Supporting Information

One-pot hydrazide-based native chemical ligation for efficient chemical synthesis and structure determination of toxin Mambalgin-1

Man Pan,[a,b‡] Yao He,[a,‡] Ming Wen,[a] Fangming Wu,[a] Demeng Sun,[a,b] Sijian Li,[c] Longhua Zhang,[a,*] Yiming Li,[b,c,*] Changlin Tian[a,*]

[a] *School of Life Sciences, University of Science and Technology of China, and High Magnetic Field Laboratory, Chinese Academy of Sciences, Hefei 230026, China. Email: zlhustc@ustc.edu.cn & cltian@ustc.edu.cn*

[b] *Tsinghua-Peking Center for Life Sciences, Tsinghua University, Beijing 100084, China.*

[c] *School of Medical Engineering, Hefei University of Technology, Hefei, Anhui 230009, China, E-mail: lym2007@mail.ustc.edu.cn*

Experimental Section

a. Preparation of N_3-Cys(Trt)-OH:

Sodium azide (10 g, 154 mmol, 10 eq) was dissolved in distilled water (25 mL), cooled in an ice bath, and dichloromethane (DCM, 40 mL) was added. Triflyl anhydride (5 mL, 30 mmol, 2 eq) was added slowly, the mixture was stirred for 2 h, transferred to a separating funnel, and organic extracts containing triflyl azide were collected and used without further purification.

H-Cys(Trt)-OH (5.46 g, 15 mmol, 1 eq), K_2CO_3 (3 g, 22.5 mmol, 1.5eq) and $CuIISO_4 \bullet 5H_2O$ (37.7 mg, 150 μmol) were dissolved in a mixture of distilled water (50 mL) and methanol (100 mL). Triflyl azide in DCM was then added and the mixture was stirred overnight. Subsequently, organic solvent was removed under reduced pressure and the aqueous slurry was diluted with water (50 mL) and acidified to pH 2 with concentrated HCl. The product was extracted three times with 20 mL ethyl acetate and purified using silica gel column chromatography to give 3.95 g N_3-Cys(Trt)-OH in 66% yield. ESI-MS (negative) 388.8 (observed, M-H) 389.12 (calculated, M). ^{1}H NMR (400 MHz, DMSO-d6): δ (ppm) 2.43 (d, 2H), 3.90 (m, 1H), 7.26-7.28 (m, 3H), 7.32-7.37 (s, 12H).^{13}C NMR (400 MHz, DMSO-d6): δ (ppm) 169.99, 143.99, 129.05, 128.18, 126.94, 66.39, 60.38, 32.99.

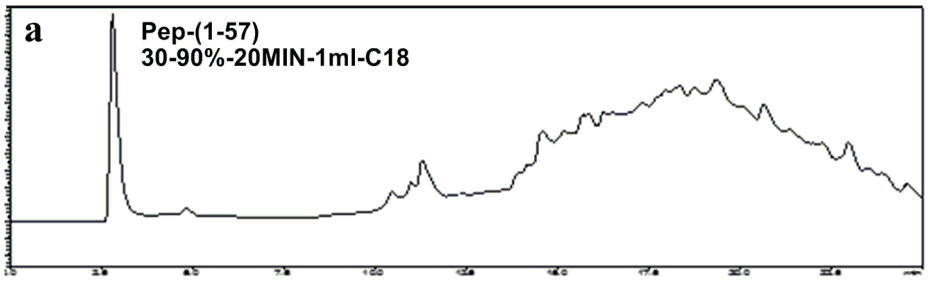


Figure S1: HPLC analysis of final synthetic product of Ma-1 by using direct SPPS. The spectra did not show any major peak of desire product.

b. Synthesis of peptide-hydrazides

Peptides used in this work were obtained from the Fmoc based solid phase peptide synthesis. The reaction vessels were purchased from commercial sources. Specific steps are shown below:

1) Preparation of hydrazine-Trt(2-Cl)resin:

Commercial 2-Cl-(Trt)-Cl resin (2 g, 1.12 mmol reactive Cl) was swollen for 10 min in DMF (5 mL) and cooled under ice bath. A mixture of triethylamine (0.48 mL, 3.36 mmol) and hydrazine hydrate (0.56 g, 11.2mmol) in DMF (1 mL) was added dropwise and the suspension was stirred for 2h

at rt. The resin was filtered, washed with DMF, water, DMF, methanol. After that, the resin was dried for 1h at rt in vacuo.

2) 2-Cl-Trityl resin (0.56 mmol/g 200 m) and hydrazine-Trt(2-Cl)resin (0.35 mmol/g 200 m) was initially swelled with DCM/DMF (1/1) for 10 min before use.

3) The amino acid coupling conditions: a pre-activated solution of 4eq protected amino acids, 3.8 eq of HCTU, 8 eq the Dipea successively dissolved in DMF was poured to the resin. The coupling time should extend from 1 h to 2 h with peptide chain extension, then the resin was washed with DMF, DCM, DMF all 5 times alternately.

4) Coupling efficiency test: Pre-prepared three reagents: 0.5 g ninhydrin was dissolved in 10 mL EtOH, 80% phenol solution. 0.4 mL of 1 mM KCN solution was diluted with piperidine to 20 mL. Transfer a few resin beads to a glass tube and add one drop of each reagent mentioned above. Mix well and heat the mixture to boil for 2 min.

5) The Fmoc deprotection: 20% piperidine in DMF swelled twice for 5 min and 10 min. Then the resin was washed with DMF, DCM, DMF 5 times alternately;

6) Peptides cleavage: After the de-protection of the Fmoc of the final amino acid, cleavage reagent (reagent K: TFA/phenol/water/thioanisole/EDT (82.5/5/5/5/2.5)) was added to the dry resin pre-washed with DCM. After 4 h, the resin was washed with another equal volume of TFA. Combined liquids were concentrated by blowing with Ar2. The crude peptides were obtained by precipitating with cold ether and centrifugation at 5000 rpm for 2 min.

7) Peptide purification: The crude peptides was dissolved in co-solvent of acetonitrile (0.1% TFA) and water (0.1% TFA), analyzed by analytical HPLC and ESI-MS and purified by semi-preparative HPLC and lyophilized immediately.

8) Notice: For 2-Cl-Trityl resin, the first amino acid was coupled with 4 eq protected amino acids, 8 eq dipea in DMF/DCM (1:1) in 4h and the resin was capped with methanol.

c. Racemization of the reduction of N_3-Cys(Trt)-OH in peptides

(1) Synthesis of model peptides:

Peptides H-CRALNKS-$NHNH_2$, H-CRALNKS-$NHNH_2$, and N_3-CRALNKS-$NHNH_2$ were synthesized using Fmoc-based SPPS.

(2) Reduction of N_3-CRALNKS-$NHNH_2$:

N_3-CRALNKS-$NHNH_2$ (2 mg, 2.4 μmol, 1 eq) was reduced with 25 μL TCEP (20 mg dissolved in 400 μL distilled water) in 10 min.

(3) Racemization:

Model peptides H-CRALNKS-$NHNH_2$ and H-CRALNKS-$NHNH_2$ and the reduction product were analyzed by RP-HPLC. Co-injection confirmed that no racemization occurred.

