Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry

Convergent Diversity-Oriented Side-Chain Macrocyclization Scan for Unprotected Polypeptides

Supporting Information

Yekui Zou,[1] Alexander M. Spokoyny,[1] Chi Zhang,[1] Mark D. Simon,[1] Hongtao Yu,[2] Yu-Shan Lin,[2] and Bradley L. Pentelute[*,1]

[1]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, United States; [2]Department of Chemistry, Tufts University, Medford, MA 02155, United States.
* E-mail: blp@mit.edu

1. Materials, Methods and General Considerations

N$^{\alpha}$-Fmoc protected L-amino acids, 2-(1H-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), and Fmoc-Rink amide linker were purchased from Chem-Impex International (Wood Dale, IL). MBHA resin was obtained from Anaspec (Fremont, CA). Hexafluorobenzene and decafluorobiphenyl were purchased from Oakwood Chemicals (West Columbia, SC) and used as received. Decafluorobiphenyl sulfide was purchased from SynQuest Laboratories (Alachua, FL). *N,N*-dimethylformamide (DMF), dichloromethane (DCM), diethyl ether, and HPLC-grade acetonitrile were purchased from VWR International (Philadelphia, PA). All other reagents were purchased from Sigma-Aldrich and used as received, unless otherwise noted.

LC-MS chromatograms and associated mass spectra were acquired using Agilent 6520 ESI-QTOF mass spectrometer equipped with Zorbax SB C_3 column: 2.1 x 150 mm, 5 μm, gradient: 0-2 minutes 5% B, 2-11 minutes 5-65% B, 11-12 minutes 65% B, flow rate: 0.8 mL/min. Data was processed using Agilent Mass Hunter software package. Deconvoluted mass spectra were obtained using maximum entropy setting. All conversions associated with peptides shown in this work were determined by measuring UV absorption at λ=214 nm using LC-MS. By use of the Agilent Mass Hunter software package, the areas for all relevant chromatographic peaks were integrated. Then, the relevant peak areas from UV absorption spectra at λ=214 nm that arose from unreacted starting material, desired peptide macrocyclic product, disulfide crosslinked byproducts, and other peptide-containing byproducts were summed. Conversion was calculated using equation: %yield = S_{pro}/S_{all}, where S_{pro} corresponds to the peak area of the desired macrocyclized peptide and S_{all} is the sum of all peak areas including the product and all byproducts.

NMR spectroscopy was conducted on a 400 MHz Bruker Avance III system and processed with TopSpin 3.1 software supplied by Bruker. Variable temperature NMR spectra were plotted using MestReNova suite, version 7.1. ^{1}H NMR spectra were referenced to residual solvent resonances in deuterated solvents. ^{19}F NMR spectra were referenced to $CFCl_3$ external standard at δ 0.0. The following abbreviations are used in the description of the acquired NMR spectra: s – singlet, d – doublet, m – multiplet.

The initial structures of the compounds for DFT calculations were generated and energy minimized with the molecular modeling software Avogadro.[2] After a full geometry optimization at the B3LYP/6-311G(d,p) level with the Gaussian 09 Rev. D.01 package,[3] the Mulliken charges were calculated.

2. Solid-Phase Supported Peptide Synthesis and Purifications

Peptides **1-14** were synthesized on a 0.2 mmol scale on MBHA resin using manual Fmoc-SPPS chemistry[1] and a home-built flow-based system. Specifically, all reagents and solvents are delivered to a stainless steel reactor containing resin at a constant flow rate using HPLC pump; temperature of the

reactor was maintained at 60 °C during the synthesis using water bath. The procedure used for coupling cycle included a 30 second coupling with 1 mmol Fmoc protected amino acids, 1.0 mmol HBTU, and 500 μL of diisopropyl ethyl amine (DIEA) in 2.5 mL of DMF, at a flow rate of 6 mL/min; 1 minute wash with DMF, flow rate was 20 mL/min; 20 seconds deprotection with 50% (v/v) piperidine in DMF, flow rate was 20 mL/min; and 1 minute wash with DMF, flow rate was 20 mL/min. The C-terminal amide was introduced using Fmoc-Rink amide linker. Side-chain protections for L-amino acids were as followed: Arg(Pbf), Cys(Trt), Glu(*t*Bu), Ser(*t*Bu), Lys(Boc), Tyr(*t*Bu), and Thr(*t*Bu). The resin was washed thoroughly with DCM and air dried after completion of the stepwise SPPS. The peptide is then simultaneously cleaved from the resin and side-chain deprotected by treatment with 2.5% (v/v) water, 2.5% (v/v) 1,2-ethanedithiol (EDT), and 1% (v/v) triisoproprylsilane in trifluoroacetic acid (TFA) for 2 hours at room temperature. The resulting solution was evaporated by blowing a stream of nitrogen gas over its surface for 20-30 minutes, then triturated and washed three times with cold diethyl ether. Obtained gummy-like solid was dissolved in 50% H_2O: 50% (v/v) acetonitrile containing 0.1% (v/v) TFA and lyophilized. ***These same solvent compositions were used in most experiments and will be referred to as A: 0.1% TFA in H_2O and B: 0.1% TFA in acetonitrile.***

