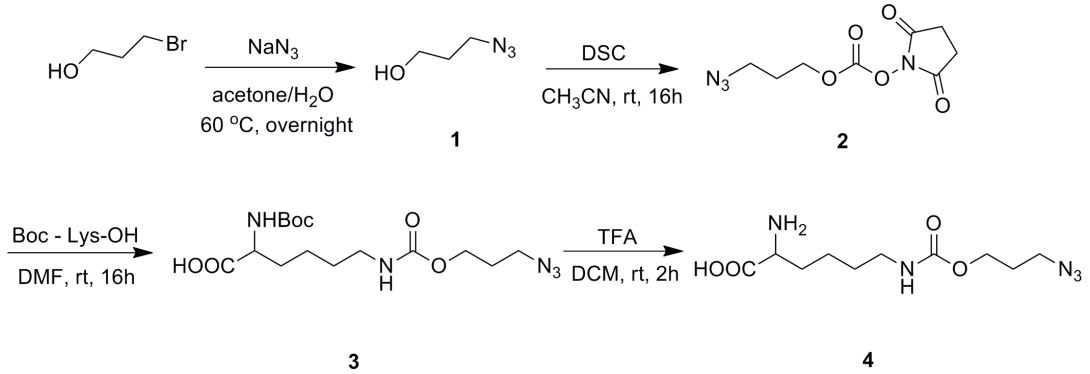
General materials and methods

The genes of TMVCP and the corresponding primers were synthesized by Sangon Biotech Co., Ltd. (shanghai, China). Enzymes were ordered from New England Biolabs (NEB). *E. coli* BL21 competent cells were purchased from TransGen Biotech Co., Ltd. (Beijing, China). 3-bromopropan-1-ol, N, N'-disuccinimidyl carbonate, and Dibenzocyclooctyne-PEG_4-biotin conjugate were obtained from Sigma Aldrich. Boc-Lys-OH was bought from GL Biochem Ltd. (Shanghai, China). Trifluoroacetic acid (TFA) was ordered from Bomaijie Technology Co., Ltd. (Beijing, China). Bathophenanthroline sulfonate sodium salt was purchased from Heowns Biotech Co., Ltd. (Tianjin, China). Azide-PEG_4-biotin conjugate was bought from Mag Biotech Co., Ltd. Uranyl acetate staining solution was bought from Zhongjingkeyi Technology Co., Ltd. (Beijing, China). Other chemical compounds were of analytical grade and purchased from Beijing Chemical Works (Beijing, China). All the reagents

were used without further purification. Dialysis bag (MW3500) was bought from Biodee Co., Ltd (Beijing, China). Desalting column (sephadex G-25) was purchased from GE Healthcare. Carbon-coated copper grids (T10023) for TEM were purchased from Xinxing Braim Technology Co., Ltd (Beijing, China). Zip-Tip pipette tipses (C_{18}) were bought from Merk Millipore China.

The concentration of TMVCP was determined by Bradford protein assay kit (Beyotime Institute of Biotechnology, Beijing, China) using BSA as a reference. ^{1}H NMR spectra were recorded on a Bruker Avance-400 instrument. Transmission Electron Microscopy (TEM) images were obtained using the JEOL JEM-2100F transmission electron microscope at 200KV accelerating voltage. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out on a Mini-Protean 3 system from Bio-Rad. MALDI-TOF mass spectra were acquired on a AXIMA-CFP plus (KRATOS Analytical, Shimadzu Group Company). Mass spectrometer equipped with a nitrogen laser (337.1 nm). Mass spectra were obtained in positive ion and linear mode with an acceleration voltage of 20kV and averaged over 100 laser shots.

Synthesis of unnatural amino acid



Scheme S1. Synthesis route of KPN.

Synthesis of 3-azidopropan-1-ol (1).

This compound was synthesized according to the procedure described by D. Quémener *et. al*[1].

3-bromopropan-1-ol. (1.28 g, 9.2 mmol, 835 μL) and sodium azide (980 mg, 15 mmol) were dissolved in acetone: H_2O (3:1, 40 mL), reflux overnight. Acetone was then evaporated in vacuum, and the residue was extracted by diethyl ether (25 mL × 3), dried with Na_2SO_4, and then filtered. The solvent was removed leaving behind the colorless liquid in 78% yield. (725 mg, 7.18 mmol). ^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 3.78 (t, 2H), 3.49 (t, 2H), 2.36 (s, 1H), 1.91-1.84 (m, 2H). IR (ν cm^{-1}): 3417, 2916, 2100, 1633, 1083, 1016, 450.

Synthesis of 3-azidopropyl (2, 5-dioxopyrrolidin-1-yl) carbonate (2).

N, N'-disuccinimidyl carbonate (DSC) (2.97 g, 11.6 mmol) and triethyl amine (2.9 mL, 20.5 mmol) were dissolved in CH_3CN (25 mL). 3-azidopropan-1-ol **1** (690 mg, 6.83 mmol) was added into the solution. The mixture was stirred at room temperature for 16 h. The solvent was removed in vacuum and the residue was chromatographed by silica gel with DCM:MeOH (20:1) as an eluent. The yellow serum (1.5 g, 5.51 mmol) was obtained in 91% yield. ^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 4.48 (t, 2H), 3.53 (t, 3H), 2.90 (s, 4H), 2.10-2.04 (m, 2H). IR (ν cm^{-1}): 2951, 2104, 1815, 1789, 1740, 1260, 1210, 1091, 649.