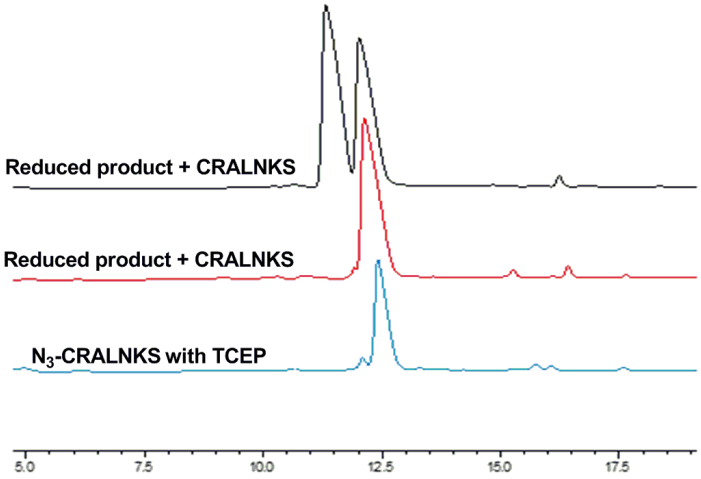


Figure S2: (a) The co-injection of H-CRALNKS-$NHNH_2$, H-CRALNKS-$NHNH_2$ and the reduction product; (b) RP-HPLC monitoring of azide-Cys assisted one-pot ligation.

d. One-pot test of model peptides

(1) Synthesis of model peptides:

Peptides H-C(Acm)RANK-$NHNH_2$, N_3-CRALNKS-$NHNH_2$, and H-CRALNKS-$NHNH_2$ were synthesized using Fmoc-based SPPS.

(1) One-pot ligation test:

N_3-CRALNKS-$NHNH_2$ (4 mg, 4.8 μmol, 1 eq) and H-CRALNKS-OH (3.8 mg, 4.8 μmol, 1 eq) were dissolved in 500 μL ligation buffer (6 M Gn-Cl, 100 mM NaH_2PO_4, pH 3.0) and cooled in an ice bath (-10°C). 20 μL of aqueous 1 M $NaNO_2$ solution (20 μmol, 4 eq) was added dropwise and the reaction was incubated for 25 min. MPAA (16 mg, 100 μmol, 20 eq) was added and the pH was adjusted to 5.0 for 5 min then to 6.3. Completion of the reaction was detected by RP-HPLC, at which point TCEP (7 mg, 25 μmol, 5 eq) was added and the mixture incubated for 10 min. The reaction mixture was combined with the oxidized H-C(Acm)RANK-$NHNH_2$ (3.7 mg, 5.0 μmol, 1.05 eq) which was synthesized as described for N_3-CRALNKS-$NHNH_2$, and similarly monitored by RP-HPLC (Figure S1b). ESI mass spectrometry confirmed that full-length H-

C(Acm)RANKCRALNKSCRALNKS-$NHNH_2$ had been prepared. The N-terminal Acm and C-terminal $NHNH_2$ groups allow for the expansion of this method.

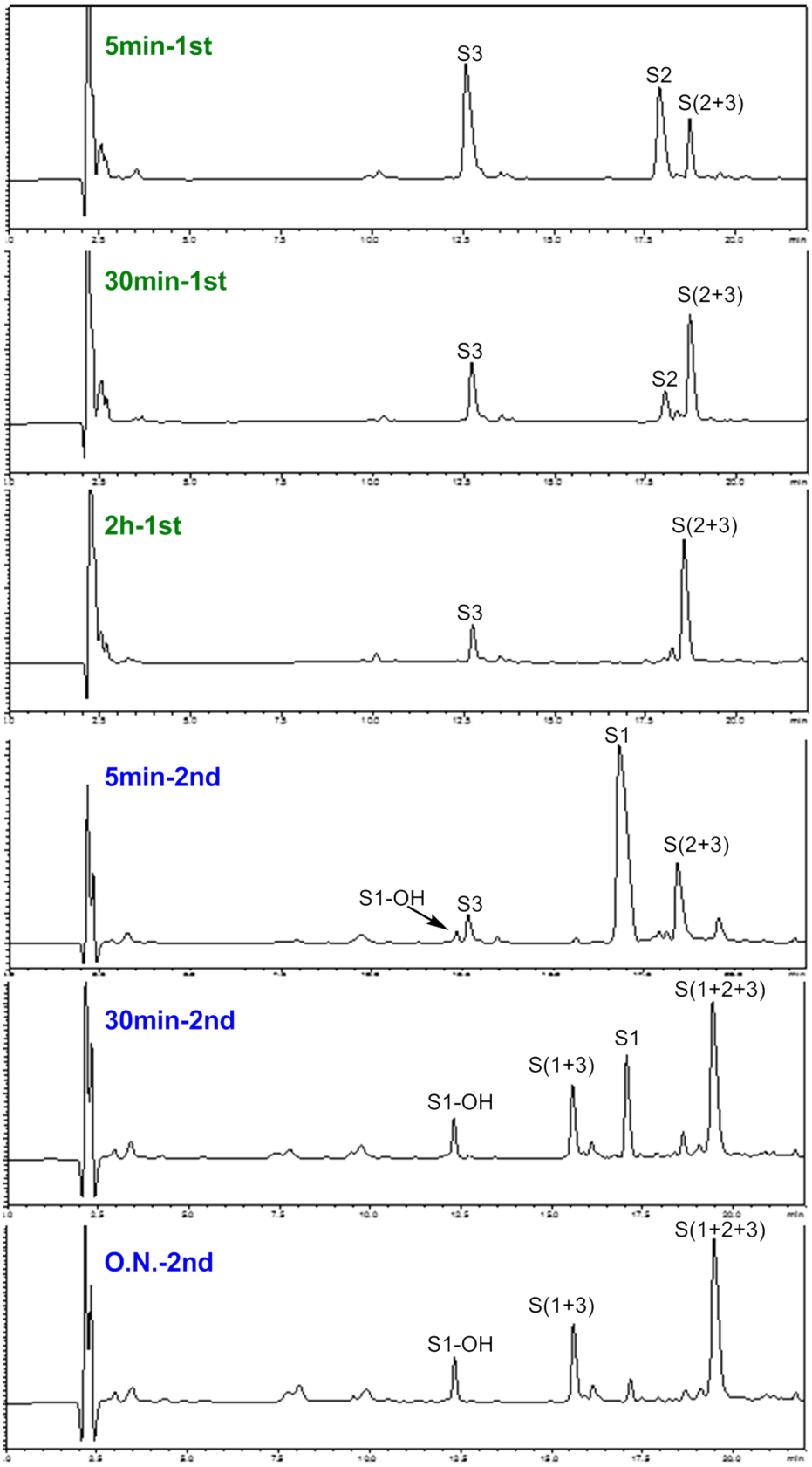


Figure S3: RP-HPLC analysis of model test on azide-switch based one pot ligation. (S1:

intramolecular thioester of H-C(Acm)RANK-$NHNH_2$; S1-OH: H-C(Acm)RANK-OH; S2: intramolecular thioester of H-CRALNKS-$NHNH_2$; S3: H-CRALNKS-$NHNH_2$; S(2+3): H-CRALNKSCRALNKS-$NHNH_2$; S(1+2+3): H-C(Acm)RANKCRALNKSCRALNKS-$NHNH_2$. 1st and 2nd means the first and second ligation.

e. One-pot total synthesis of Mambalgin peptides

(1) Synthesis of model peptides:

