Supporting Information

Cucurbit[8]uril Facilitated Michael Addition for Regioselective Cysteine Modification

Gao Li,[*a] Jun Hu,[b] Huai Chen,[b] Yong-Xiang Chen,[b] and Yan-Mei Li[*bcd]

a. Institute of Oceanography, Minjiang University, Fuzhou, Fujian 350108, China

b. Key Laboratory of Bioorganic Phosphorus Chemistry and Chemical Biology (Ministry of Education), Department of Chemistry, Tsinghua University, Beijing 100084, China

c. Beijing Institute for Brain Disorders, Beijing 100069, China

d. Center for Synthetic and Systems Biology, Tsinghua University, Beijing 100084, China

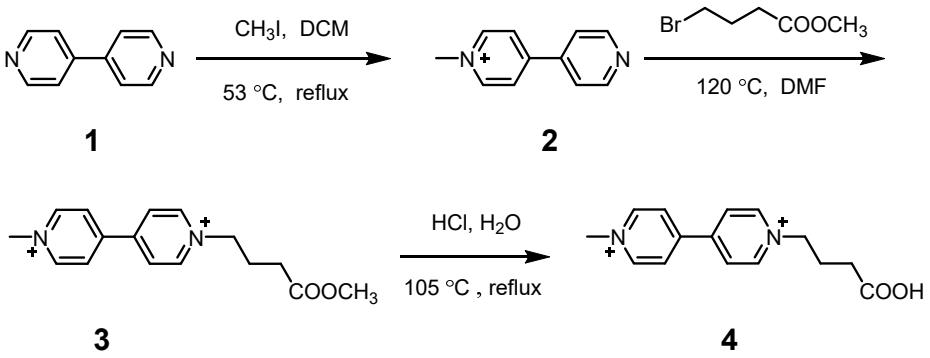
Email: ligao2019@mju.edu.cn; liym@mail.tsinghua.edu.cn.

Contents

Materials and Reagents

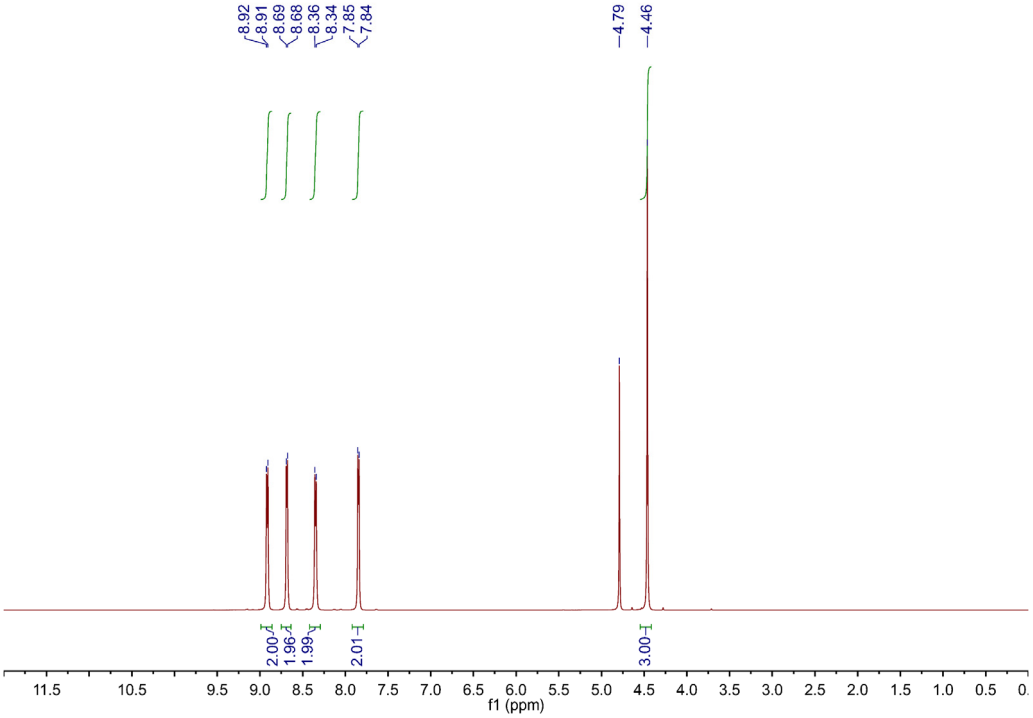
All the chemical reagents were bought from Innochem (Beijing, China). Peptides (except **Dha-MV**) with ≥95% purity were purchased from GL Biochem (Shanghai, China).

Synthesis of 4-(3-carboxypropyl)-4'-methyl viologen[1]

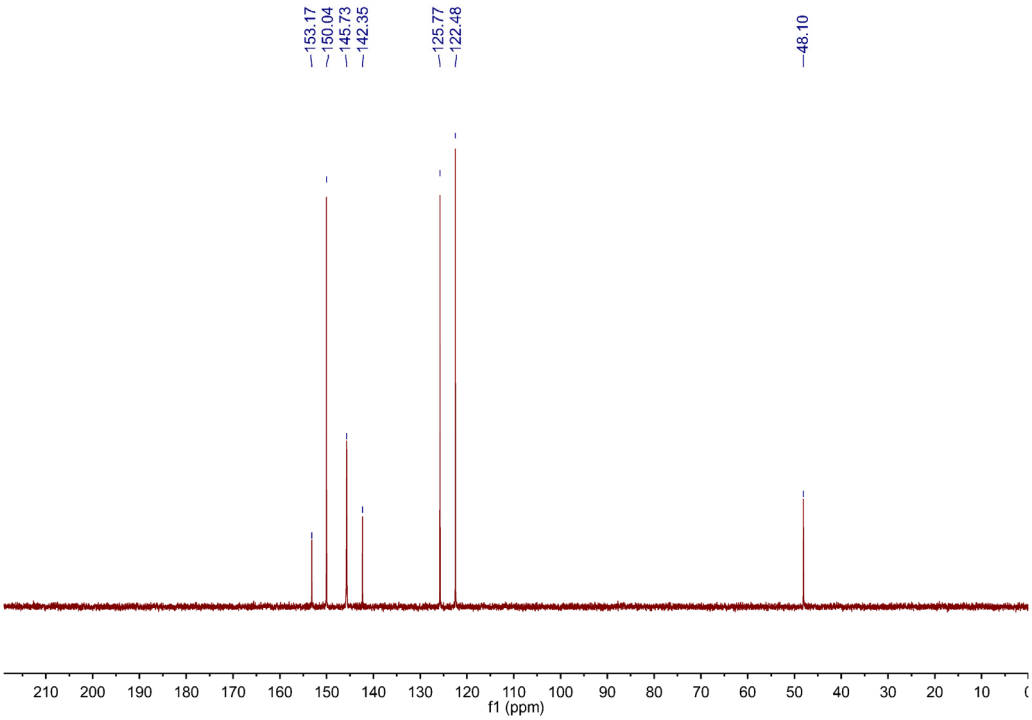


Scheme S1. Synthetic route for 4-(3-carboxypropyl)-4'-methyl viologen

16 g (102 mmol) 4,4'-bipyridine (**1**) was dissolved in 250 mL dichloromethane (DCM) and 5 mL (80 mmol) CH_3I was diluted in 50 mL DCM. Diluted CH_3I was added dropwise to 4,4'-bipyridine in a duration of 1 h at room temperature (r.t.). Then the temperature was raised to 53 °C and reacted for 2 h. The reaction mixture was cooled to r.t. and filtrated through a 0.45 μm microfiltration membrane. The solid was solubilized in CH_3OH and re-precipitated in ethyl acetate. The mixture was filtrated again and 6 g precipitation was collected as 1-methyl-[4,4'-bipyridin]-1-ium iodide (**2**), yield 25%. ^{1}H-NMR (D_2O, 400 MHz) δ (ppm): 8.92 (d, J = 4.0 Hz, 2H), 8.69 (d, J = 4.0 Hz, 2H), 8.35 (d, J = 8.0 Hz, 2H), 7.85 (d, J = 4.0 Hz, 2H), 4.46 (s, 3H). ^{13}C-NMR (D_2O, 100 MHz) δ (ppm): 153.17, 150.04, 145.73, 142.35, 125.77, 122.48, 48.10.