Synthesis of 6-(((3-azidopropoxy) carbonyl) amino)-2-((tert-butoxycarbonyl) amino) hexanoic acid (3).

3-azidopropyl (2,5-dioxopyrrolidin-1-yl) carbonate **2** (484 mg, 2.0 mmol) was dissolved in dry DMF (8 mL), Boc-lysine (639 mg, 2.6 mmol) was added to the solution and stirred at room temperature for 16 h. Then ice-cold water (100 mL) was added to the solution. The mixture was extracted with DCM (50 mL × 2) and the organic layers were combined and subsequently washed with brine (25 mL × 3). The organic layer was then dried over Na_2SO_4, filtered, and evaporated. The yellow solid (594 mg, 1.59 mmol) was obtained in 80% yield. ^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 7.96 (s, 1H), 4.82 (br, 1H), 4.23 (br, 1H), 4.15-4.12 (m, 1H), 4.08 (t, 2H), 3.32 (t, 2H), 3.11 (t, 2H), 1.82 (t, 2H), 1.61 (m, 2H), 1.47 (t, 2H), 1.38 (s, 9H), 1.25-1.19

(m, 2H). LC-ESI-MS: m/z calcd for $C_{15}H_{27}N_5O_6$ $[M+Na]^+$: 396; found 396, $[M+H\text{-}boc]^+$: 274.2; found 274.

Synthesis of 2-amino-6-(((3-azidopropoxy) carbonyl) amino) hexanoic acid (4).

Compound **4** was synthesized according to the procedure described by D.P. Nguyen *et. al*[2].

Compound **3** (594 mg, 1.59 mmol) was dissolved in DCM (6 ml), and TFA (6 ml) was slowly added to the solution. The mixture was stirred at room temperature for 2 h, and excess DCM was evaporated under vacuum. The residue was redissolved in MeOH (2 ml) and precipitated into diethyl ether, affording pure **4** in 90% yield (391 mg, 1.43 mmol). ^{1}H NMR (400 MHz, D2O) δ (ppm): 4.06 (t, 2H), 3.64 (t, 1H), 3.35 (t, 2H), 3.05 (t, 2H), 1.85-1.75 (m, 4H), 1.50-1.43 (m, 2H), 1.37-1.29 (m, 2H). ESI-MS: m/z calcd for $C_{10}H_{19}N_5O_4$ $[M+Na]^+$: 296; found 296, $[M+H]^+$: 274; found 274.

Construction of protein expression plasmids

Wild type TMVCP gene was firstly synthesized and cloned into the *SmaI* blunt terminal of pUC57 clone vector by Sangon Biotech Co., Ltd. (shanghai, China), and then was subcloned into the *NdeI* and *EcorI* restriction sites of PET-23a expression plasmid to generate the plasmid of pET23a-TMVCP. The amber mutation of Phe11TAG was introduced by site-directed mutagenesis with primers Phe11TAGF (5'-GGTGTTCTTGTCATCA-3') and Phe11TAGR (5'-TACTGAGATGGAGTAG-3') to construct TMVCP-KPN11TAG-pET23a expression plasmid.

Construction of the pEVOL- KPNRS-tRNApyl CUA plasmid:

KPNRS was amplified by the primers 5'-GGAAGATCTATGGATAAAAAA CCGCTGGATG-3' and 5'-ACGCGTCGACTTACAGGTTCGTGCTAATG-3'. The PCR product was digested with restriction enzymes BglII and SalI and ligated into pEVOL vector, affording pEVOL-1. KPNRS was again amplified by the primers 5'-GTCCAGTCATATGGATAAAAAACCGCTGGATGTG-3' and 5'-ATCAGAC

TCGAGCAGGTTCGTGCTAATGCCGTTATA-3'. The PCR product was digested with restriction enzymes NdeI and PstI and ligated into pEVOL-1 vector, affording pEVOL-2. MbtRNApyl CUA was amplified with primers 5'-GTGCACGGCTAA CTAAGCGGCCTGC-3' and 5'-CTCGAGCATGCAAAAAAGCCTGCTC-3'. The PCR product was digested with restriction enzymes ApaLI and XhoI and ligated into pEVOL-2 vector, affording pEVOL- KPNRS-tRNApyl CUA.