H-LKCYQHGKVVTCHRDMKF-$NHNH_2$, H-LKCFQHGKVVTCHRDMKF-$NHNH_2$, N_3-CYHNTGMPFRNLKLILQGCSSS-$NHNH_2$, N_3-CYHNTGMPFRNLKLILQGCSSS-$NHNH_2$ and H-CSETENNKCCSTDRCNK-OH were synthesized using Fmoc-based SPPS

(2) One-pot ligation of Mambalgin-1:

N_3-CYHNTGMPFRNLKLILQGCSSS-$NHNH_2$ (25 mg, 12 μmol, 1 eq) and H-CSETENNKCCSTDRCNK-OH (24 mg, 12 μmol, 1 eq) were dissolved in 1 mL ligation buffer (6 M Gn-Cl，100 mM NaH_2PO_4, pH 3.0) and cooled in an ice bath (-10°C). 48 μL of 1M aqueous $NaNO_2$ solution (48 μmol, 4 eq) was added dropwise and the reaction was incubated for 25 min. MPAA (40.36 mg, 240 μmol, 20 eq) was added and the pH was adjusted to 5.0 for 5 min then to 6.3. Completion was again detected by RP-HPLC, and TCEP (16.8 mg, 60 μmol, 5 eq) was added and the reaction incubated for 10 min. The reaction was combined with oxidized H-LKCYQHGKVVTCHRDMKF-$NHNH_2$ or H-LKCFQHGKVVTCHRDMKF-$NHNH_2$ (28 mg, 12.6 μmol, 1.05 eq) which were prepared the same as N_3-CRALNKS-$NHNH_2$, and similarly monitored by RP-HPLC. This efficient ligation produced full-length mambalgin-1 and mambalgin-2 in good yield (37%) following HPLC purification.

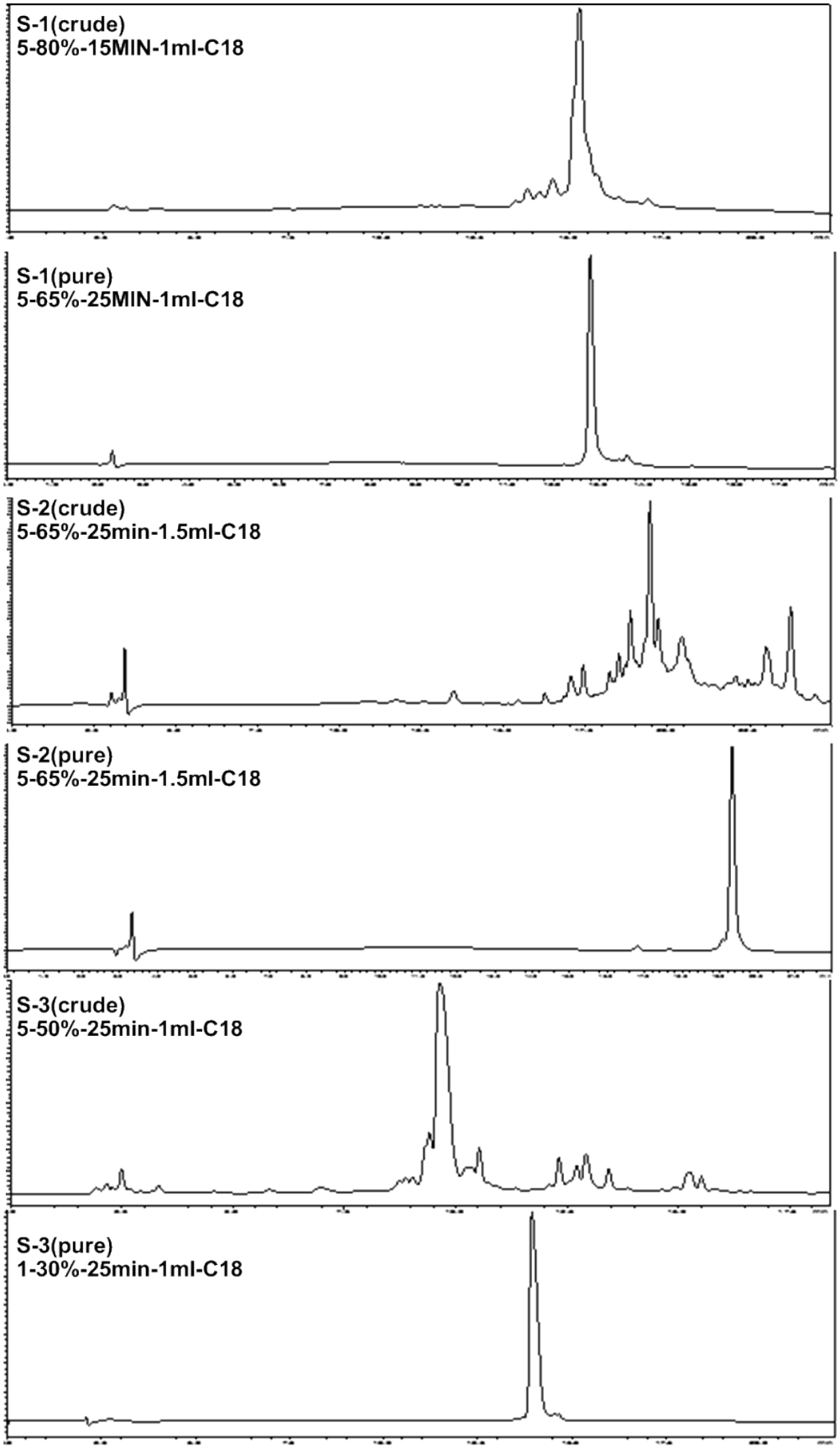


Figure S4: RP-HPLC analysis of the synthesis of MA-1[1-18]-$NHNH_2$ (S-1), MA-1[19-40]-$NHNH_2$ (S-2) and MA-1[41-57] (S-3).