The crude peptide was dissolved in 95% A: 5% B with 6 M guanidinium hydrochloride and purified by preparative RP-HPLC (Agilent Zorbax SB C_{18} column: 10 x 250 mm, 7 μm, linear gradient: 5-50% B over 90 min, flow rate: 5 mL/min). HPLC fractions were spotted with MALDI using alpha-cyano-4-hydroxycinnamic acid (CHCA) matrix in 50% A: 50% B and checked for fractions with desired molecular mass. The purity of fractions was confirmed by analytical RP-HPLC (Agilent Zorbax SB C_3 column: 2.1 x 150 mm, 5 μm, gradient: 0-2 minutes 5% B, 2-11 minutes 5-65% B, 11-12 minutes 65% B, flow rate: 0.8 mL/min). HPLC fractions containing only product material (screened by MALDI) were combined and lyophilized.

3. Synthesis of Linkers

L_d: To a magnetically stirred solution of ethanedithiol (0.27 g, 3 mmols) in 30 mL of anhydrous acetonitrile in a Schlenk flask cooled on ice-bath under positive flow of argon gas was added sodium phosphate (0.98 g, 6 mmol), followed by neat hexafluorobenzene (22 g, 75 mmol). Resulting suspension was warmed to room temperature and left stirring for 5 hours under inert atmosphere. Product mixture was then filtered through a pad of Celite™ on a fritted glass filter, evaporated *in vacuo* and subjected to purification on a silica gel column (2:1 hexanes/dichloromethane). Fractions containing product were combined and dried *in vacuo* affording **L_d** as white solid (0.23 g, 18% yield). ^{1}H NMR (400.1 MHz, $CDCl_3$, 298 K): δ 3.01 (s, 4H, Alkyl-CH_2); ^{13}C NMR* (100.6 MHz, $CDCl_3$, 298 K): δ 34.6 (s, alkyl-**C**H_2); ^{19}F NMR (376.5 MHz, $CDCl_3$, 298 K): δ −132.1 (m, 4F), − 151.4 (m, 2F), − 160.3 (m, 4F).
* due to complex ^{13}C-^{19}F coupling patterns, these resonances exhibited very low signal-to-noise quality in ^{13}C NMR spectrum precluding their clear identification.

L_g was prepared according to the procedure analogous to the one used for the synthesis of **L_d** (*vide supra*) on a 3 mmol scale from 1,4-benzenedimethanethiol (Sigma-Aldrich, CAS Number: 105-09-9) affording **L_g** as a white solid (0.31 g, 20% yield). ^{1}H NMR (400.1 MHz, $CDCl_3$, 298 K): δ 7.10 (s, 4H,

Ar-H), 4 (s, 4H, Benzyl-CH_2); ^{13}C NMR* (100.6 MHz, CD_2Cl_2, 298 K): δ 136.3 (s, aryl-C), 128.9 (s, aryl-CH), 38.7 (benzyl-CH_2); ^{19}F NMR (376.5 MHz, $CDCl_3$, 298 K): δ −132.4 (m, 4F), − 152.1 (m, 2F), − 161.1 (m, 4F). *due to complex ^{13}C-^{19}F coupling patterns, these resonances exhibited very low signal-to-noise quality in ^{13}C NMR spectrum precluding their clear identification.

L_f: $CuSC_6F_5$ (1.5 g, 5.8 mmol)[4] and 1,4-diiodobenzene (0.66 g, 2 mmol) were combined in a 15 mL oven-dried test-tube equipped with a small magnetic stir bar. Reaction vessel was sealed with Teflon screw-cap septum, evacuated under vacuum and back-filled with argon gas. Anhydrous dimethylformamide (5 mL) was added to the reaction vessel and the resulting mixture was allowed to stir at 140 °C for 2.5 hours. Product mixture was cooled to room temperature, quenched with 10% solution of aqueous hydrochloric acid and extracted with diethyl ether (2 times, 30 mL total). Organic fractions were combined, dried *in vacuo* and subjected to column chromatography purification on silica gel eluted with 5% dichloromethane in hexanes solvent mixture. Collected product fractions were combined, dried *in vacuo* and recrystallized from hexanes at − 20 °C yielding white solid product (0.58 g, 61% yield). ^{1}H NMR (400.1 MHz, $CDCl_3$, 298 K): δ 7.27 (s, 4H, Aryl-H); ^{13}C NMR* (100.6 MHz, CD_2Cl_2, 298 K): δ 133.1 (s, aryl-C), 131.0 (s, aryl-CH); ^{19}F NMR (376.5 MHz, $CDCl_3$, 298 K): δ −131.4 (m, 4F), − 150.4 (m, 2F), − 159.9 (m, 4F). *due to complex ^{13}C-^{19}F coupling patterns, these resonances exhibited very low signal-to-noise quality in ^{13}C NMR spectrum precluding their clear identification.