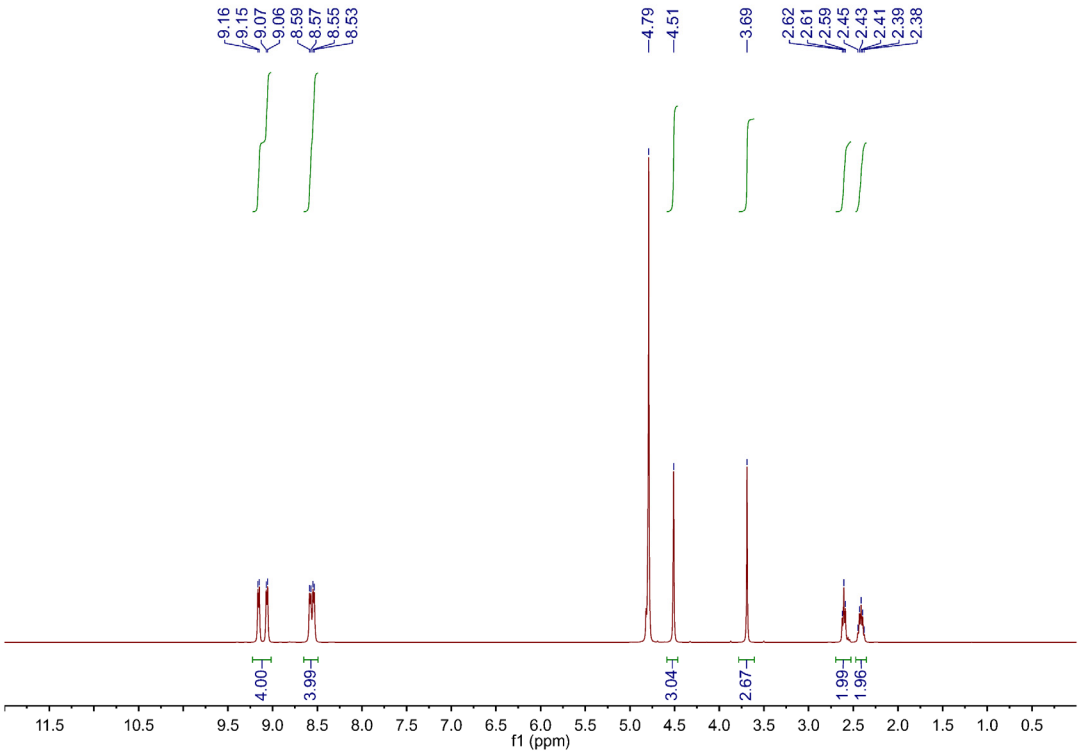


400 MHz ^{1}H NMR of **2** in D_2O

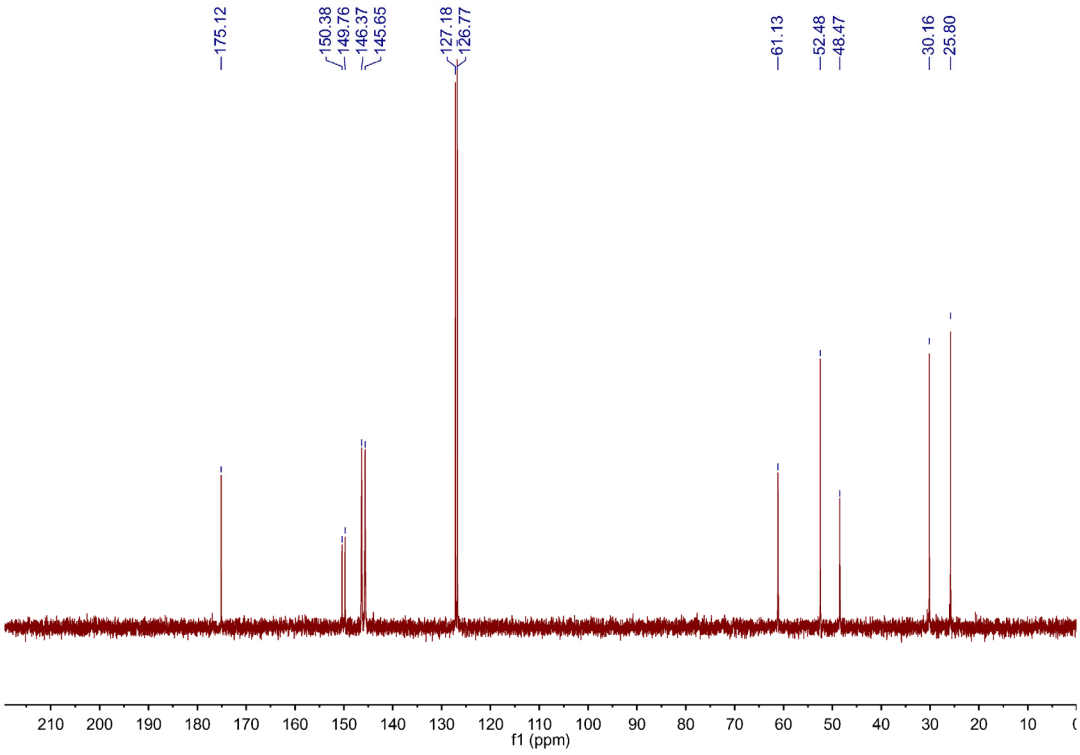


100 MHz ^{13}C NMR of **2** in D_2O

2 g **2** was solubilized in 120 °C DMF and 10 g methyl 4-bromobutanoate was added. The reaction was carried out for 48 h. Precipitations were observed during the reaction. The mixture was filtrated immediately but the precipitation A was not our desired product. The filtrate was cooled to r.t. and new precipitation B was observed. Precipitation B was solubilized in water and re-precipitated in acetone. ^{1}H-NMR confirmed that precipitation B was our desired product **3**. 0.4 g, yield 13%. ^{1}H-NMR of 4-(3-carboxypropyl)-4'-methyl viologen in D_2O, 400 MHz. δ (ppm): 9.11 (dd, J = 36, 4 Hz, 4H), 8.56 (dd, J = 16, 4 Hz, 4H), 4.51 (s, 3H), 3.69, (s, 3H), 2.61 (t, J = 8 Hz, 2H), 2.42 (m, J = 8 Hz, 2H). Please note that the peak for CH_2 linker to N was partially buried in the H_2O peak[1]. ^{13}C-NMR (D_2O, 100 MHz) δ (ppm): 175.12, 150.38, 149.76, 146.37, 145.65, 127.18, 126.77, 61.13, 52.48, 48.47, 30.16, 25.80.