The synthesized wild type TMVCP gene sequence:

5'-

CATATGTCTTACAGTATCACTACTCCATCTCAGTTCGTGTTCTTGTCATCAG
CGTGGGCCGACCCAATAGAGTTAATTAATTTATGTACTAATGCCTTAGGA
AATCAGTTTCAAACACAACAAGCTCGAACTGTCGTTCAAAGACAATTCAG
TGAGGTGTGGAAACCTTCACCACAAGTAACTGTTAGGTTCCCTGACAGTG
ACTTTAAGGTGTACAGGTACAATGCGGTATTAGACCCGCTAGTCACAGCA
CTGTTAGGTGCATTCGACACTAGAAATAGAATAATAGAAGTTGAAAATCA
GGCGAACCCCACGACTGCCGAAACGTTAGATGCTACTCGTAGAGTAGACG
ACGCAACGGTGGCCATAAGGAGCGCGATAAATAATTTAATAGTAGAATTG
ATCAGAGGAACCGGATCTTATAATCGGAGCTCTTTCGAGAGCTCTTCTGGT
TTGGTTTGGACCTCTGGTCCTGCAACTTGAGAATTC-3'

The mutant TMVCP gene sequence:

5'-

CATATGTCTTACAGTATCACTACTCCATCTCAGTAGGTGTTCTTGTCATCA
GCGTGGGCCGACCCAATAGAGTTAATTAATTTATGTACTAATGCCTTAGG
AAATCAGTTTCAAACACAACAAGCTCGAACTGTCGTTCAAAGACAATTCA
GTGAGGTGTGGAAACCTTCACCACAAGTAACTGTTAGGTTCCCTGACAGT
GACTTTAAGGTGTACAGGTACAATGCGGTATTAGACCCGCTAGTCACAGC
ACTGTTAGGTGCATTCGACACTAGAAATAGAATAATAGAAGTTGAAAATC
AGGCGAACCCCACGACTGCCGAAACGTTAGATGCTACTCGTAGAGTAGAC
GACGCAACGGTGGCCATAAGGAGCGCGATAAATAATTTAATAGTAGAATT

GATCAGAGGAACCGGATCTTATAATCGGAGCTCTTTCGAGAGCTCTTCTG GTTTGGTTTGGACCTCTGGTCCTGCAACTTGAGAATTC-3’

Note: The yellow highlighted residues are restriction sites; and the green highlighted residues are the Phe11 and the substituted amber codon, respectively.

Protein expression and purification

The two plasmids of pEVOL-KPNRS-tRNApyl CUA (chloromycetin resistance) and pET23a-TMVCP-Phe11TAG (ampicillin resistance) were co-transformed into *E.coli* BL21 and cultured on the plate which contains 100 μg/mL ampicillin and 50 μg/mL chloromycetin at 37 °C overnight. Then a single colony was selected and inoculated in 5 mL LB medium containing 100 μg/mL ampicillin and 50 μg/mL chloromycetin at 37 °C, 220 rpm for 12h. After that, 1 mL of this culture was added to 100 mL fresh LB medium supplemented with 100 μg/mL ampicillin and 50 μg/mL chloromycetin at 37 °C, 220 rpm overnight, then 10 mL of this overnight culture was used to inoculate in 1 L of the same LB medium mentioned above. KPN, L-arabinose and isopropyl β-D-1-thiogalactopyranoside (IPTG) were added to the final concentration of 1 mM, 0.2%, and 1 mM respectively when the culture was grown at 37 °C to an OD 600 ~ 0.8. Then the protein was induced to express at 16 °C, 220 rpm overnight.

Cells were harvested by centrifugation at 4000 g for 20 min at 4 °C, and the pellets were frozen by liquid nitrogen. 10 mL buffer (100 mM Tris-HCl pH 8.0, 50 mM EDTA, 200 mM NaCl, 0.1% 2-mercaptoethal) was used to resuspend the frozen pellet. When it was uniformly dispersed in the solution, it was sonicated with a 6s-on and 6s-off cycle for 30 min. Clarified supernatant was collected by centrifugation at 13000 g for 20 min 4 °C. The TMVCP in the supernatant was precipitated by the addition of ammonium sulfate to 80%, and then the precipitate was dissolved in 15% ammonium sulfate. The mixture was dialyzed against 2 L buffer which contains 100 mM sodium acetate pH 5.2 at 4 °C overnight. The precipitate was removed by centrifugation at

13000 g for 20 min at 4 °C, and the supernatant was further dialyzed against 50 mM citrate buffer pH 3.5. The precipitate was collected by centrifugation at 4 °C, 13000 g for 20 min and dissolved in 20 mM Tris-HCl pH 7.6, then the resulting supernatant was collected by centrifugation at 4 °C, 13000 g for 20 min.

Self-assembly of KPN11-TMVCP and KPN11-TMVCP-biontin into disks and rods

The final concentration of the protein used in previously descripted experiments was 1 mg/mL. Rod-like assemblies were produced upon dialysis against 100 mM sodium acetate buffer at 4 °C, pH 5.5 for 24 h. For the formation of disk aggregates, the protein was dialyzed against 400 mM Tris-HCl pH 7.6 at 4 °C for 96 h, the dialysis buffer was changed every 12h. After each dialysis, two identical samples were taken to do the TEM analysis.

Site-specific biotin labelling of proteins by copper-free cycloaddition reactions

500 μL of 1 mg/mL KPN11-TMVCP and 7.2 μL of 20 mM Dibenzo- cyclooctyne-PEG_4-biotin conjugate (dissolved in water) were incubated for 12 h at 4 °C in 20 mM PBS pH 8.0. The excess substrate was separated by passing through a desalting column.