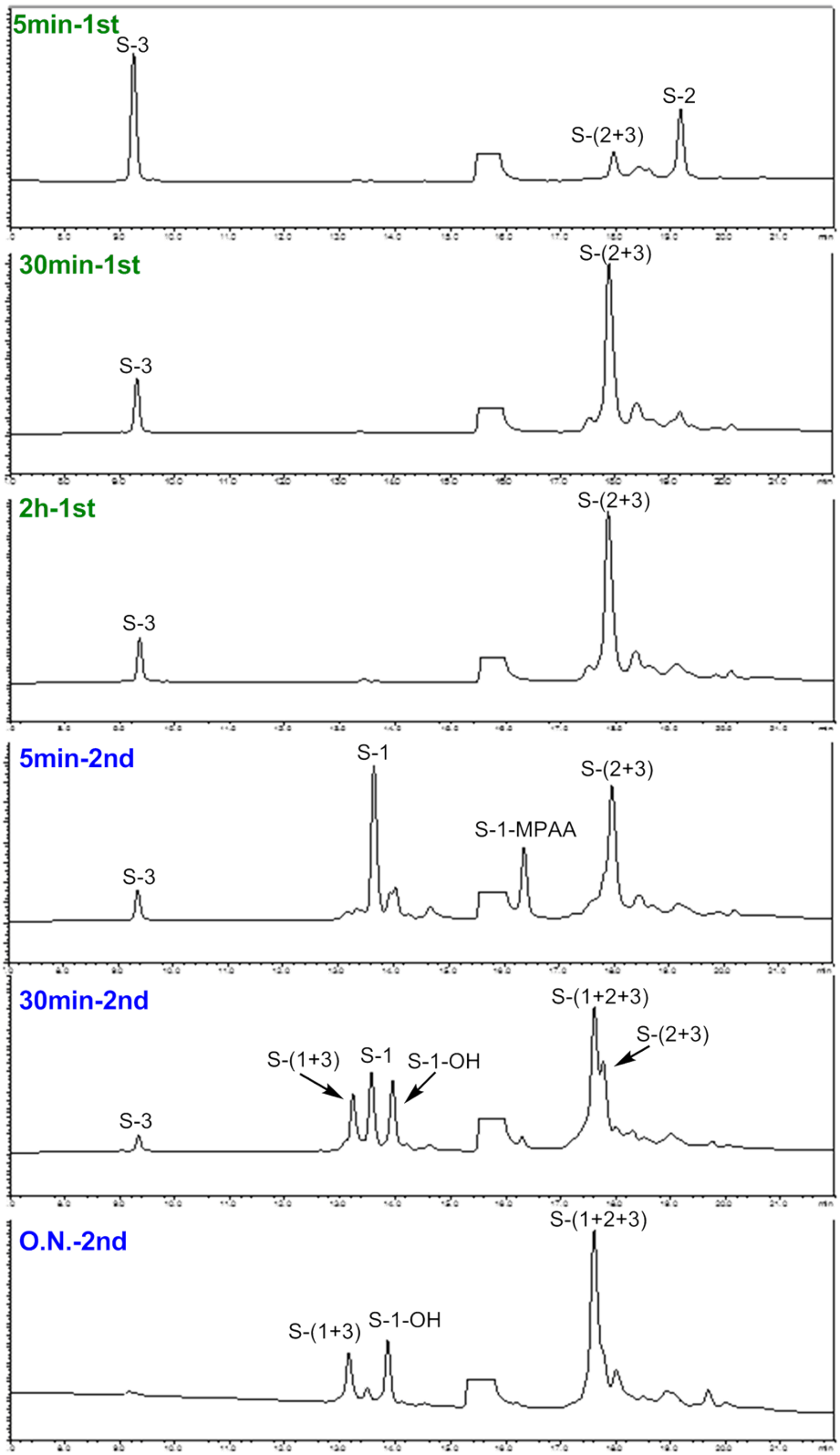


Figure S5: RP-HPLC analysis of azide-switch based one pot ligation of Mambalgin-1. (S-1: intramolecular thioester of MA-1[1-18]; S-1-MPAA: MPAA thioester of MA-1[1-18]; S-1-OH: hydrolysis product of MA-1[1-18]; S-2: intramolecular thioester of MA-1[19-40]; S-3: MA-1[41-57]; S-(2+3): MA-1[19-57]; S-(1+2+3): final ligation product of MA-1[1-57]. 1st and 2nd means the first ligation and second ligation.

f. Protein folding

We next optimized in vitro folding of the polypeptide chain for the formation of the four disulfide bridges. Folding process were found more efficient in oxidation solution containing 10 μM peptide with 1000 μM GSH and 100 μM GSSH, pH =7.8, 25℃ and stirred under 60 r/min over night. Under these conditions, the reaction reached equilibrium in 24 h with very little precipitation. Then, the pH of the folding buffer was adjusted to 2 and lyophilized immediately. The crude folding product was dissolved in water (0.1% TFA) and the folded Mambalgin-1 was purified by reverse phase HPLC obtained 55% isolated yield.

g. CD Spectroscopy

CD spectra were obtained using a Pistar π-180 CD spectrometer (Applied Photophysics Ltd.) with scanning between 200 nm and 260 nm in a 0.2 cm path length cell at 24°C. Each protein sample was dissolved to a final concentration of 0.3 mg/mL in 10 mM PBS (pH 7.4). The results were averaged from three scans and plotted as mean residue ellipticity [θ] (mdeg cm^2 $dmol^{-1}$).

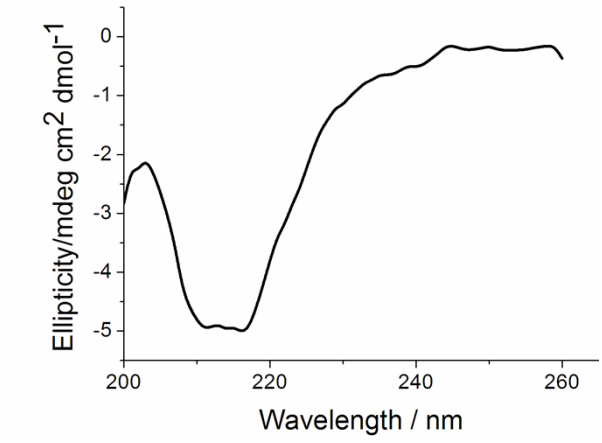


Figure S6: CD analysis of synthetic Mambalgin-1

h. NMR structure determination

The solution NMR sample of Mambalgin-1 (final concentration of ~3.0 mM) was prepared in 50 mM PBS buffer (pH 6.5) containing 10% D_2O (Cambridge Isotope Laboratories). A set of homonuclear DQF-COSY, TOCSY and NOESY spectra were recorded on a Varian 700 MHz spectrometer equipped with a cryo-probe. The TOCSY spectra were recorded at different temperatures

(298 K and 310 K) with different mixing times (20 ms, 50 ms and 80 ms) to distinguish ambiguous signals. The DQF-COSY and NOESY spectra were recorded at 310 K and the mixing time of the NOESY experiment was 200 msec. Spectra were processed using NMRPipe (Delaglio et al 1995) and analyzed using SPARKY (Goddard & Kneller 2006) The sequential assignment procedures developed by Wüthrich (Wuthrich 1986) were used for resonance assignments based on TOCSY and NOESY spectra. Distance restraints were derived from NOE peak intensities and used for structure calculations using Xplor-NIH (Schwieters et al 2003). A total of 740 NOE restraints were applied. Simulated annealing was processed with 30,000 steps during cooling and 200 structures were generated. Structure convergences were analyzed for the 20 lowest energy structures. The program MOLMOL (Koradi et al 1996) was used to assess structure quality, and PROCHECK-NMR (Laskowski et al 1996) was used for structure ensemble statistical analysis and Ramachandran diagram generation. PyMol (DeLano 2010) was used to visualize the peptide structures and surfaces.

Table 1. Structural statistics of the final 20 conformers of Mambalgin-1.

	Manba
Number of Distance Constrains	
NOE distance constraints	740
intraresidue ($i-j=0$)	159
sequential ($i-j=1$)	238
medium-range ($2 \leq i-j \leq 4$)	102
long-range ($i-j \geq 5$)	241
Rmsd for experimental restraints	
NOE distance constraints	0.007±0.001
Rmsd from idealized covalent geometry	
Bonds (Å)	0.001±0.000
Angles (°)	0.367±0.002
Impropers (°)	0.218±0.005
Average RMSD of Atomic Coordinates (Å)	Backbone/Heavy atoms
Residues in secondary structure regions[a]	0.306/0.848
All residues	1.216/2.148
Ramachandran plot analysis (%)	
most favorable region	55.0
additional allowed regions	38.8
generously allowed regions	4.8
disallowed regions	1.4

[a] Residues in the secondary structure regions corresponds to residues 3-4, 9-10, 18-22, 34-38, and

47-50).