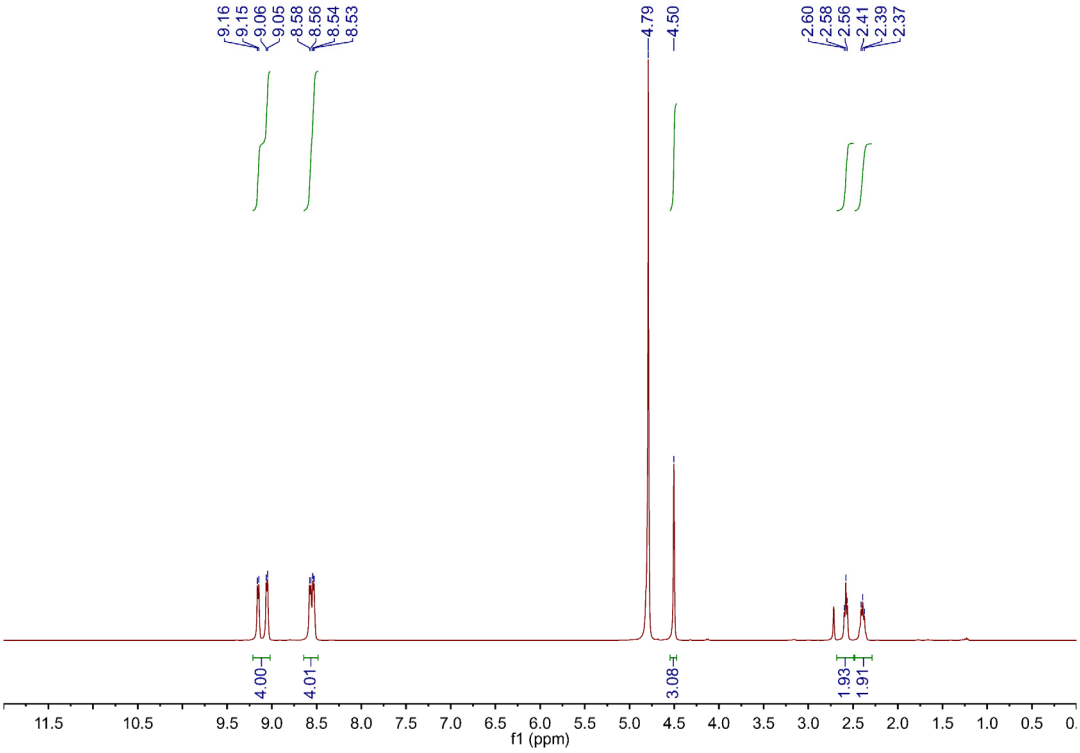
400 MHz ^{1}H NMR of **3** in D_2O



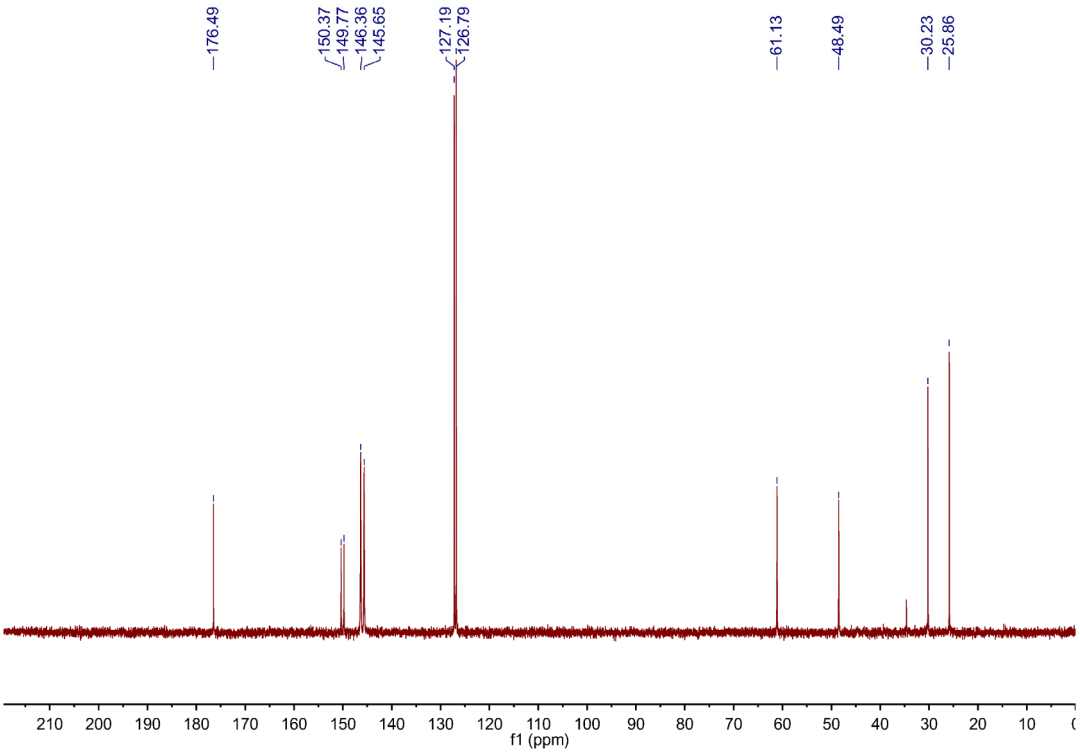
100 MHz ^{13}C NMR of **3** in D_2O

0.4 g **3** was dissolved in 12 mL concentrated hydrochloric acid and 4 mL H_2O, reaction was conducted on 105 °C for 12 h. Ethanol was added to the mixture and the solvent was removed by rotary evaporation. The solid was solubilized in water and lyophilized. 0.3 g, yield 77%. ^{1}H-NMR of 4-(3-carboxypropyl)-

4’-methyl viologen in D_2O, 400 MHz. δ (ppm): 9.11 (dd, J = 40, 4 Hz, 4H), 8.56 (dd, J = 16, 4 Hz, 4H), 4.50 (s, 3H), 2.58 (t, J = 8 Hz, 2H), 2.40 (m, J = 8 Hz, 2H). Please note that the peak for CH_2 linker to N was partially buried in the H_2O peak[1]. ^{13}C-NMR (D_2O, 100 MHz) δ (ppm): 176.49, 150.37, 149.77, 146.36, 145.65, 127.19, 126.79, 61.13, 48.48, 30.23, 25.86.



400 MHz ^{1}H NMR of **4** in D_2O

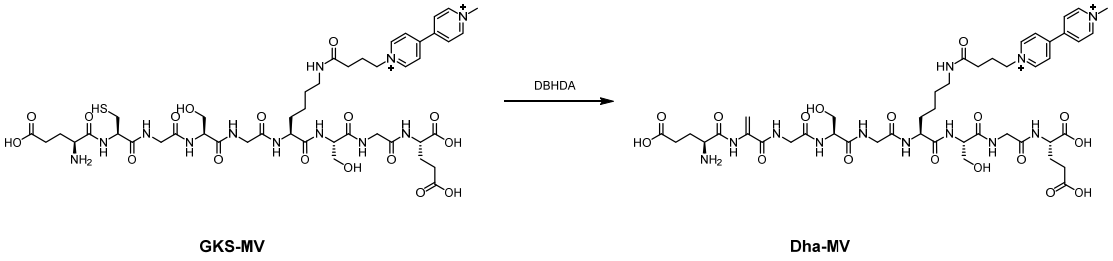


100 MHz ^{13}C NMR of **4** in D_2O

Synthesis of Dha-MV

0.075 mmol Fmoc-Glu(OtBu)-Wang resin was used for the peptide synthesis. As previously described[2], sequential coupling of Fmoc protected amino acids was achieved with HATU, HOAT and DIEA. Note that the side chain of lysine was protected with ivDde group and the N-terminal amino group was protected with Boc. The resin was incubated with 2% NH_2NH_2 in DMF for 3×20 min. Then 0.15 g 4-(3-carboxypropyl)-4'-methyl viologen was coupled to lysine with HATU, HOAT and DIEA. The peptide was deprotected and cleaved from the resin with 10 mL TFA/triisopropylsilane/H_2O=95/2.5/2.5. The crude peptide was purified by HPLC, we got 12.5 mg purified **GKS-MV** after lyophilized.

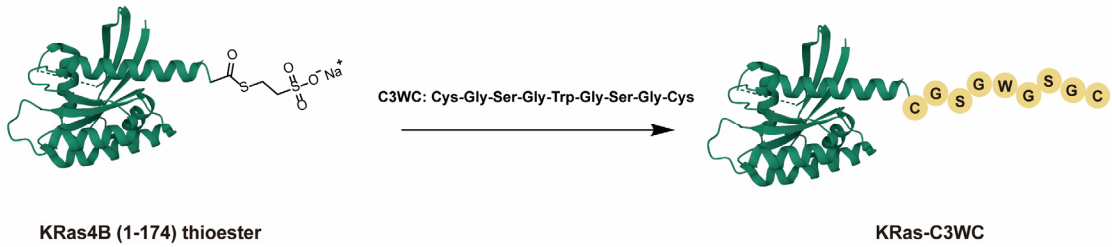
7.79 mg **GKS-MV** and 0.5 mg TCEP was solubilized in 7 mL 50 mM pH=8.5 PB buffer. Then we adjusted the pH to 8.8 by adding some NaOH. 108 mg DBHDA (2,5-Dibromohexane diacetamide) in 0.8 mL DMF was then added to the solution. After incubating at 25 °C for 1h, the mixture was shaking at 300 rpm, 37 °C for 5 h[3]. HPLC was used to purify the product and we got 4.70 mg **Dha-MV**.



Scheme S2. Conversion of GKS-MV to Dha-MV

Synthesis of KRas-C3WC

KRas4B (1-174)-intein CBD plasmid was transformed into BL21 *Escherichia coli* and purified through chitin resin[4]. After treatment with 250 mM MESNa buffer (50 mM Hepes, 500 mM NaCl, 2 mM $MgCl_2$ and 250 mM MESNa), KRas4B (1-174) thioester protein was obtained. MESNa buffer prevents the hydrolysis of KRas4B thioester protein. Native chemical ligation was performed using the method previously described[4], the reaction buffer was 250 mM MESNa buffer. 6.1 mg/mL KRas4B (1-174) thioester protein (300 μM) and 5 times excess peptide **C3WC** were mixed in the condition of 100 mM TCEP and 200 mM MPAA (16 °C, 5 h). The product **KRas-C3WC** was purified through AKTApurifier^{TM} UPC 10. The purified protein was concentrated by ultrafiltration and its concentration was determined to be 68 μM (in PB buffer) by NanoDrop (Ext. coefficient = 17420). The protein solution was aliquoted, quick-freezing in liquid nitrogen and stored at -80 °C.