Sample preparation for TEM

Protein samples were prepared for TEM analysis by dropping 20 μL solution of protein on TEM grids. A filter paper was then used to absorb the remaining liquid, following by rinsing with pure water. The grid was then negatively stained with uranyl acetate solution for 1 min.

MALDI-TOF MS analysis

The protein sample was denatured in the solution of 6 M guanidine hydrochloride for 10 min.

The Zip-tip C_{18} column was attached to the 10μL micropipette. Then the Zip-Tip was washed 15 times with 0.1% trifluroacetic acid (TFA)/acetonitrile, followed by 15 additional washes with 0.1% TFA / 50% acetonitrile. Then the Zip-Tip was equilibrated with 0.1% TFA / water by pipetting in and out 30 times. After that the Zip-Tip was carefully filled with the denatured protein sample solution, and then slowly pipetted in and out of the tube. This step was repeated 30 times to ensure that most of the peptides were retained in the Zip-Tip. Afterwards, the Zip-Tip was washed with 50μL 0.1% TFA / water 30 times to remove the salts. And then the protein sample was eluted from the Zip-Tip in 10μL 0.1% TFA / 50% acetonitrile 30 times.

The sinapinic acid matrix was prepared by dissolving 10 mg in 1 mL of 30:70 acetonitrile / water containing 0.1% TFA. The service of MODLI-TOF sample analysis was offered by Centre of Medical and Health Analysis, Peking University.

Test of TEM calibration

Native TMV with the diameter of 18 nm was used to calibrate the size of the large disk. Both native TMV and TMVCP assemblies were negatively stained and visualized by TEM with the same magnification. As shown in Figure S1 B, the diameter for a native TMV is 18 nm, which matches well to the previously report[3], and the disk with around 30 nm in diameter can be clearly seen in Figure S1 A.

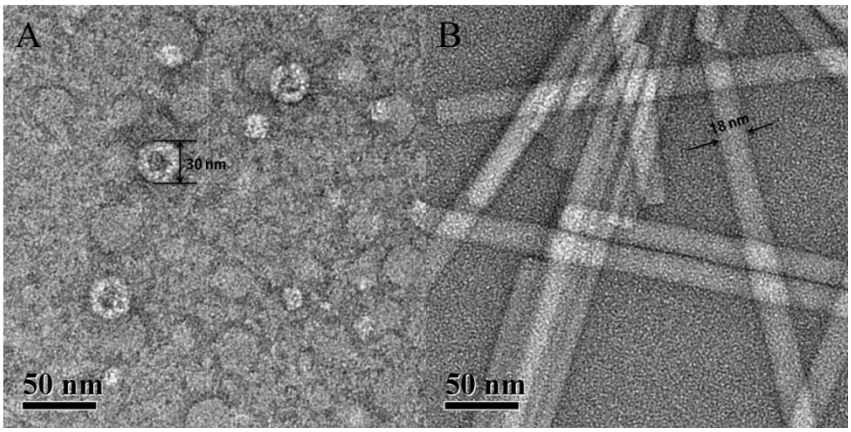


Fig. S1 Test of TEM calibration. (A) TEM image of the self-assembly KPN11-TMVCP-Biotin into large

disks; (B) TEM image of native TMV.