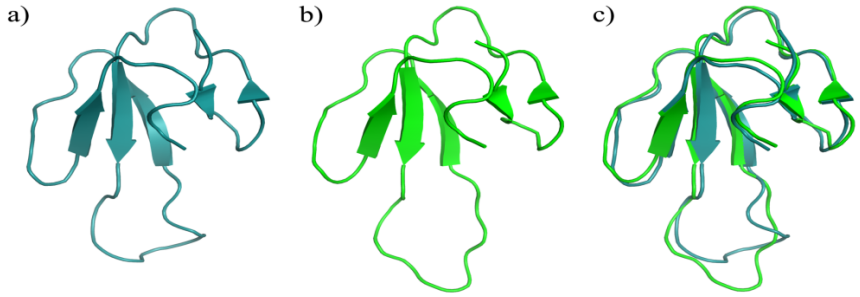


Figure S7. Ribbon representation of Mambalgin-1 (a, PDB: 2MJY) and Mambalgin-2 (b, PDB: 2MFA (Schroeder et al 2014)). The backbone alignment of shows the high similarity: the RMSD is 0.51 for backbone atoms in secondary structure regions (residue 3-4, 9-10, 18-22, 34-38 and 47-50) and 1.44 for the backbone atoms of all other residues.

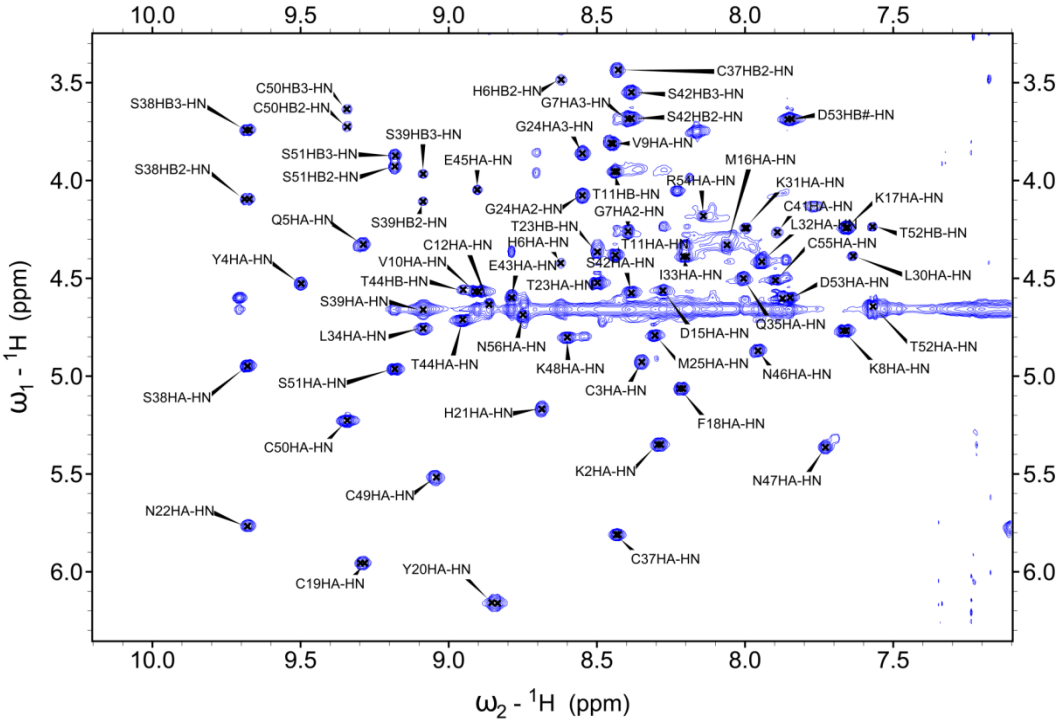


Figure S8: HN-HA region of the synthetic Mamba-1 TOCSY spectra.

i. Electrophysiology of hASIC1a and Mambalgin-1 inhibition

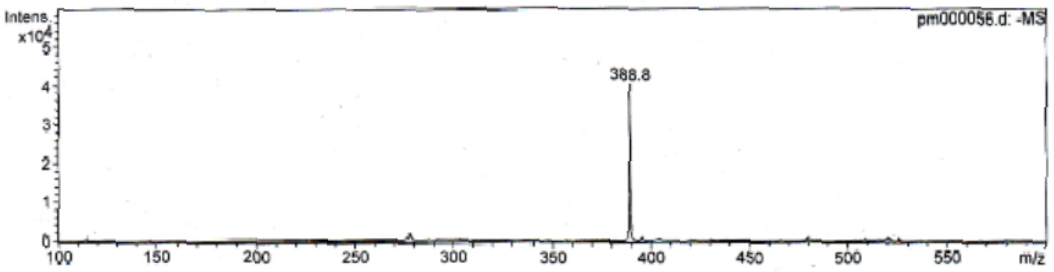
Recording of recombinant human ASIC1a channel current was performed in Chinese hamster ovary (CHO-K1) cells transfected with the ASIC1a gene. The full-length human ASIC1a coding sequence was amplified and inserted into the pcDNA3.1/Zeo (+) vector. The complete ASIC1a cDNA was sequenced to verify that nonspecific mutations were not introduced. Cotransfection of human

ASIC1a and GFP (also in the pcDNA3.1 vector) were performed using PolyFect transfection reagent. Electrophysiological experiments were conducted 24-48 hr after cells were transferred to acid-washed cover slips pre-coated with 0.05% poly-L-lysine for experiments. For whole-cell recordings, the standard bath solution contained the following (in mM): 150 NaCl, 4 KCl, 2.0 $CaCl_2$, 1.0 $MgCl_2$, 10 HEPES (pH 7.4; 300 ~ 312 mOsm). The pipette solution contained the following (in mM): 10 NaCl, 140 KCl, 2.0 $CaCl_2$, 1.0 $MgCl_2$, 10 HEPES, 5 EGTA (pH 7.33; 297 mOsm). All chemicals were purchased from Sigma except mambalgin-1. All the experiments were carried out using the whole-cell patch-clamp technique. Patch pipettes were pulled from thick-walled borosilicated glass with filament (Sutter Instruments, USA) using a four-stage P-1000 puller (Sutter Instruments, USA) with a resistance of 2-5 MΩ when filled with intracellular solution. Recordings were performed using an EPC-10 amplifier (HEKA Electronic) at room temperature (22-24°C) and PatchMaster software was used for data acquisition. All cells were held at -70 mV. The access resistance and capacitance were monitored throughout the recording and if they changed by more than 10%, the results were discarded. Acid currents or toxin effects were elicited through a Y-tube perfusion system that allowed local and rapid changes of solutions at 2 min intervals to allow for a complete recovery of the channels from desensitization. Toxins were applied before the pH drop and during the pH application to prevent non-specific adsorption. Signals were analyzed off-line using Clampfit in the pClamp 10 software. Concentration-response curves were fitted using the Hill equation: $I = I_{min} + (I_{max} - I_{min})(C^b / (C^b + IC_{50}^b))$ where I is the amplitude of the relative current, C is the toxin concentration, and IC_{50} is the toxin concentration that half-maximally inhibits the current.