L_e was prepared according to the procedure analogous to the one used for the synthesis of **L_f** (*vide supra*) on a 2 mmol scale from 1,4-dibromo-2,3,5,6-tetrafluorobenzene (Sigma Aldrich, CAS Number: 344-03-6) affording **L_e** as an off-white solid (0.66 g, 60% yield). ^{19}F NMR (376.5 MHz, C_6D_6, 298 K): δ −132.8 (d, 4F, $J_{F\text{-}F}$ = 22 Hz), −132.9 (s, 4F), − 150.6 (m, 2F), − 161.2 (m, 4F). Due to complex ^{13}C-^{19}F coupling patterns, these resonances exhibited very low signal-to-noise quality in ^{13}C NMR spectrum precluding their clear identification.

4. Syntheses of Peptides 7a', 7b' and 14a':

To a solid sample of peptide **7** (5 μmoles) in a plastic Eppendorf tube dissolved in 20 mM Tris base in 1 mL of DMF was added 500 μmoles of hexafluorobenzene or decafluorobiphenyl for peptide **7a'** or peptide **7b'** respectively. The tube was vortexed for 1 minute and left at room temperature for 1 hour. The reaction mixture was characterized by LC-MS analysis till complete. Resulting reaction mixture was quenched by addition of 20 mL of 95% A: 5% B. Peptide **7a'** and **7b'** were purified by RP-HPLC. Fractions were characterized by LC-MS analysis. Pure fractions containing product were collected and lyophilized.

To a solid sample of peptide **14** (5 μmoles) in a plastic Eppendorf tube dissolved in 20 mM Tris base in 1 mL of DMF was added 500 μmoles of hexafluorobenzene for peptide **14a'**. The tube was vortexed for 1 minute and left at room temperature for 1 hour. The reaction mixture was characterized by LC-MS analysis till complete. Resulting reaction mixture was quenched by addition of 20 mL of 95% A: 5% B. Peptide **14a'** were purified by RP-HPLC. Fractions were characterized by LC-MS analysis. Pure fractions containing product were collected and lyophilized.

5. General Protocols for Peptide Macrocyclization Scan

Note that all the peptide stock solutions were made fresh and used immediately. Deoxygenated DMF was used.

Reactions with Linker L_a. To a solution of peptide (10 mM, 5 μL) in a plastic Eppendorf tube was added hexafluorobenzene (20 mM in DMF, 20 μL) and TRIS base (50 mM in DMF, 20μL). The reaction was mixed by pipetting the solution up and down 20 times and left at room temperature for 2 hours. Reaction progress was monitored via LC-MS. For each LC-MS sample, 1 μL of the reaction was diluted with 20 μL of 50% A: 50% B.

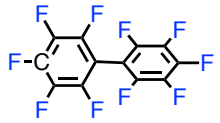
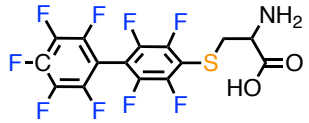
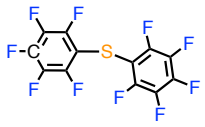
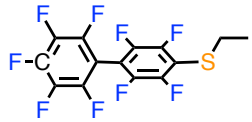
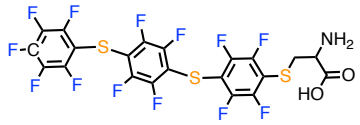
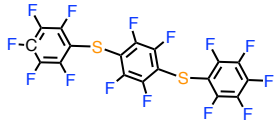
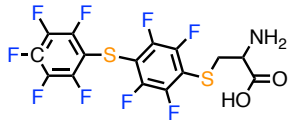
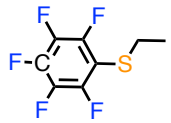
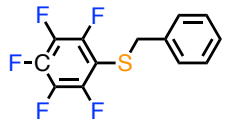
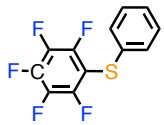
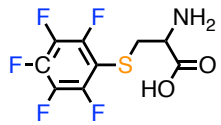
Reactions with Linker L_b-L_g. To a stock solution of peptide dissolved in DMF (10 mM, 8 μL) in a plastic Eppendorf tube was added linkers $\mathbf{L_b}$-$\mathbf{L_g}$ (20 mM in DMF, 5 μL) and TRIS base (50 mM in DMF, 16 μL) and additional 11 μL of DMF. The reaction was mixed by pipetting the solution up and down 20 times and left at room temperature for 2 hours. Reaction progress was monitored via LC-MS. For each LC-MS sample, 1 μL of the reaction was diluted with 20 μL of 50% A: 50% B.

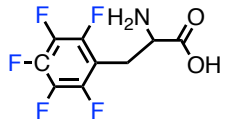
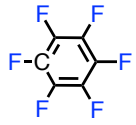
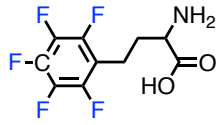
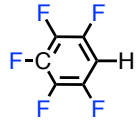
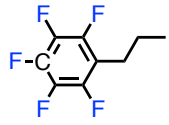
Reactions with Linker L_b and L_c at diluted concentrations. To a stock solution of peptide dissolved in DMF (10 mM, 1 μL) in a plastic Eppendorf tube was added linkers $\mathbf{L_b}$ or $\mathbf{L_c}$ (20 mM in DMF, 0.5 μL) and TRIS base (50 mM in DMF, 20 μL) and additional 78.5 μL of DMF. The reaction was mixed by pipetting the solution up and down 20 times and left at room temperature for 2 hours. Reaction progress was monitored via LC-MS. For each LC-MS sample, 1 μL of the reaction was diluted with 20 μL of 50% A: 50% B.