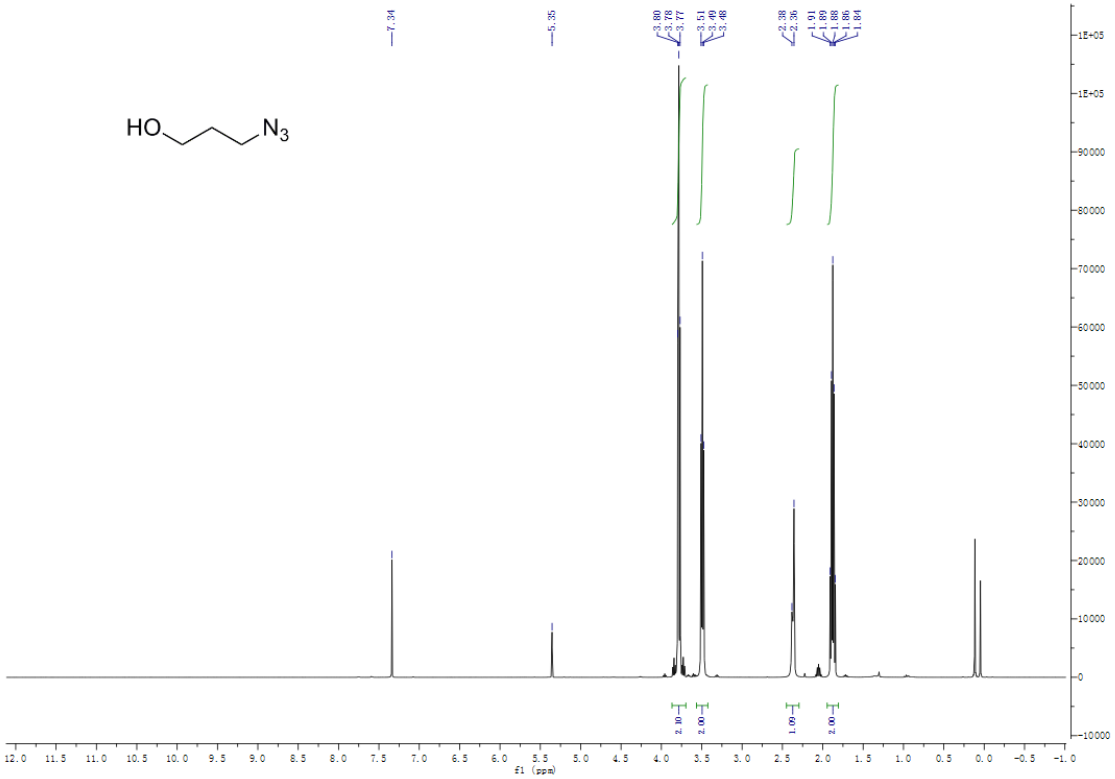


Fig. S2 ^{1}H NMR spectra of product **1.**

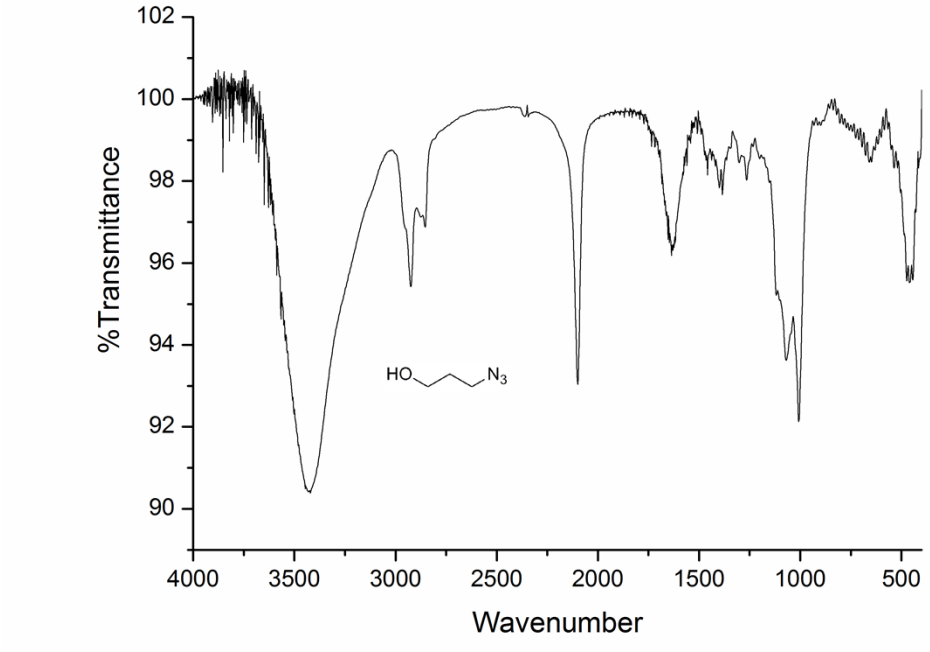


Fig.S3 FT-IR spectra of product **1.**

Fig. S4 ^{1}H NMR spectra of product **2.**

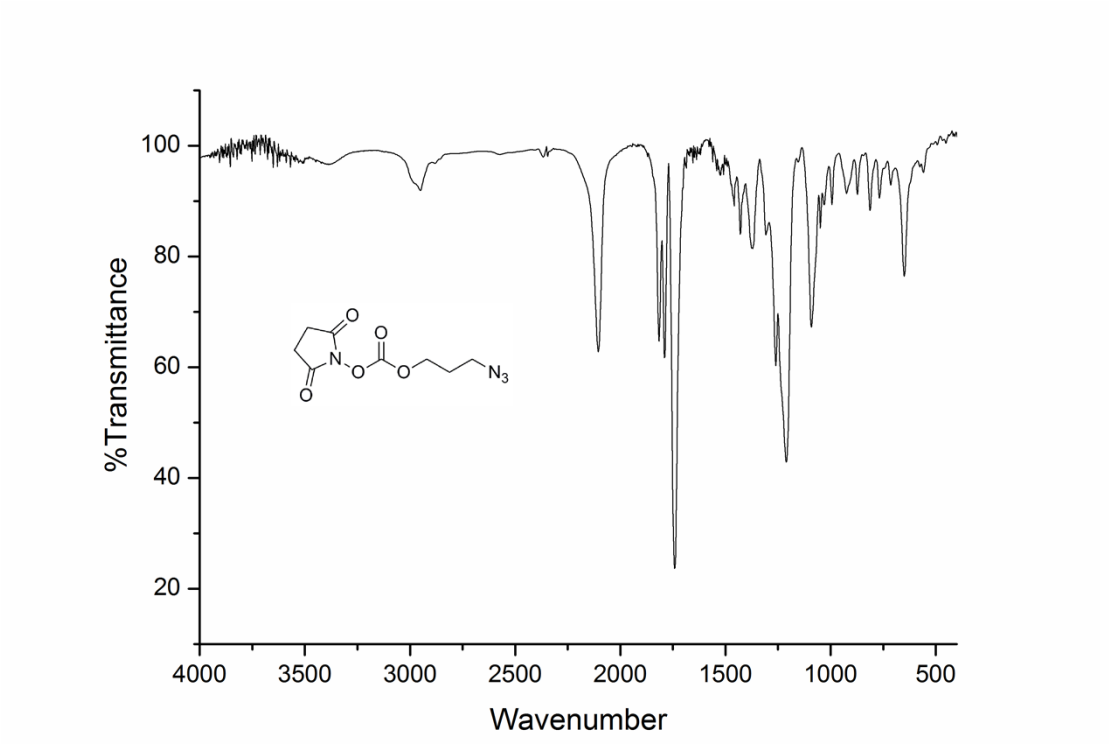