Scheme S3. Synthetic route for KRas-C3WC

General HPLC and MS procedures

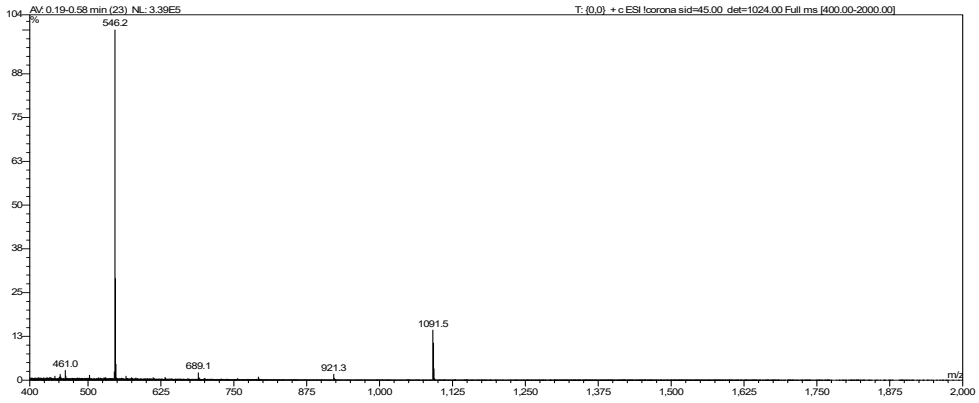
Shimadzu (Japan) LC-6AD and LC-2010A HT were used for HPLC. We used ODS column (YMC-Pack, 250×20 mml.D., S-5 μm, 12 nm or 30 nm) to purify **GKS-MV** and **Dha-MV**. Their purity was confirmed with an analytical ODS column (YMC-Exphere 150×4.6 mml.D., S-5 μm, 12 nm). For KRas

proteins, an analytical Proteonavi protein column (Col. No. J20AB01180) was used. Mobile phase A: water with 0.06% trifluoroacetate (TFA), mobile phase B: 80% acetonitrile and 20% water with 0.06% TFA. The flow rates were 10 mL/min for purification and 0.8 mL/min for purity and product analysis.

A Thermo MSQ Plus mass spectrometer was used for general ESI-MS analysis of peptides. The mobile phase of water, acetonitrile and methanol contained 0.06% formic acid and a flow rate of 0.4 mL/min was used. ESI-MS of KRas proteins and deconvolution were performed on Shimadzu LCMS-2020. MALDI-TOF-MS and MS/MS were performed at the Center of Biomedical Analysis, Tsinghua University with a 4800 Plus MALDI TOF/TOF^{TM} Analyzer. Trypsin digestion and LC-MS/MS analysis of KRas proteins were also performed at the Center of Biomedical Analysis, Tsinghua University. Proteins were reduced with dithiothreitol and alkylated with iodoacetamide before digestion with sequencing grade-modified trypsin. A Thermo-Dionex Ultimate 3000 HPLC system, which was directly interfaced with the Thermo Orbitrap Fusion mass spectrometer, was used for LC-MS/MS.

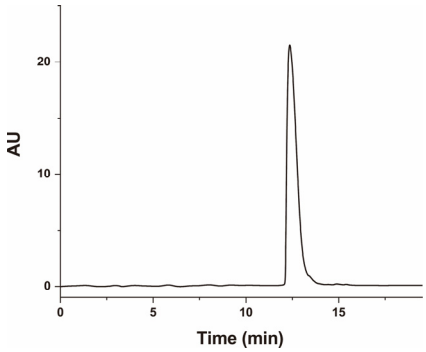
Characterization of the peptides and KRas-C3WC

GKS-MV, $C_{46}H_{68}N_{12}O_{17}S^{2+}$, Exact Mass = 1092.45, found $[GKS\text{-}MV]^{+}$ 1091.5, $[GKS\text{-}MV]^{2+}$ 546.2. The purified peptide was directly used for reaction with DBHDA and analytical HPLC spectrum was not recorded.

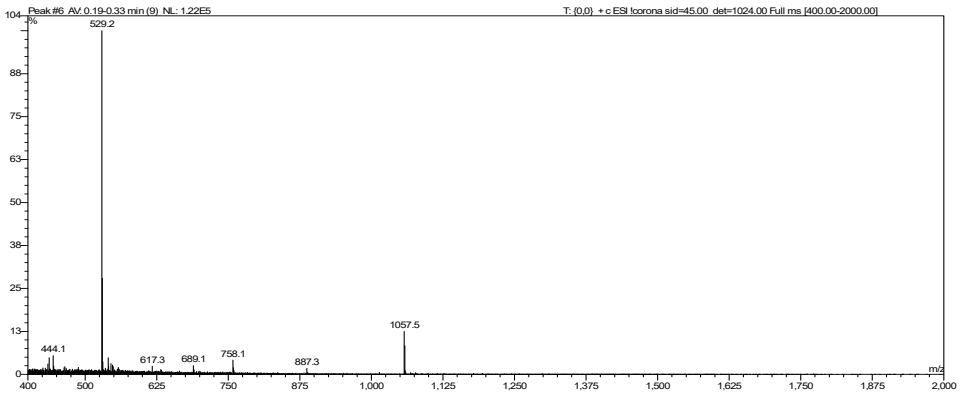


ESI-MS spectrum of **GKS-MV**

Dha-MV, $C_{46}H_{66}N_{12}O_{17}^{2+}$, Exact Mass = 1058.47, found $[Dha\text{-}MV]^{+}$ 1057.5, $[Dha\text{-}MV]^{2+}$ 529.2. Analytical HPLC spectrum, gradient: 5%-12.5% B, 15 min.



Analytical HPLC spectrum of **Dha-MV**



ESI-MS spectrum of **Dha-MV**

KRas-C3WC, Mw=20559.19. Sequence:

10 20 30 40 50 60

MTEYKLVVVG AGGVGKSALT IQLIQNHFVD EYDPTIEDSY RKQVVIDGET **C**LLDILDTAG

70 80 90 100 110 120

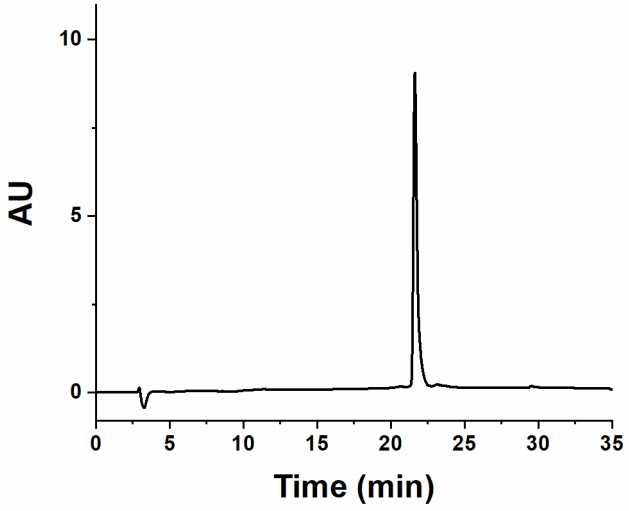
QEEYSAMRDQ YMRTGEGFL**C** VFAINNTKSF EDIHHYREQI KRVKDSEDVP MVLVGNK**C**DL

130 140 150 160 170 180

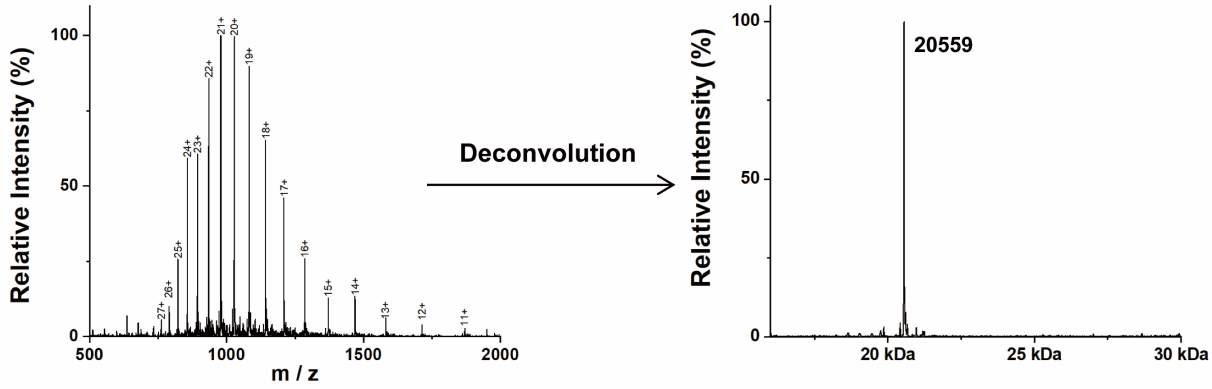
PSRTVDTKQA QDLARSYGIP FIETSAKTRQ GVDDAFYTLV REIRKHKEKM SKDG**C**GSGWG

SG**C**

Analytical HPLC spectrum, gradient: 20%-80% B, 30 min.