Reactions with dithiol linkers. To a stock solution of bis-perfluoroarylated peptide dissolved in DMF (10 mM, 2 μL) in a plastic Eppendorf tube was added 1,4-benzenedimethanethiol (40 mM in DMF, 1 μL) or 1,4-butanedithiol (40 mM in DMF, 1 μL) and TRIS base (50 mM in DMF, 8 μL) and additional 9 μL of DMF. The reaction was mixed by pipetting the solution up and down 20 times and left at room temperature for 2 hours. Reaction progress was monitored via LC-MS. For each LC-MS sample, 1 μL of the reaction was diluted with 20 μL of 50% A: 50% B.

Supplementary Tables and Figures:

Table S1. Calculated Mulliken atomic charges of different perfluoroaryl-containing molecules.

Compound	Charge on C (e)	Charge on F (e)	Total charge of C-F (e)
	0.199	-0.183	0.016
	0.199	-0.183	0.016
	0.197	-0.182	0.015
	0.198	-0.185	0.013
	0.196	-0.183	0.013
	0.196	-0.183	0.013
	0.195	-0.184	0.011
	0.195	-0.187	0.008
	0.185	-0.177	0.008
	0.193	-0.187	0.006
	0.192	-0.188	0.004

	0.192	-0.191	0.001
	0.188	-0.188	0.000
	0.190	-0.192	-0.002
	0.187	-0.189	-0.002
	0.189	-0.193	-0.004

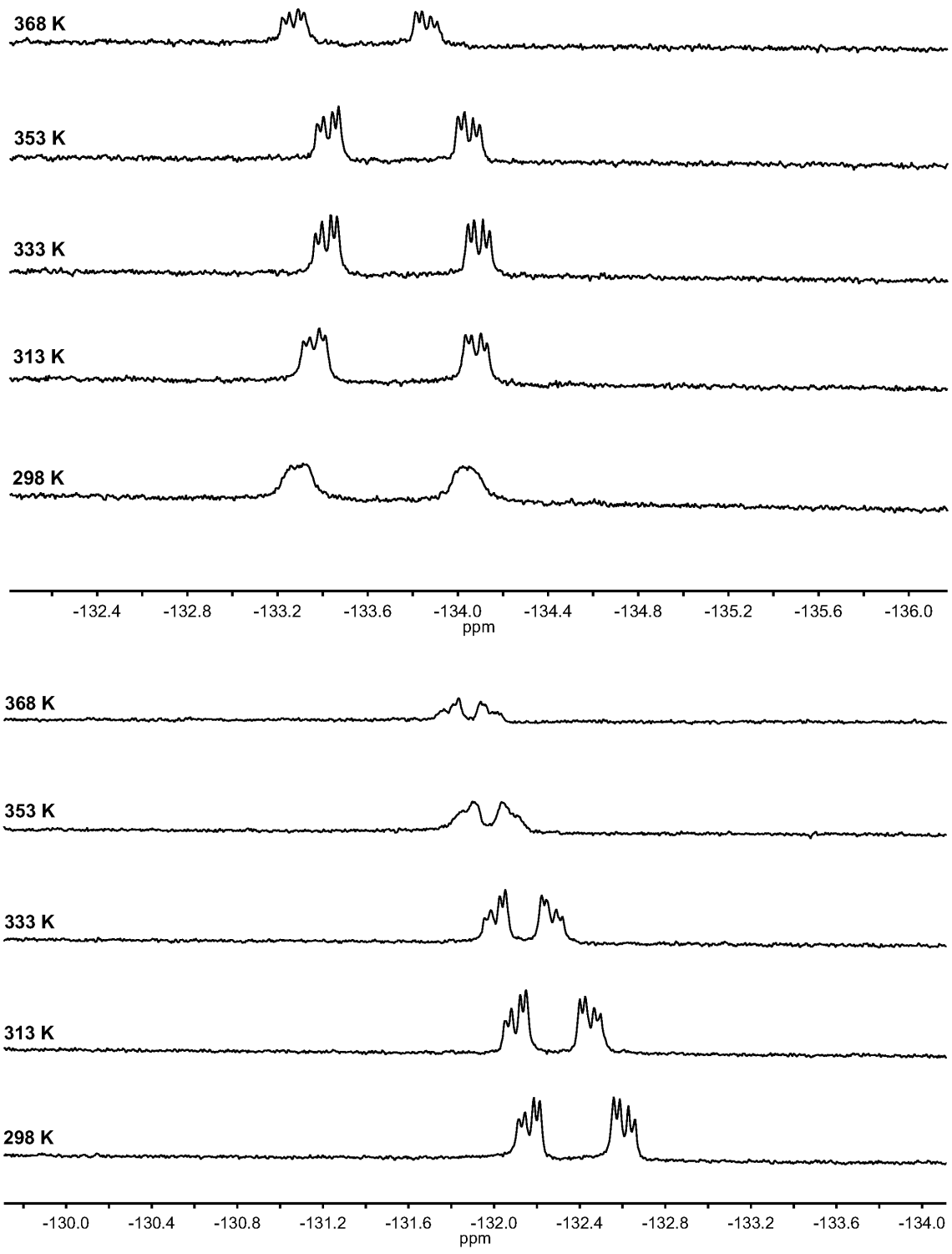


Figure S1. Variable temperature (VT) ^{19}F-NMR spectra of purified peptide **7a**: spectrum acquired in d_6-DMF (top) and in D_2O (bottom) at different temperatures.