Fig. S5 FT-IR spectra of product **2.**

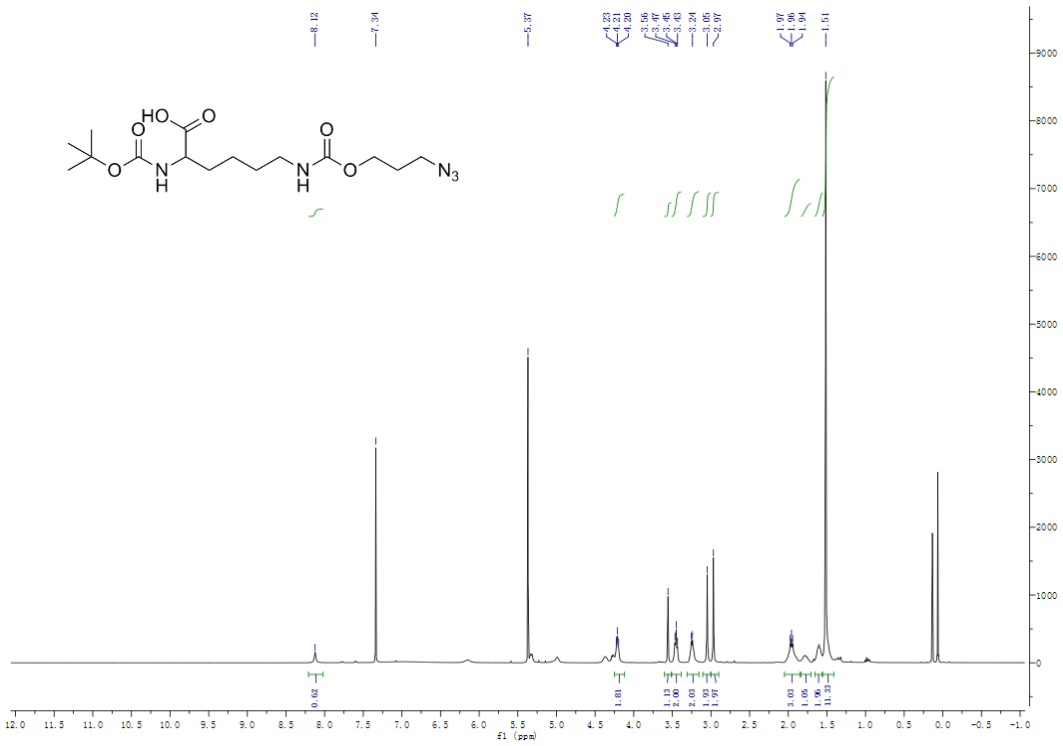


Fig. S6 ^{1}H NMR spectra of product **3.**

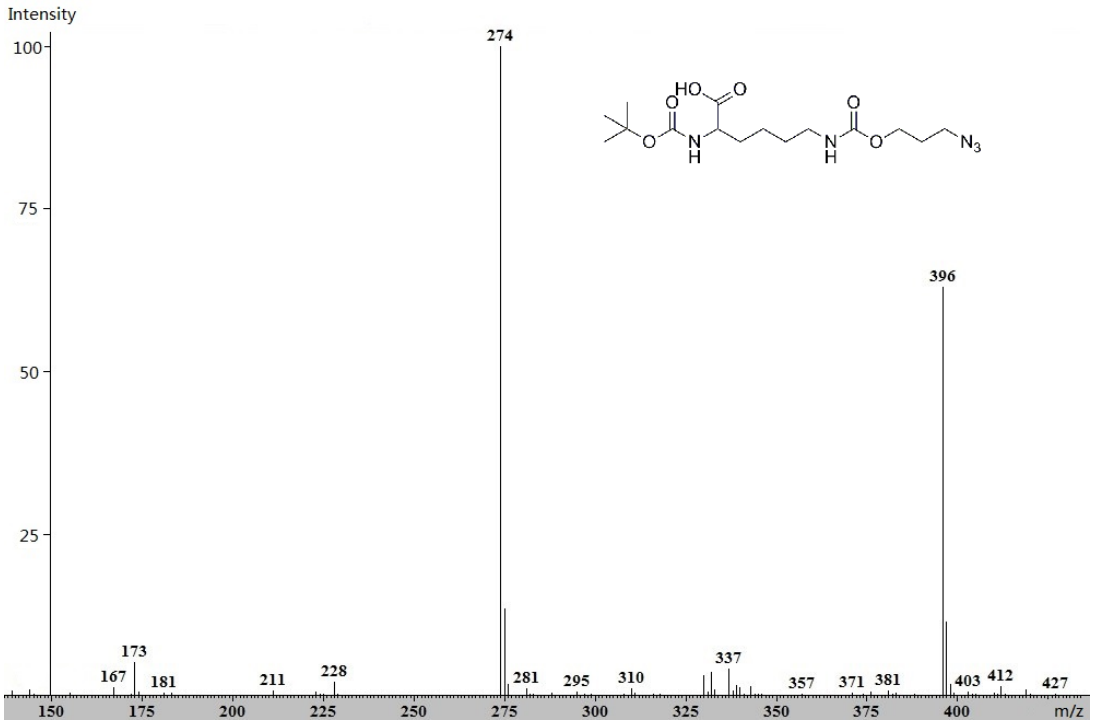


Fig. S7 LC-ESI-MS spectra of product **3.**

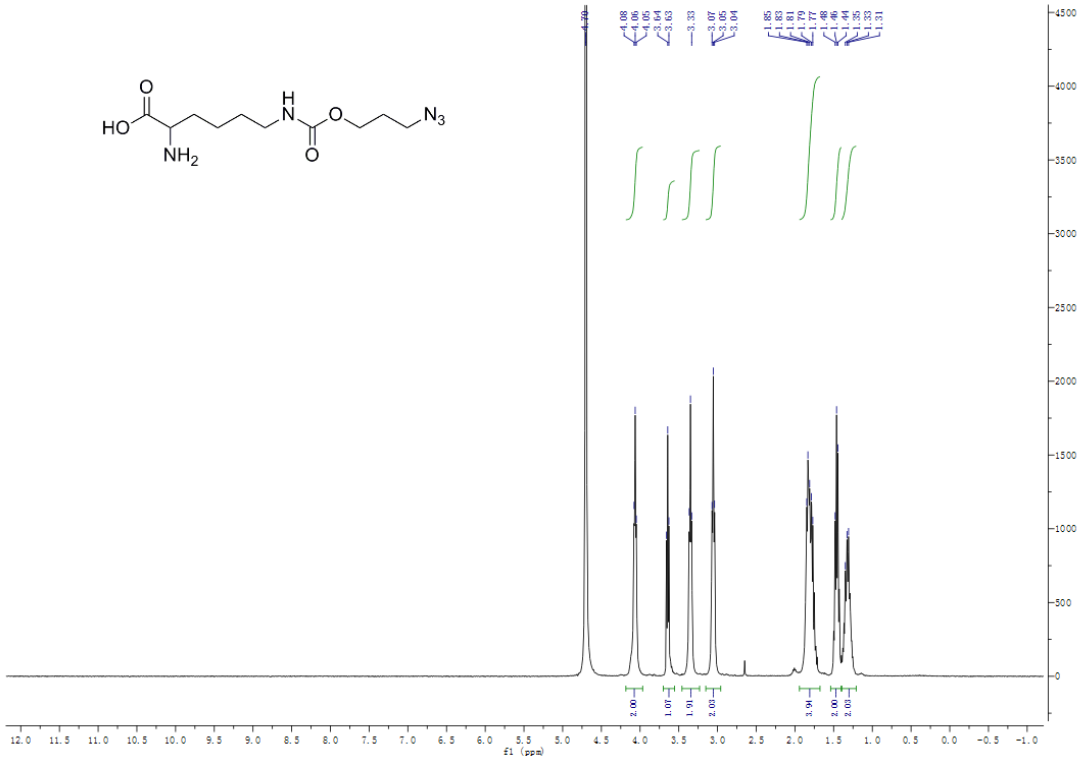