Analytical HPLC spectrum of **KRas-C3WC**



ESI-MS and deconvolved spectrum of **KRas-C3WC**

General procedures for CB[8]-facilitated Michael addition

Dha-MV and Cys-containing peptides were solubilized in water to 2 mM. TCEP was solubilized in water to 1.25 mM. CB[8] was solubilized in water to 0.25 mM. TCEP was used to prohibit disulfide bond formation. A typical formula for the reaction was showed in the following table.

	2 mM **Dha-MV**	2 mM **C3W**	1.25 mM TCEP	0.25 mM CB[8]	H_2O	10 mM PB buffer
Volume (μL)	2.5	2.5	1	40	0	54

Table S1. A typical formula for the CB[8]-facilitated reaction.

We have checked the pH for the reaction mixture to be 7.0-7.2, close to the pH of 10 mM PB buffer (pH = 7.2). The reaction was conducted in Eppendorf tube (size = 600 μL) at 37 °C and 300 rpm. The reaction was terminated by quick-freezing in liquid nitrogen and stored at -80 °C until HPLC analysis.

Supplementary Figures

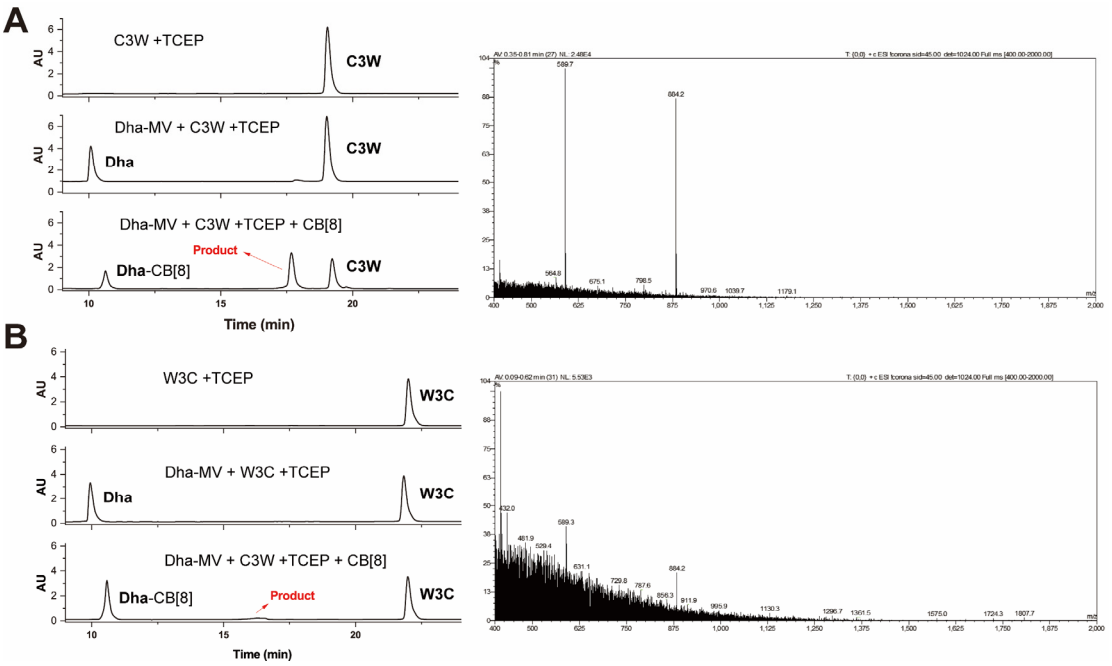


Figure S1. Left panel: Representative HPLC analysis of the reaction mixture of 50 μM (A) **C3W** or (B) **W3C** with 12.5 μM TCEP (Top), and 50 μM **Dha-MV** (middle), and 100 μM CB[8] (bottom) at 37 °C and 300 rpm for 2 h. HPLC gradient, 5%-35% B, 30 min. Note that **Dha-MV** could bind with CB[8] under HPLC solutions and result in the shift of the peak for **Dha-MV**.

Right panel: ESI-MS of the respective HPLC peak for the product. Chemical formula of the product $C_{74}H_{105}N_{21}O_{28}S^{2+}$, Exact Mass =1767.71. ESI-MS, found $[Product]^{2+}$ 884.2, $[Product]^{3+}$ 589.7 for **C3W** + **Dha-MV**. ESI-MS, found $[Product]^{2+}$ 884.2, $[Product]^{3+}$ 589.3 for **W3C** + **Dha-MV**.

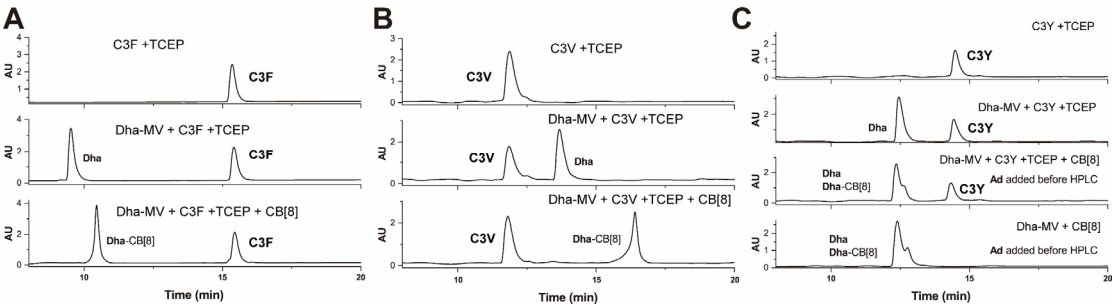


Figure S2. Representative HPLC analysis of the reaction of **Dha-MV** with (A) **CFG**, (B) **C3V** and (C) **C3Y** in the presence or absence of CB[8] for 2 h. No significant product peak could be found. The HPLC gradients were 5%-35% B, 30 min; 5%-11% B, 20 min; 5%-13% B, 20 min, respectively.

Note: (1) Peptide **C3V** showed very low absorbance at 215 nm and 254 nm (wavelength of the UV lamp of HPLC). 250 μM **C3V** was used in the experiment to get visible and comparable HPLC peak.

(2) Peptide **C3Y** and **Dha-MV** had approximative retention time (Figure S2C, 2 nd panel). After the addition of CB[8], the peak of **Dha**-CB[8] would overlap with the peak of **C3Y** (data not shown). We added excess amantadine (Ad) to the reaction mixture before HPLC analysis to prohibit the interaction between CB[8] and **Dha-MV**.

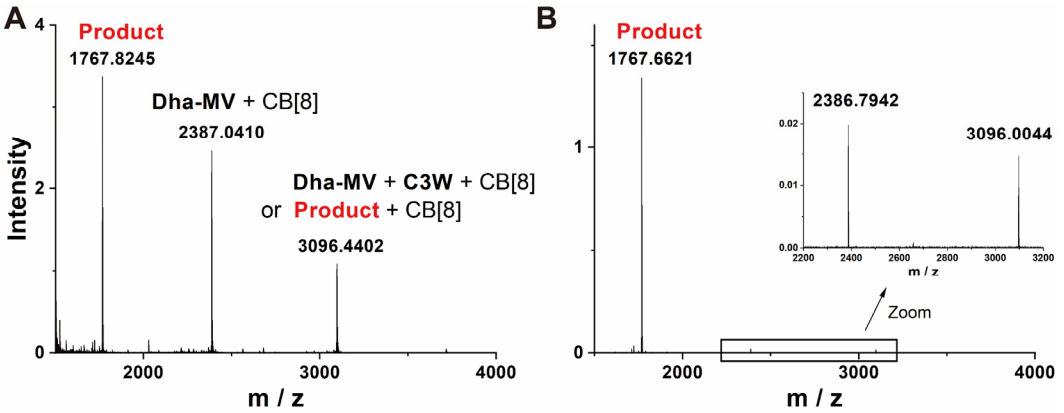


Figure S3. (A) The reaction mixture of 50 μM **Dha-MV**, 50 μM **C3W**, 100 μM CB[8] and 12.5 μM TCEP at 37 °C, 300 rpm for 2 h was directly used for MALDI-TOF-MS. (B) 1 mM amantadine was added to the above 2 h reaction mixture and then MALDI-TOF-MS was performed.