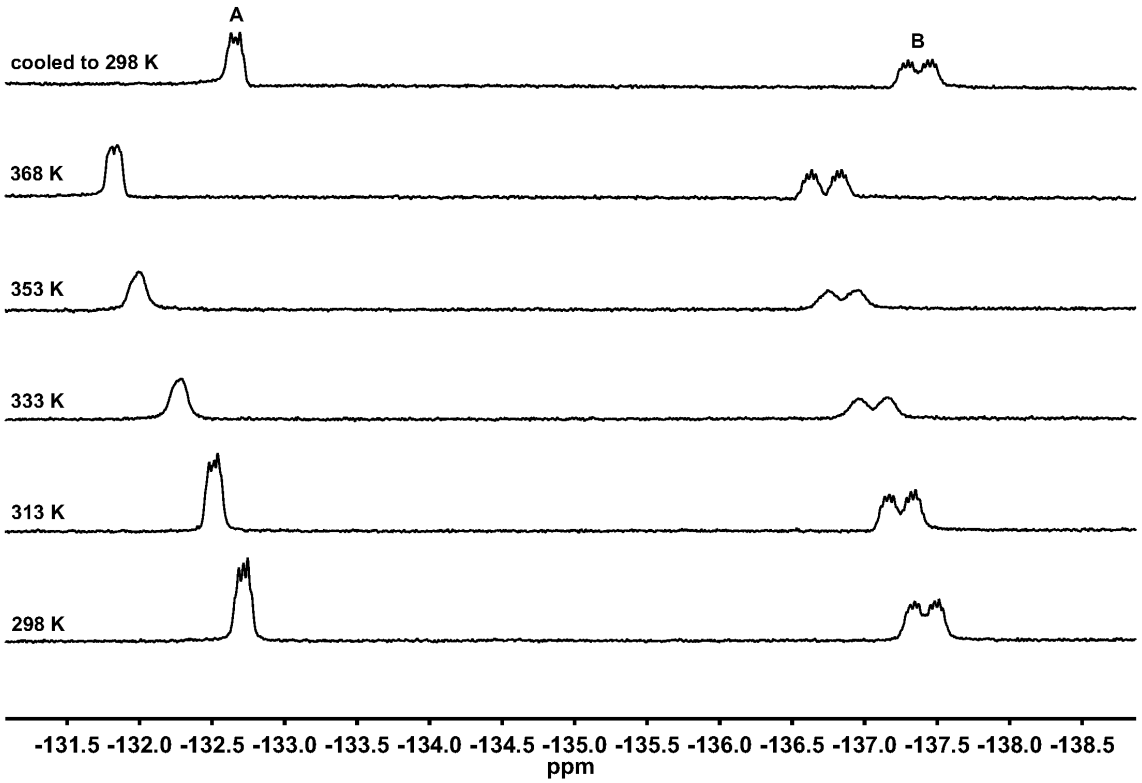


Figure S2. Variable temperature (VT) ^{19}F-NMR spectrum of purified peptide **7b** in D_2O at different temperatures.