Fig. S8 ^{1}H NMR spectra of product **4.**

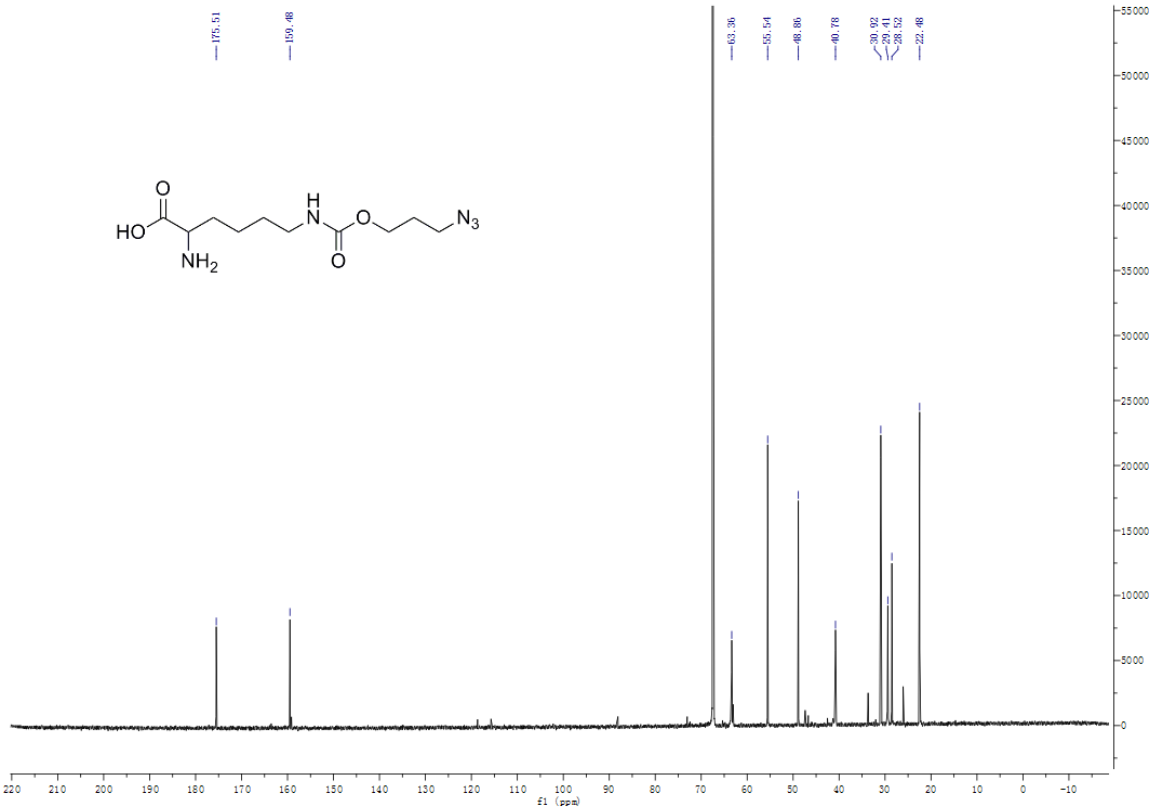


Fig. S9 ^{13}C NMR spectra of product **4.**

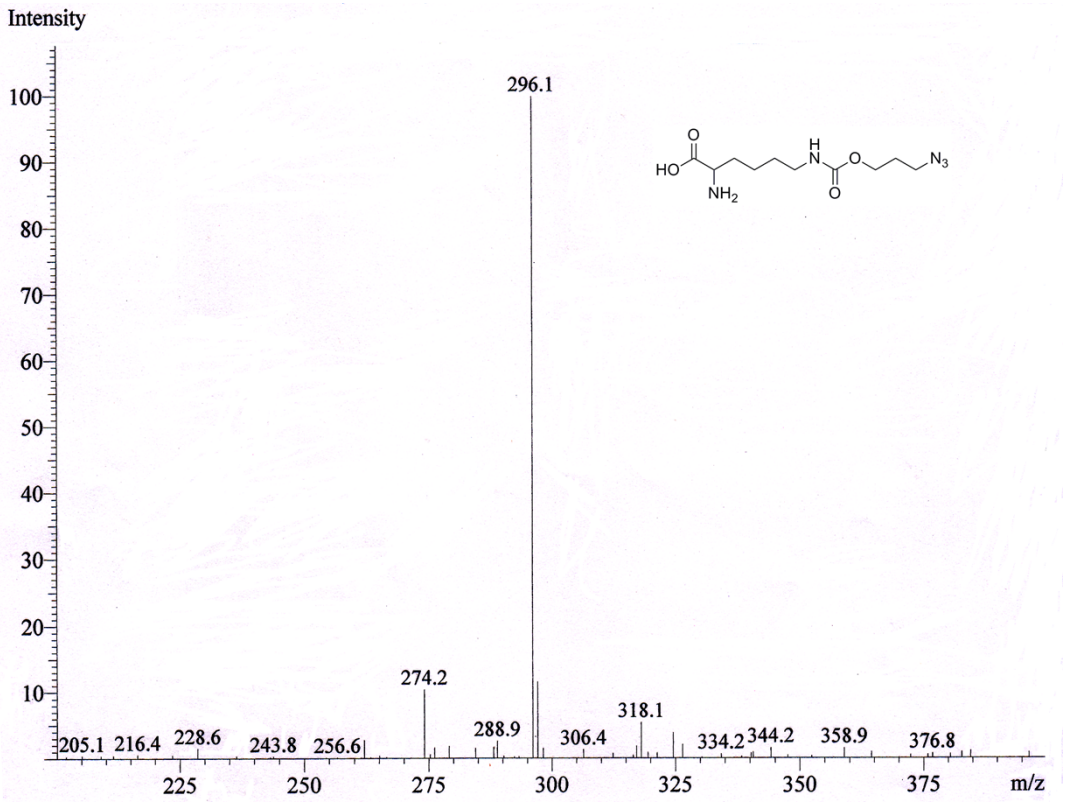


Fig. S10 ESI-MS spectra of product **4.**

References

1. D. Quémener, T. P. Davis, C. B. Kowollik and M. H. Stenzel, *Chem Commun.,* 2006, **48**, 5051-5053.
2. D. P. Nguyen, H. Lusic, H. Neumann, P. B. Kapadnis, A. Deiters and J. W. Chin, *J. Am. Chem. Soc.*, 2009, **131**, 8720.
3. R. J. Tseng , C. Tsai, L. Ma, J. Ouyang, C. S. Ozkan and Y. Yang, *Nat. Nanotech.*, 2006, **1**, 72.