3. Spectra

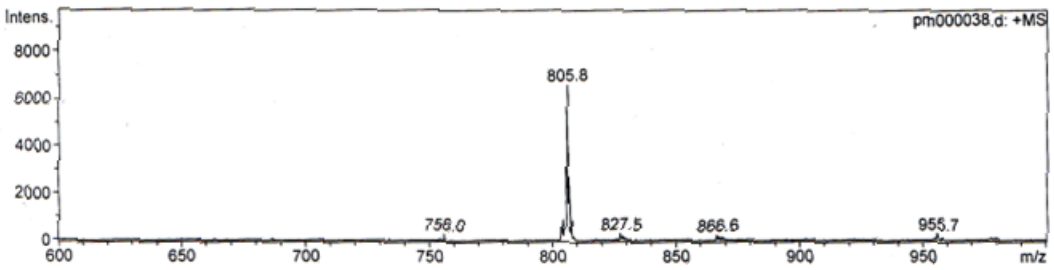
a. Mass spectra

1) N_3-Cys(Trt)-OH



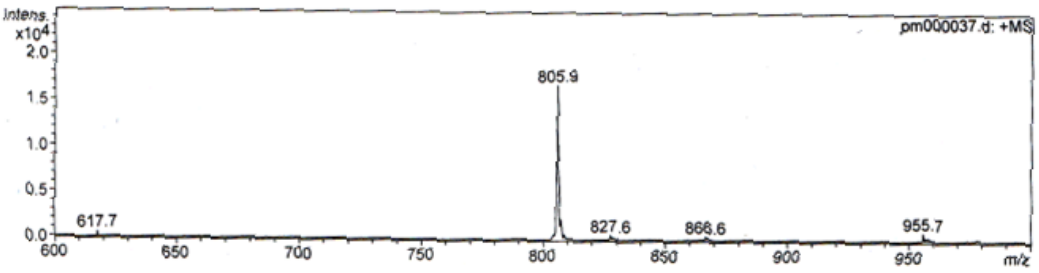
ESI-MS (negative) 388.8(observed, M-H) 389.1 (calculated, M).

2) H-CRALNKS-$NHNH_2$



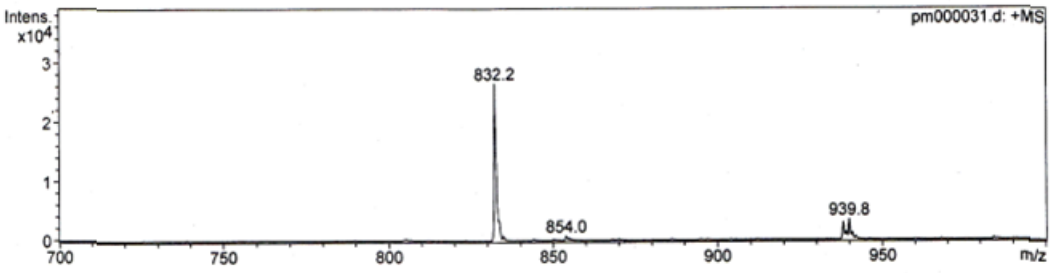
ESI-MS (positive) 805.8(observed, M+H) 804.96 (calculated, M).

3) H-cRALNKS-$NHNH_2$



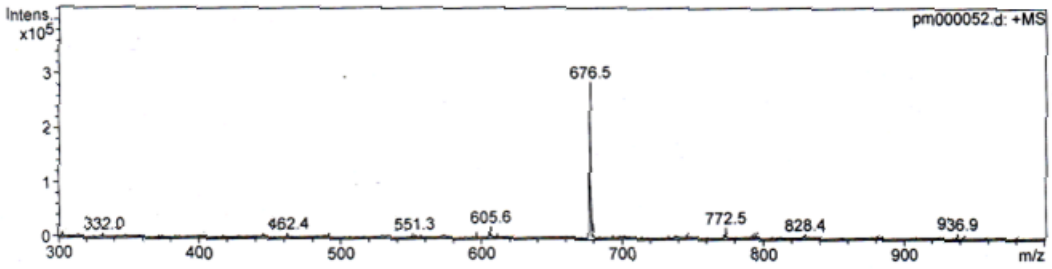
ESI-MS (positive) 805.9(observed, M+H) 804.96 (calculated, M).

4) N_3-CRALNKS-$NHNH_2$



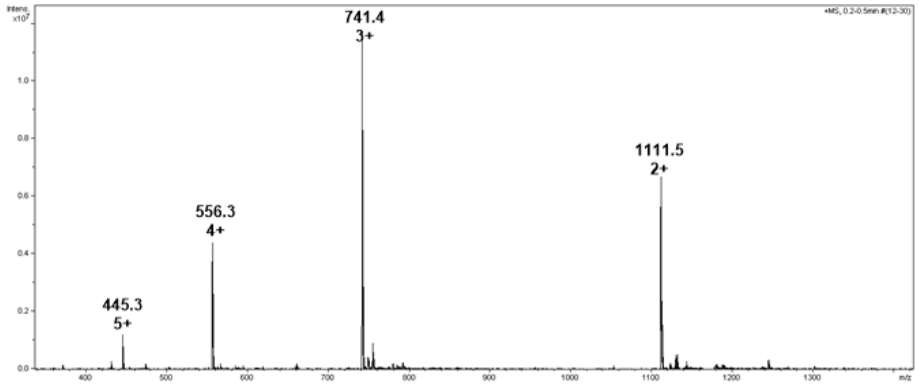
ESI-MS (positive) 832.2(observed, M+H) 830.96 (calculated, M).

5) H-C(Acm)RANK-$NHNH_2$



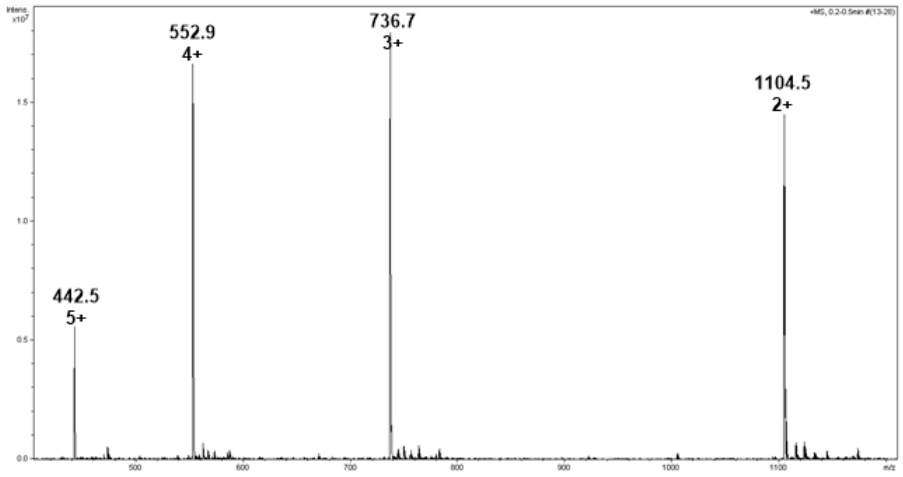
ESI-MS (positive) 676.5(observed, M+H) 675.2 (calculated, M).