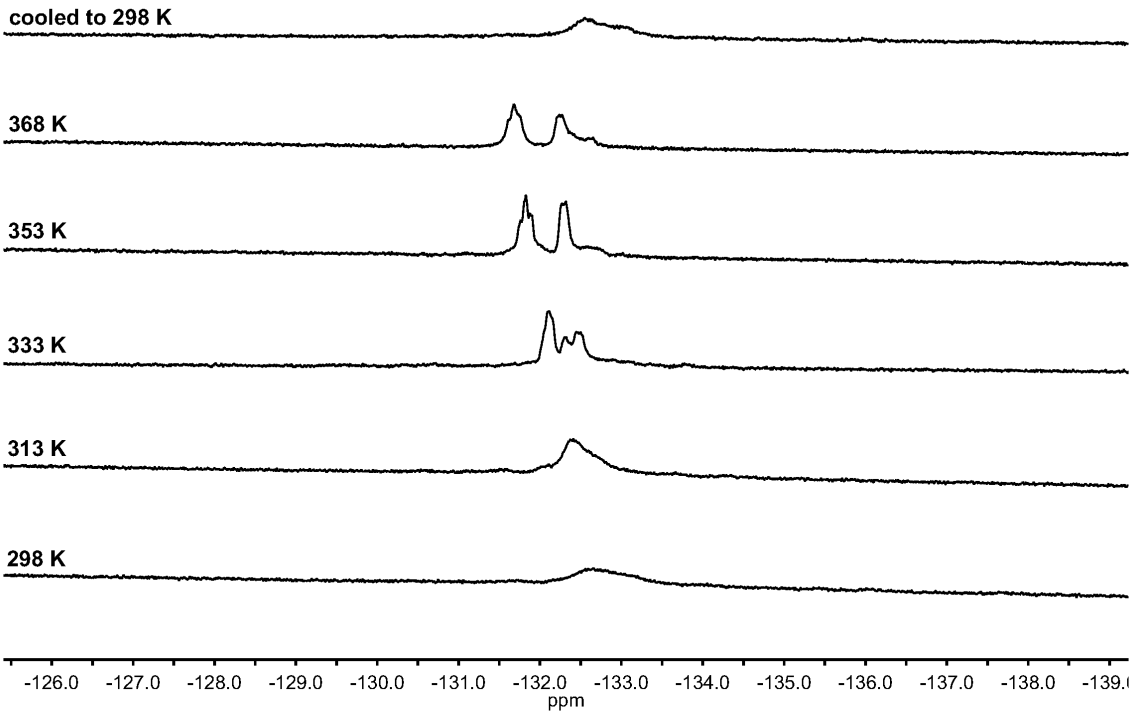


Figure S3. Variable temperature (VT) ^{19}F-NMR spectra of purified peptide **7c** in D_2O at different temperatures.

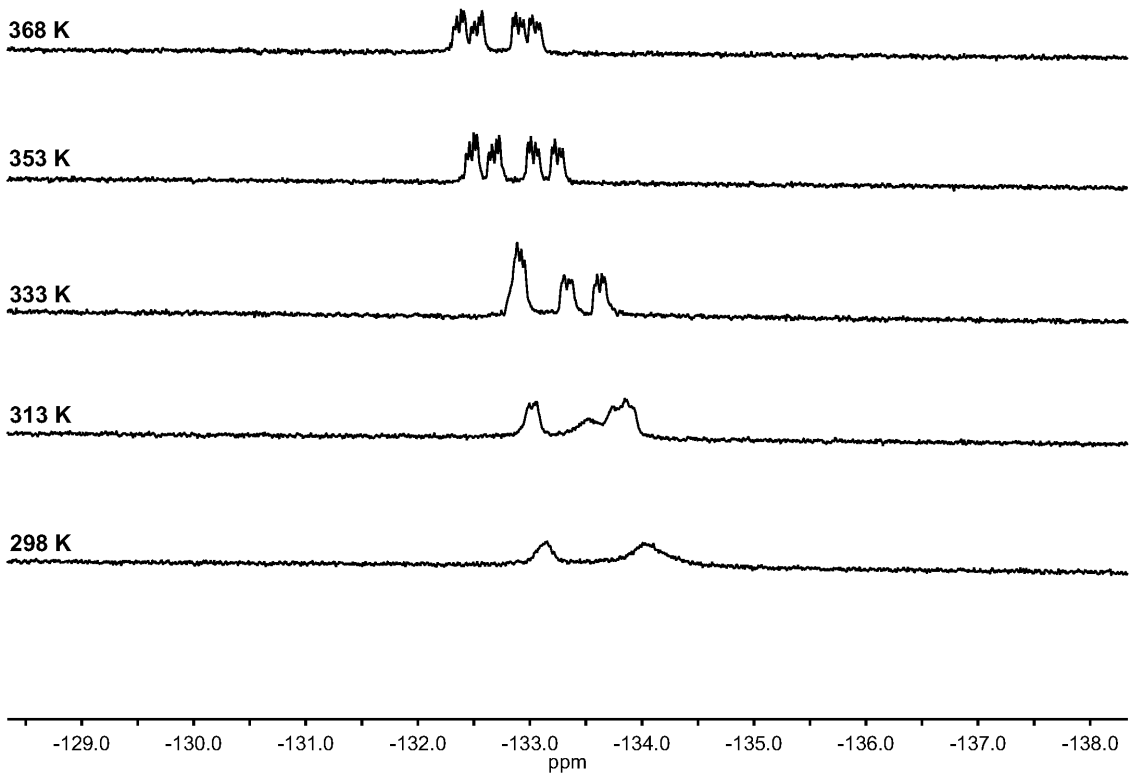


Figure S4. Variable temperature (VT) ^{19}F-NMR spectra of purified peptide **7d** in D_2O at different temperatures.