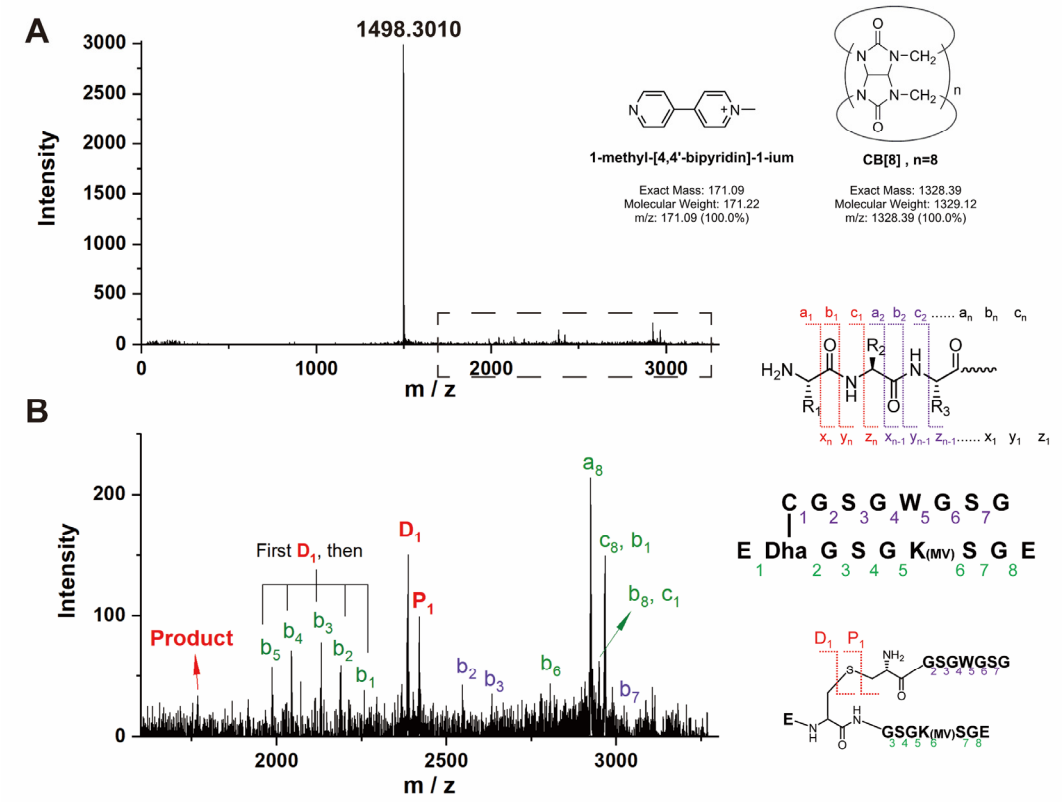


Figure S4. The peak 3096.4402 in Figure S3A was further assayed by MALDI-TOF-MS/MS.

(A) The full spectrum of the MS/MS result. The peak 1498.3010 should be the complex of 1-methyl-[4,4'-bipyridin]-1-ium (**MBP**) with CB[8]. The **MBP** ion with m/z = 171 was obseverd in the MS/MS experiments below (Figure S5 and S6, without the presence of CB[8]).

(B) Zoom of the MS/MS spectrum from 1600-3300. Except the possible peak for **Product** (Exact Mass=1767.71, found m/z = 1766), all the other peaks labeled were in complex with CB[8]. The peak **D1** could be the unreacted **Dha-MV** or the fragment ion from the product. The peak **P1**, whose m/z = **D1**+32, could not be derived from the unreacted **Dha-MV**, further confirmed the existence of product-CB[8] complex in the peak 3096.4402 in Figure S3A. The **D1**/**P1** pattern in MS/MS was also observed in experiments below (Figure S5 and S6, without the presence of CB[8]).

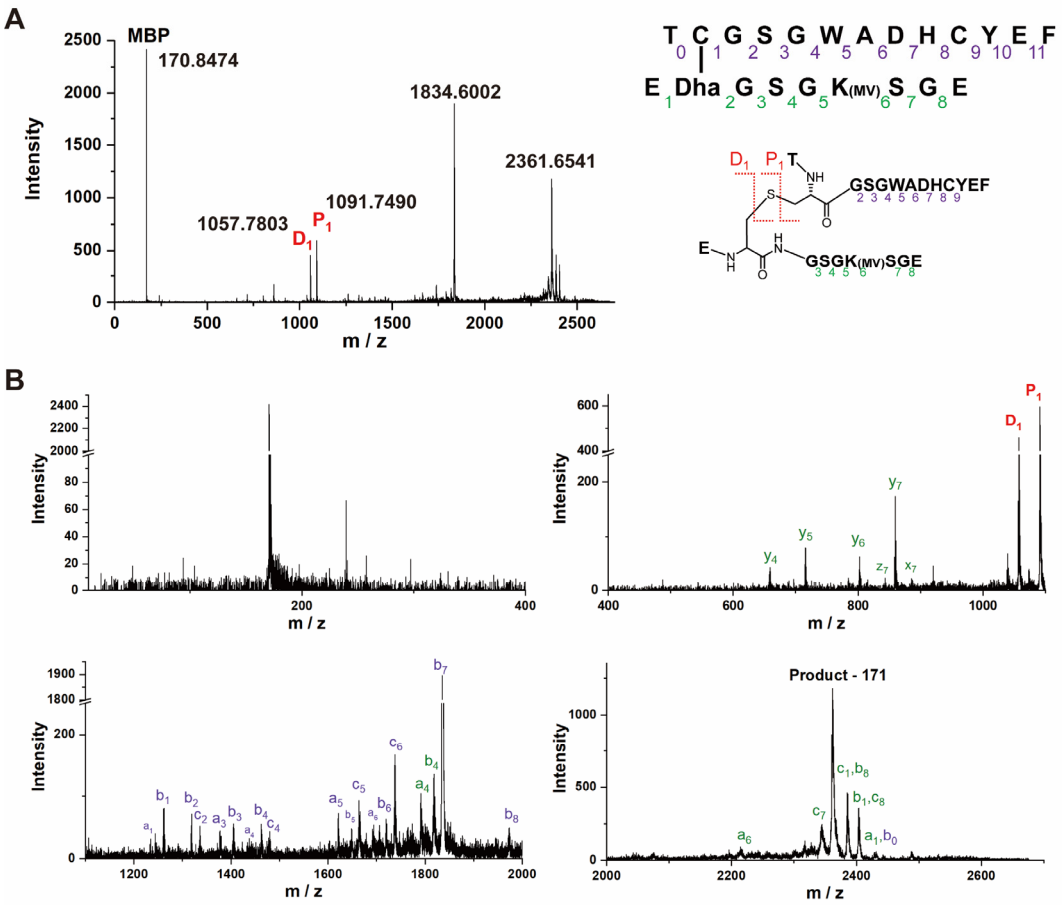


Figure S5. MALDI-TOF-MS/MS analysis of the 17.2-min peak for **TC3XC** reaction in Figure 4A.
(A) The full MS/MS spectrum. Note the m/z = 170.8474 corresponded to the above mentioned **MBP**. The m/z = 2361.6541 corresponded to **Product** subtracts **MBP**.
(B) Zoom of the MS/MS spectrum and the peak analysis. Note that almost all the labelled fragment ions contained MV, suggesting that ions with MV would be easier to analyze by mass spectrometry.