6) H-C(Acm)RANKCRALNKSCRALNKS-$NHNH_2$



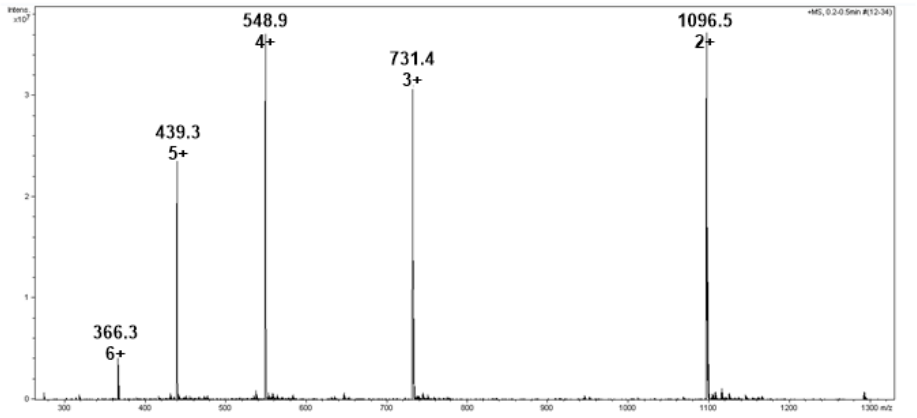
ESI-MS (positive) 1111.5(observed, M+2H) 2220.16 (calculated, M).

7)H-LKCYQHGKVVTCHRDMKF-$NHNH_2$



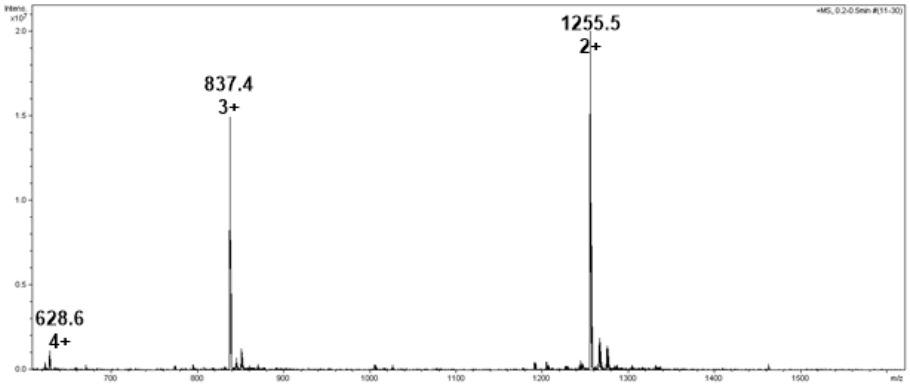
ESI-MS (positive) 1104.5(observed, M+2H) 2207.6 (calculated, M).

8)H-LKCFQHGKVVTCHRDMKF-$NHNH_2$



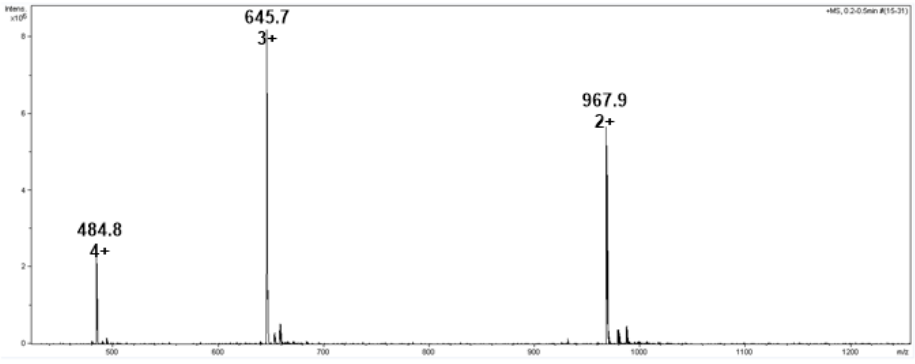
ESI-MS (positive) 1096.5(observed, M+2H) 2191.6 (calculated, M).

9)N_3-CYHNTGMPFRNLKLILQGCSSS-$NHNH_2$



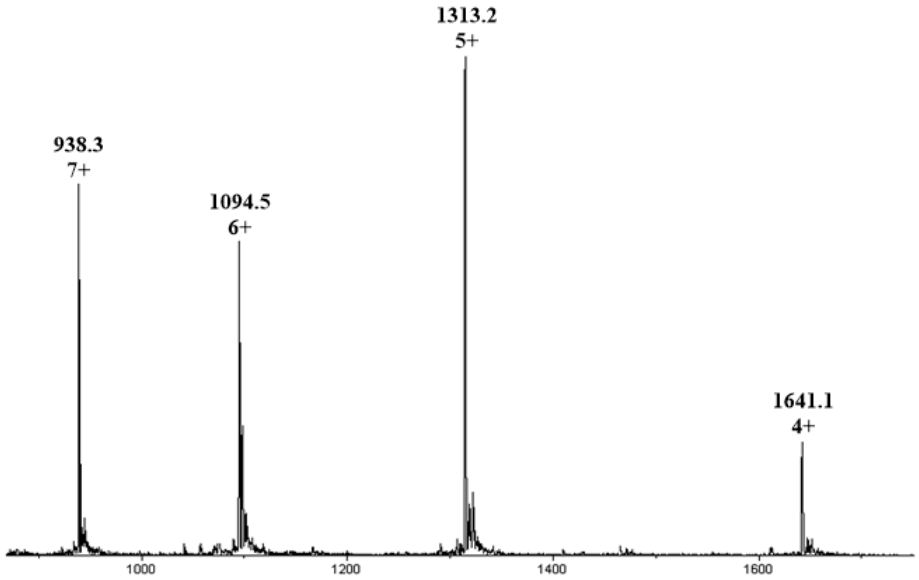
ESI-MS (positive) 1255.5 (observed, M+2H) 2509.9 (calculated, M).

10)H-CSETENNKCCSTDRCNK-OH



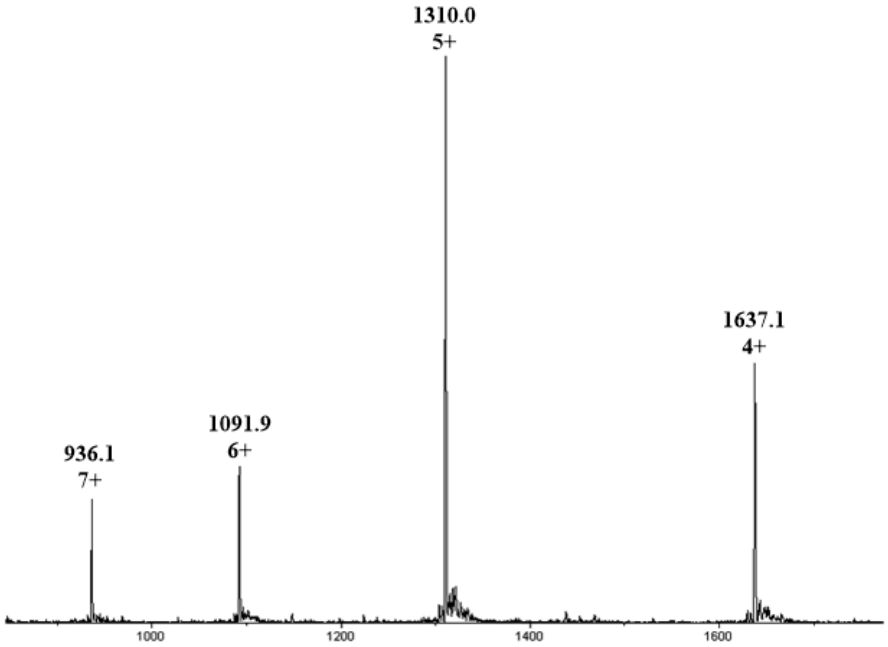
ESI-MS (positive) 967.9 (observed, M+2H) 1934.1 (calculated, M).