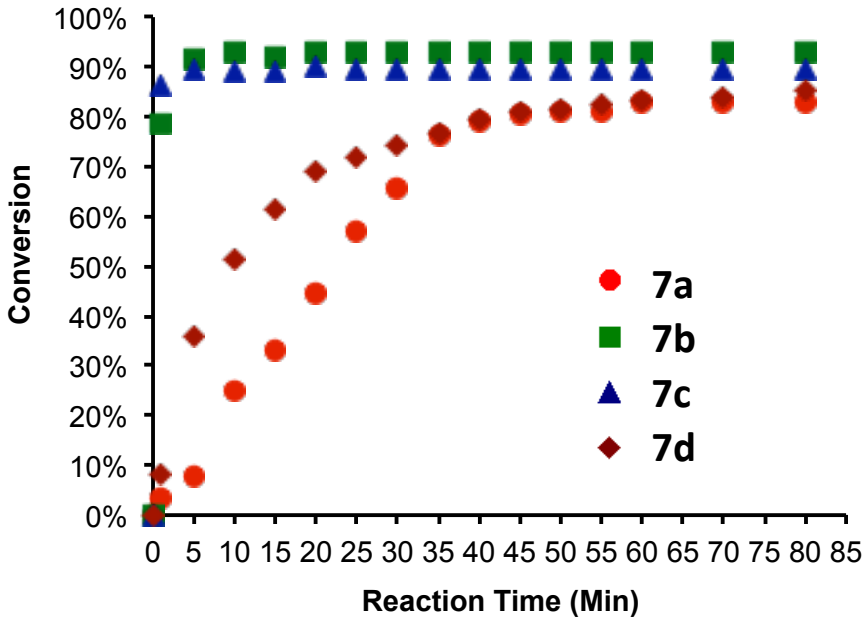


Figure S5. Rates of peptide **7** macrocyclization with linkers $\mathbf{L_a}$, $\mathbf{L_b}$, $\mathbf{L_c}$, and $\mathbf{L_d}$ measured by LC-MS. Reactions were performed as described in the general protocols. For each time point, 1 µL of the resulting reaction mixture was diluted with 20 µL of 50% A: 50% B and subjected to LC-MS analysis.

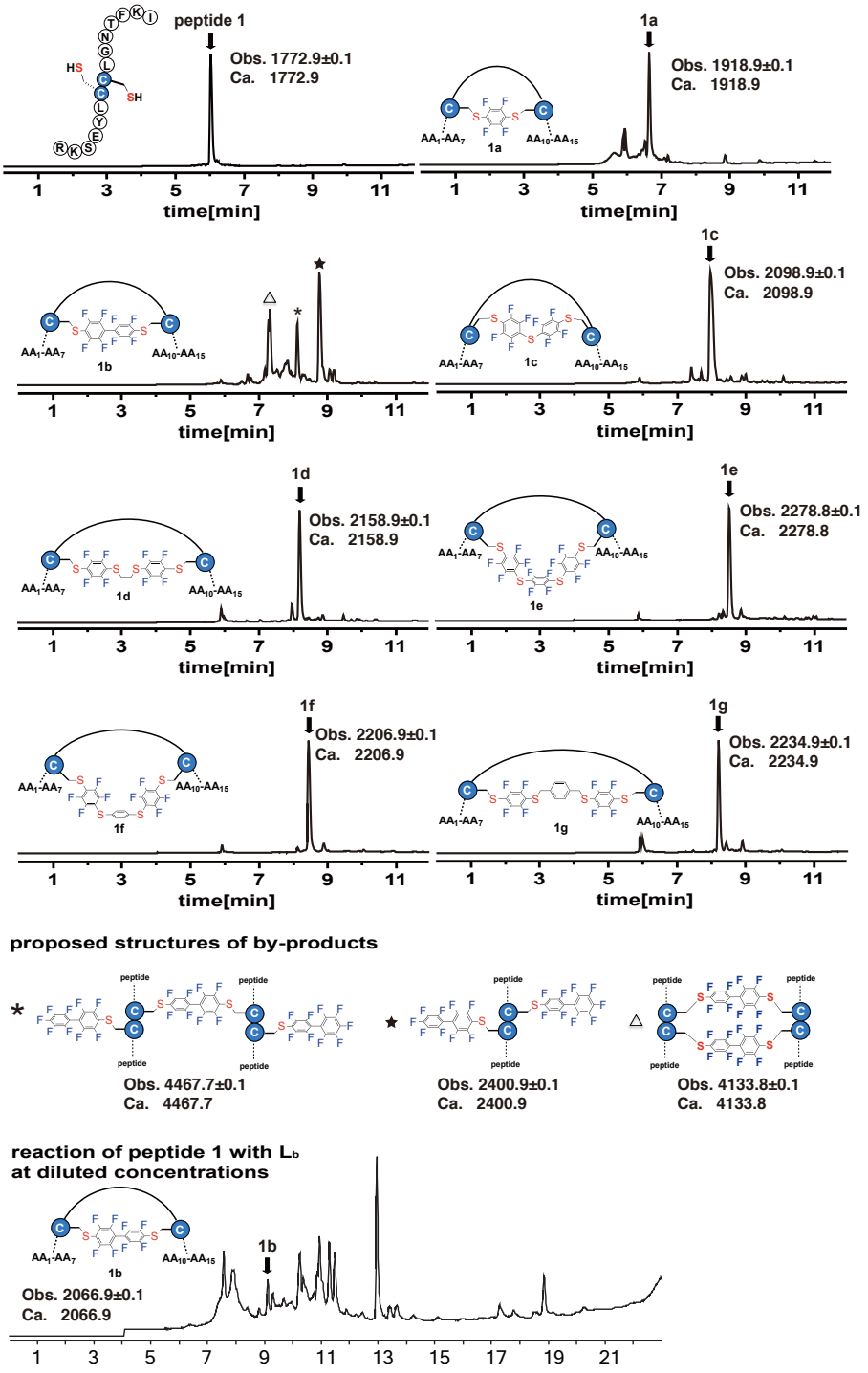


Figure S6. LC-MS chromatograms (total ion current) of purified peptide **1** and corresponding macrocyclization reactions with linkers $\mathbf{L_a}$-$\mathbf{L_g}$ analyzed *in situ*. The peak labeled as * is the by-product generated from the crosslinking of two peptides via one linker with the other two cysteines arylated. The peak labeled as ★ is the dual-arylated by-product. The peak labeled as Δ represent the by-product generated from the crosslinking of two peptides with two linkers. Reaction conditions at diluted concentrations: 0.1 mM peptide, 0.1 mM linker $\mathbf{L_b}$, 20 mM Tris, DMF, room temperature, and 60 minutes. Mass (Da) shown was calculated at the highest intensity of the TIC peak.