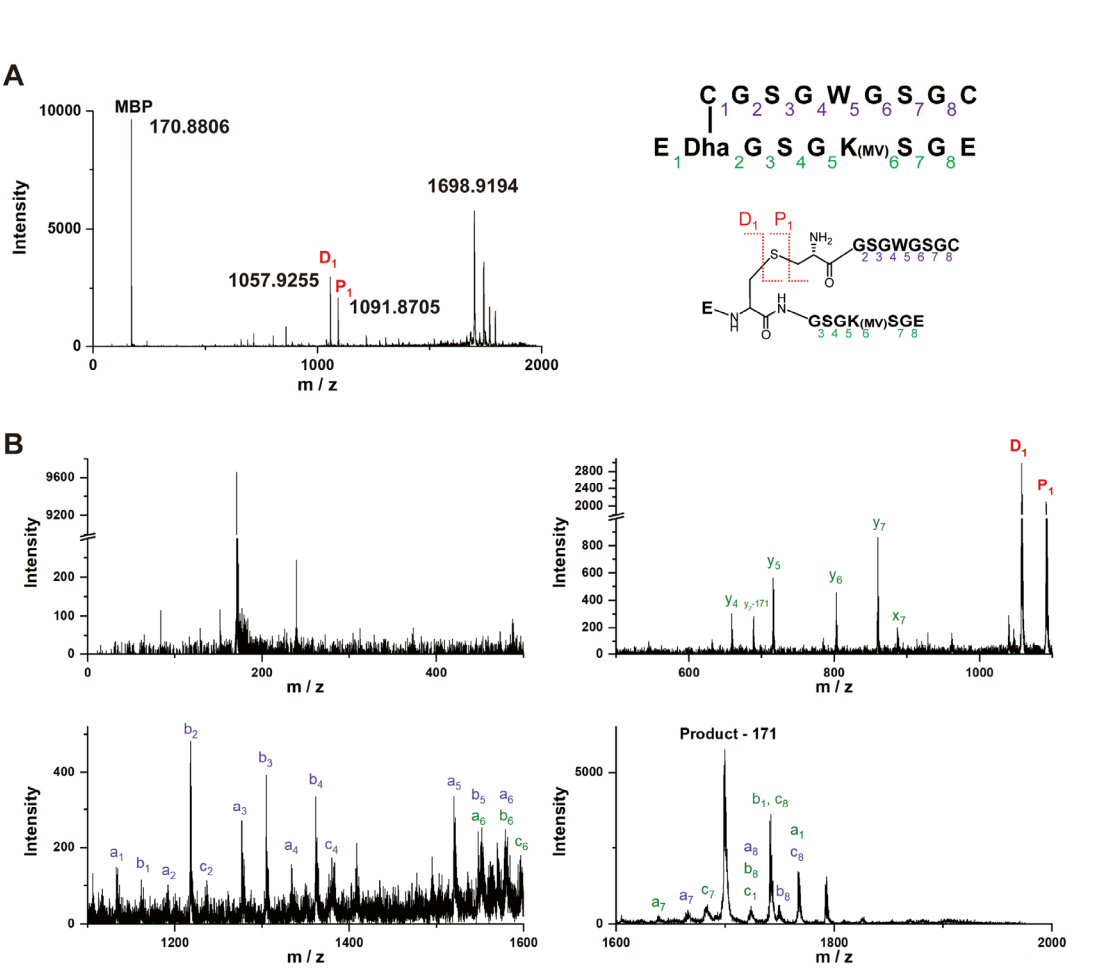


Figure S6. MALDI-TOF-MS/MS analysis of the 21.2-min HPLC peak for **C3WC** reaction in Figure 4A. (A) The full MS/MS spectrum. Note the peaks for **MBP**, **D_1**, **P_1** **and** **Product** subtracts **MBP**.
(B) Zoom of the MS/MS spectrum and the peak analysis. For the two products of **Dha-MV** reacted with **C3WC**, only the a/x fragment ions derived from the chain of **C3WC** (purple **1-8**) will show much difference in MS/MS results which were successfully found and labelled.

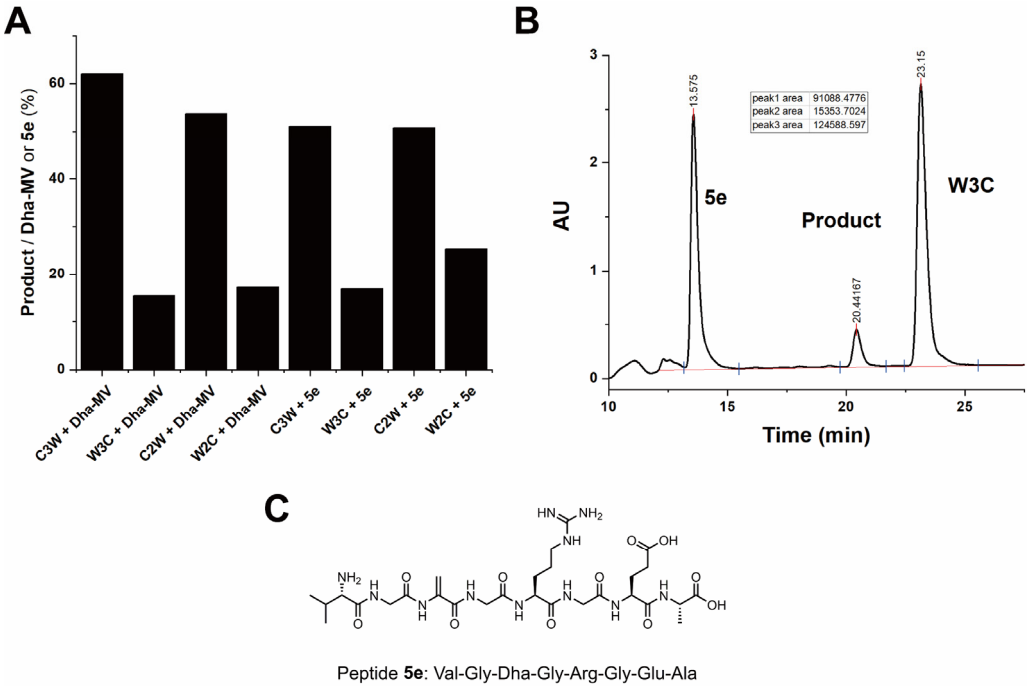


Figure S7. (A) 400 μM **C3W**, **C2W**, **W3C** or **W2C** and 400 μM **Dha-MV** or peptide **5e** were reacted in 50 mM PB buffer (pH=8.5) with 100 μM TCEP at 37 °C, 300 rpm for 2 h. Please note that the Y-axis is not conversion (%) but the ratio of **Product** / **Dha-MV** or **5e** (%). For the CB[8]-facilitated reaction of **C3W** (Figure 3A), this ratio is greater than 100%. (B) Representative HPLC analysis of the 2 h reaction mixture of 400 μM **W3C** and 400 μM **5e**. The peaks were integrated and the ratio of **Product** / **5e** could be calculated. (C) Molecular structure of peptide **5e**[5].

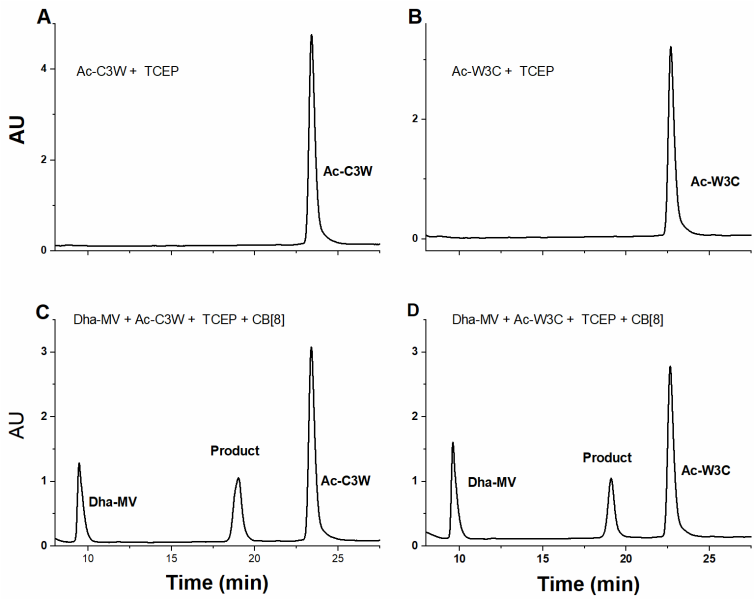


Figure S8. HPLC analysis of the reaction mixture of 50 μM (A) **Ac-C3W** or (B) **Ac-W3C** with 12.5 μM TCEP in pH=7.2 PB buffer, at 37 °C and 300 rpm for 2 h. HPLC analysis of the reaction mixture of 50 μM **Dha-MV**, 100 μM CB[8], 12.5 μM TCEP and 50 μM (C) **Ac-C3W** or (D) **Ac-W3C** in pH=7.2 PB buffer, at 37 °C and 300 rpm for 2 h.