11) mambalgin-1 before folding



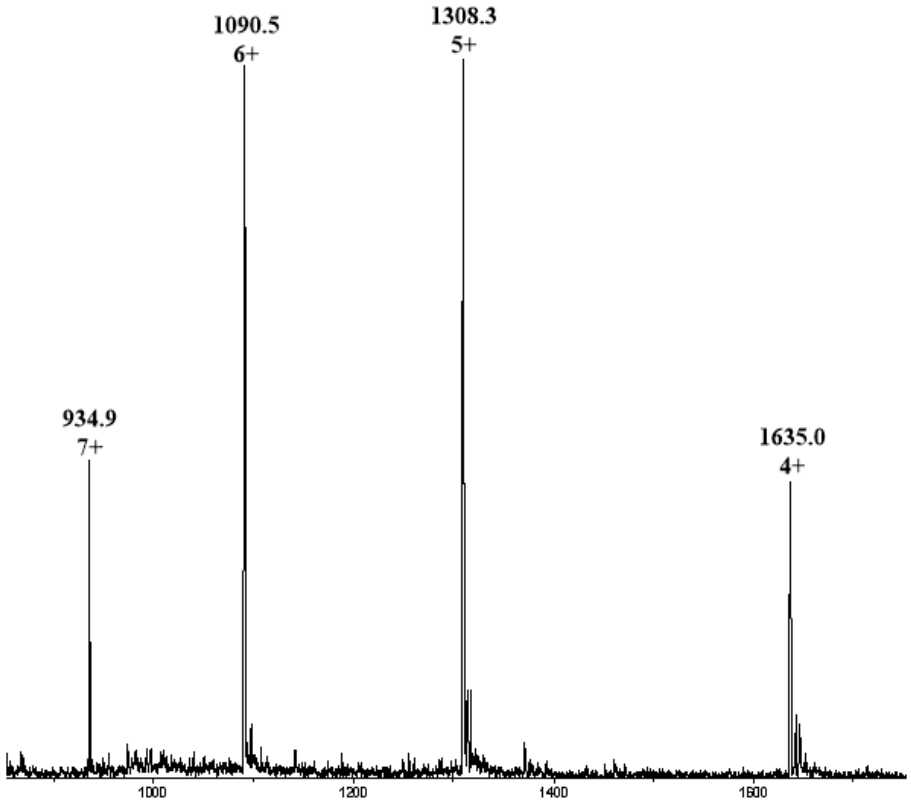
ESI-MS (positive) 1313.2(observed, M+5H) 6562.6 (calculated, M).

12) mambalgin-2 before folding



ESI-MS (positive) 1310.0 (observed, M+5H) 6546.6 (calculated, M).

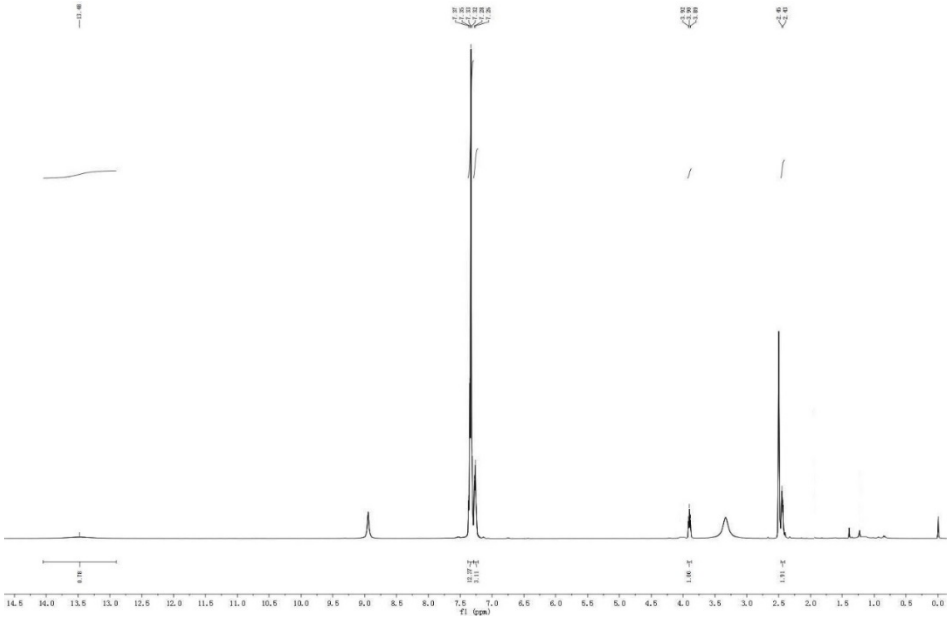
13) mambalgin-2 after folding



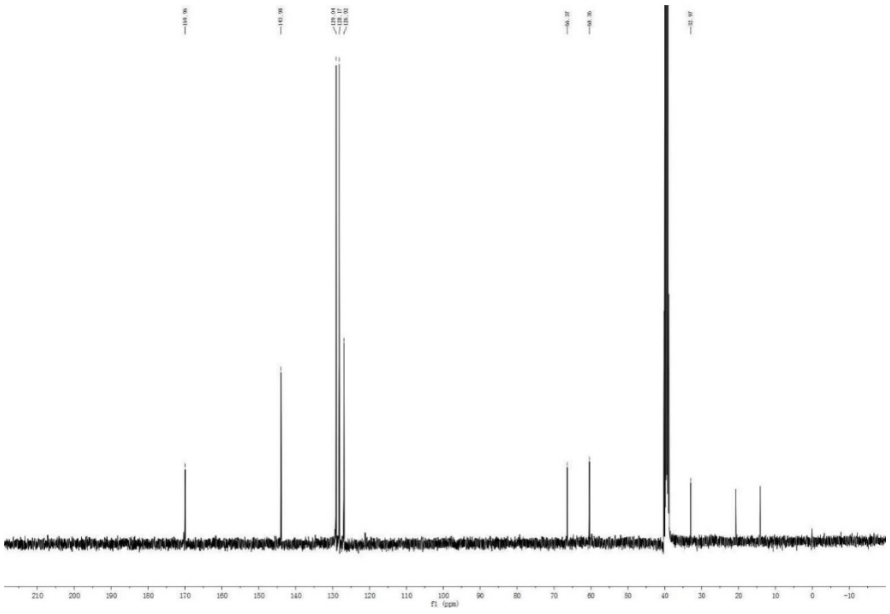
ESI-MS (positive) 1308.3 (observed, M+2H) 6538.6 (calculated, M).

b. NMR spectra

N_3-Cys(Trt)-OH ^{1}H NMR



N_3-Cys(Trt)-OH ^{13}C NMR



Reference

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6: 277-93

DeLano W. 2010. The PyMOL Molecular Graphics System. San Carlos, CA: DeLano Scientific; 2002. Accessed 6/25/2007. Available at http://www. pymol, org

Goddard T, Kneller D. 2006. Sparky—NMR assignment and integration software. *University of California, San Francisco*

Koradi R, Billeter M, Wuthrich K. 1996. MOLMOL: a program for display and analysis of macromolecular structures. *Journal of molecular graphics* 14: 51-5, 29-32

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8: 477-86

Schroeder CI, Rash LD, Vila-Farres X, Rosengren KJ, Mobli M, et al. 2014. Chemical Synthesis, 3D Structure, and ASIC Binding Site of the Toxin Mambalgin-2. *Angew Chem Int Ed Engl* 53: 1017-20

Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. 2003. The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160: 65-73

Wuthrich K. 1986. NMR of proteins and nucleic acids. New York: Wiley