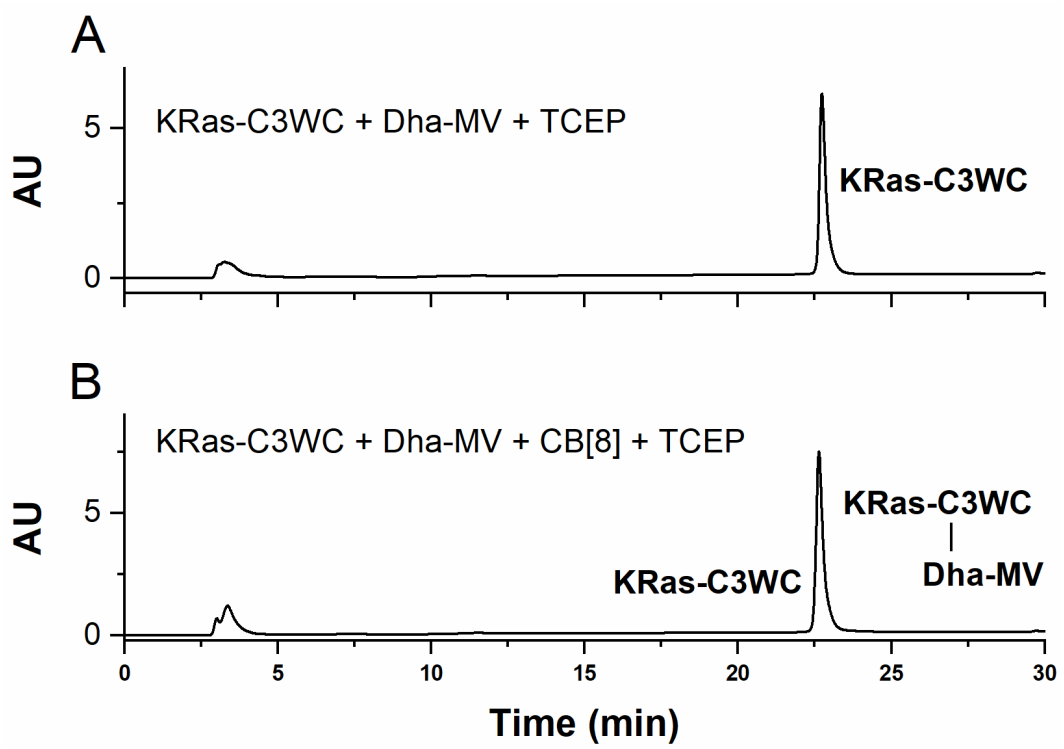


Figure S9. 50 μM **Dha-MV**, 17 μM **KRas-C3WC** and 12.5 μM TCEP were reacted in the (A) absence or (B) presence of 100 μM CB[8] at 37 °C, 300 rpm in pH=7.2 PB buffer for 4.5 h, and then HPLC analysis was performed. The peaks for the proteins were collected and used for subsequent MS analysis.

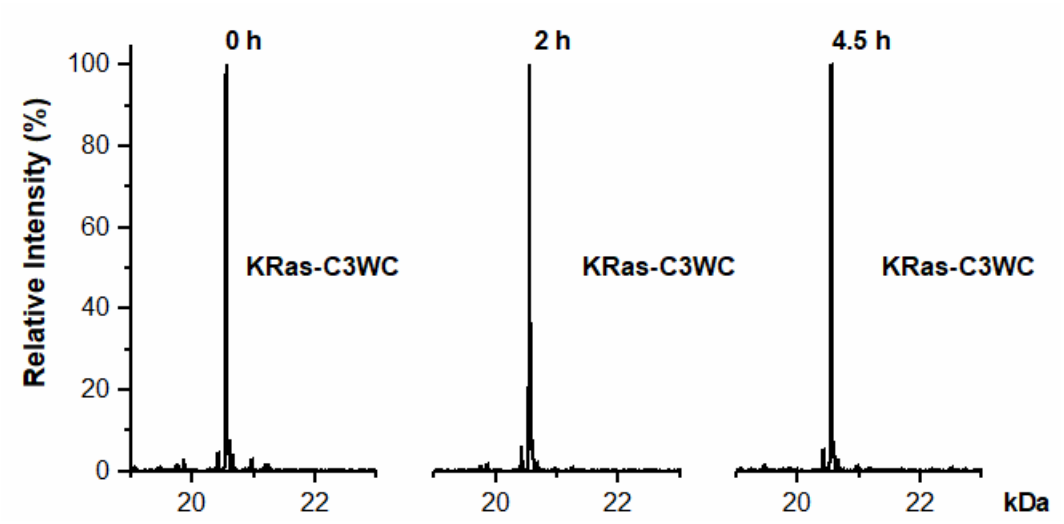


Figure S10. 50 μM **Dha-MV**, 17 μM **KRas-C3WC**, and 12.5 μM TCEP were reacted at 37 °C, 300 rpm in pH=7.2 PB buffer for different time and deconvolved MS spectra were recorded after HPLC.

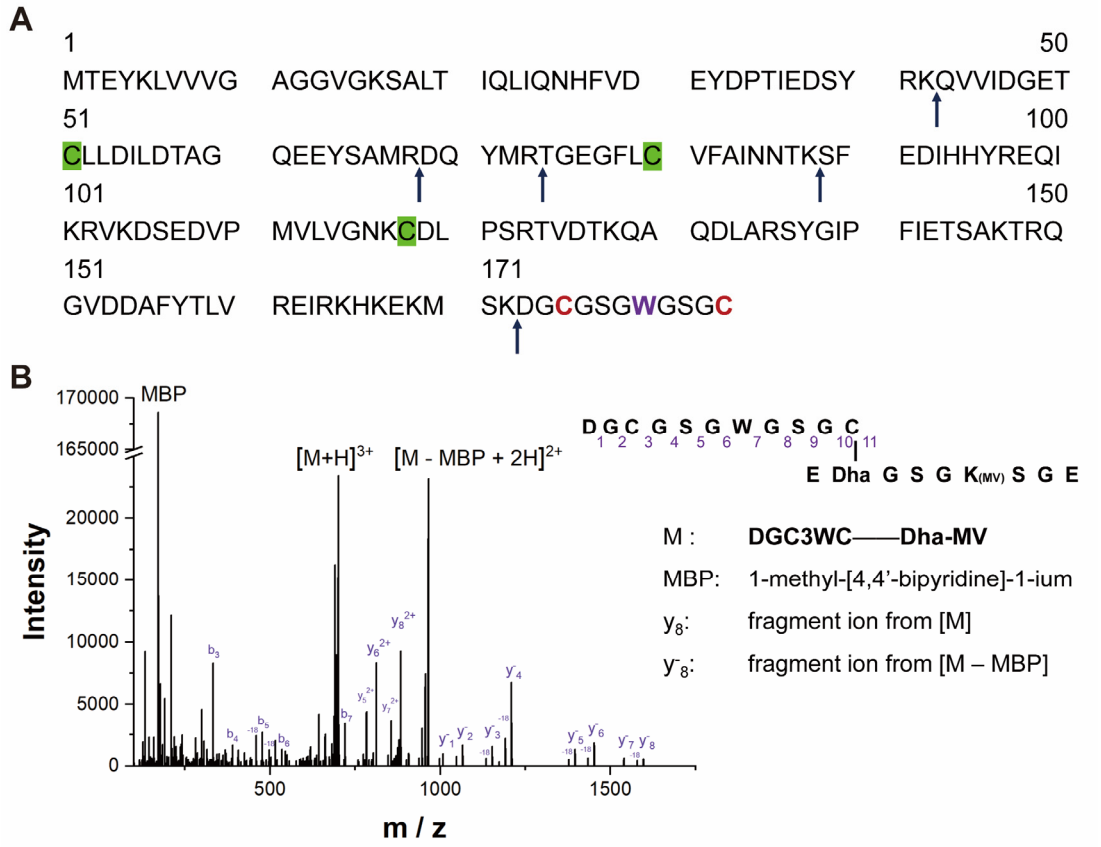


Figure S11. (A) Sequence of **KRas-C3WC** and possible trypsin digestion sites. (B) 100 μM CB[8], 50 μM **Dha-MV**, 17 μM **KRas-C3WC** and 12.5 μM TCEP were reacted for 4.5 h, and then HPLC analysis was performed. The peak for the proteins was collected and used for subsequent digestion and LC-MS/MS analysis. The MS/MS peaks for reaction at the C-terminal cysteine of **KRas-C3WC** were identified. No peaks could be found for reaction at other cysteine residues.

Note: different MS instruments were used in MALDI-MS/MS and LC-MS/MS analysis, which caused the different pattern of fragment ions (compared to Figure S4-S6).

Reference

(1) Van Emon, J.;Hammock, B.;Seiber, J. N. Enzyme-linked immunosorbent assay for paraquat and its application to exposure analysis. *Anal. Chem.* **1986**, *58*, 1866-1873.

(2) Li, G.;Yang, W. Y.;Zhao, Y. F.;Chen, Y. X.;Hong, L.;Li, Y. M. Differential Modulation of the Aggregation of N-Terminal Truncated A beta using Cucurbiturils. *Chem. Eur. J.* **2018**, *24*, 13647-13653.

(3) Chen, H.;Zhang, Y.;Li, Q. Q.;Zhao, Y. F.;Chen, Y. X.;Li, Y. M. De Novo Design To Synthesize Lanthipeptides Involving Cascade Cysteine Reactions: SapB Synthesis as an Example. *J. Org. Chem.* **2018**, *83*, 7528-7533.

(4) Chen, Y. X.;Koch, S.;Uhlenbrock, K.;Weise, K.;Das, D.;Gremer, L.;Brunsveld, L.;Wittinghofer, A.;Winter, R.;Triola, G.;Waldmann, H. Synthesis of the Rheb and K-Ras4B GTPases. *Angew. Chem. Int. Ed.* **2010**, *49*, 6090-6095.

(5) He, P. Y.;Chen, H.;Hu, H. G.;Hu, J. J.;Lim, Y. J.;Li, Y. M. Late-stage peptide and protein modifications through phospha-Michael addition reaction. *Chem. Commun.* **2020**, *56*, 12632-12635.