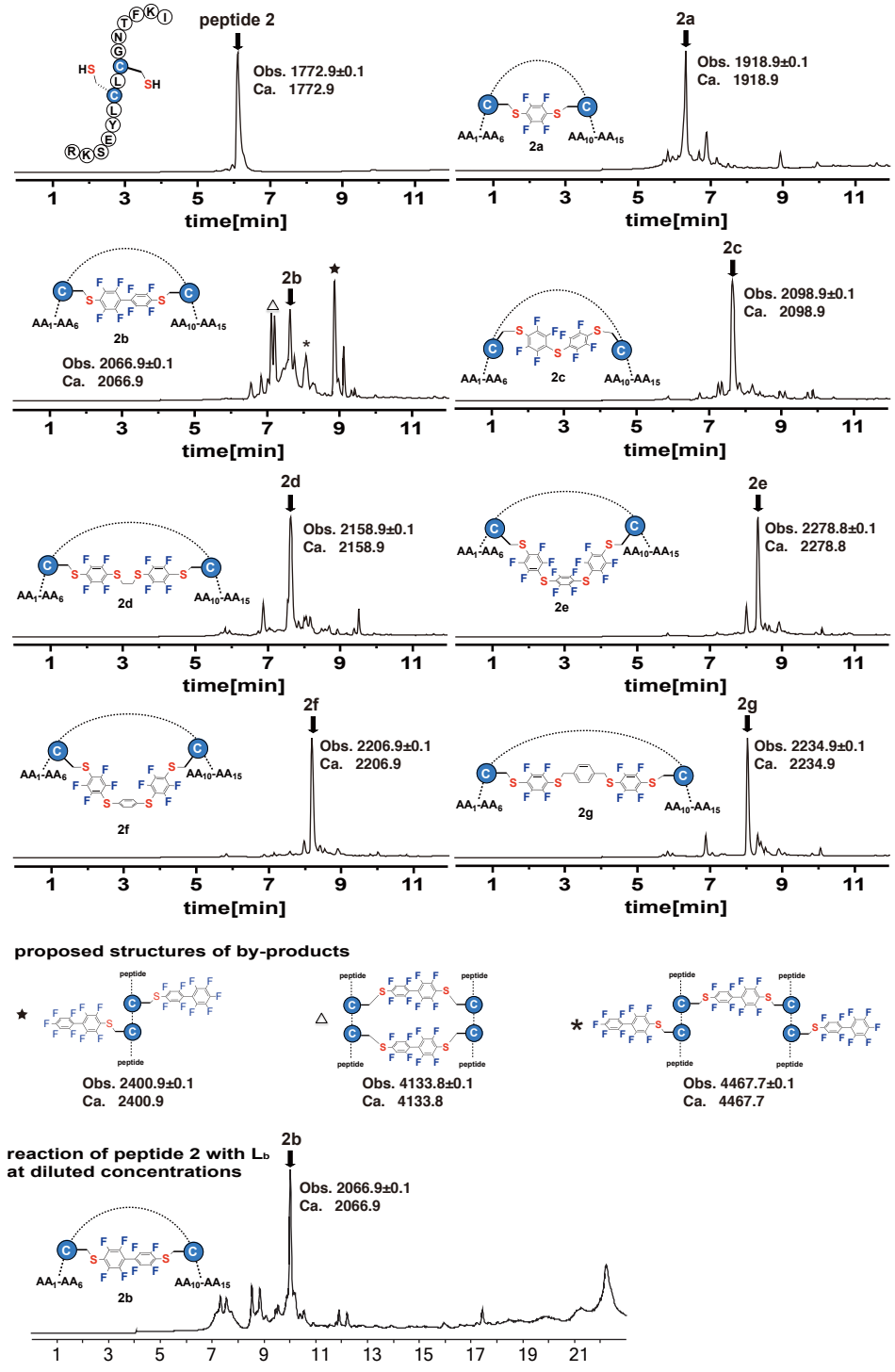


Figure S7. LC-MS chromatograms (total ion current) of purified peptide **2** and corresponding macrocyclization reactions with linkers $\mathbf{L_a}$-$\mathbf{L_g}$ analyzed *in situ*. The peak labeled as * is the by-product generated by the crosslinking of two peptides with one linker and the other two cysteines are arylated. The peak labeled as ★ is the dual-arylated by-product. The peaks labeled as Δ are the by-products generated from crosslinking of two peptides with two linkers. Reaction conditions at diluted concentrations: 0.1 mM peptide, 0.1 mM linker $\mathbf{L_b}$, 20 mM Tris, DMF, room temperature, and 60 minutes. Mass (Da) shown was calculated at highest intensity of the TIC peaks.

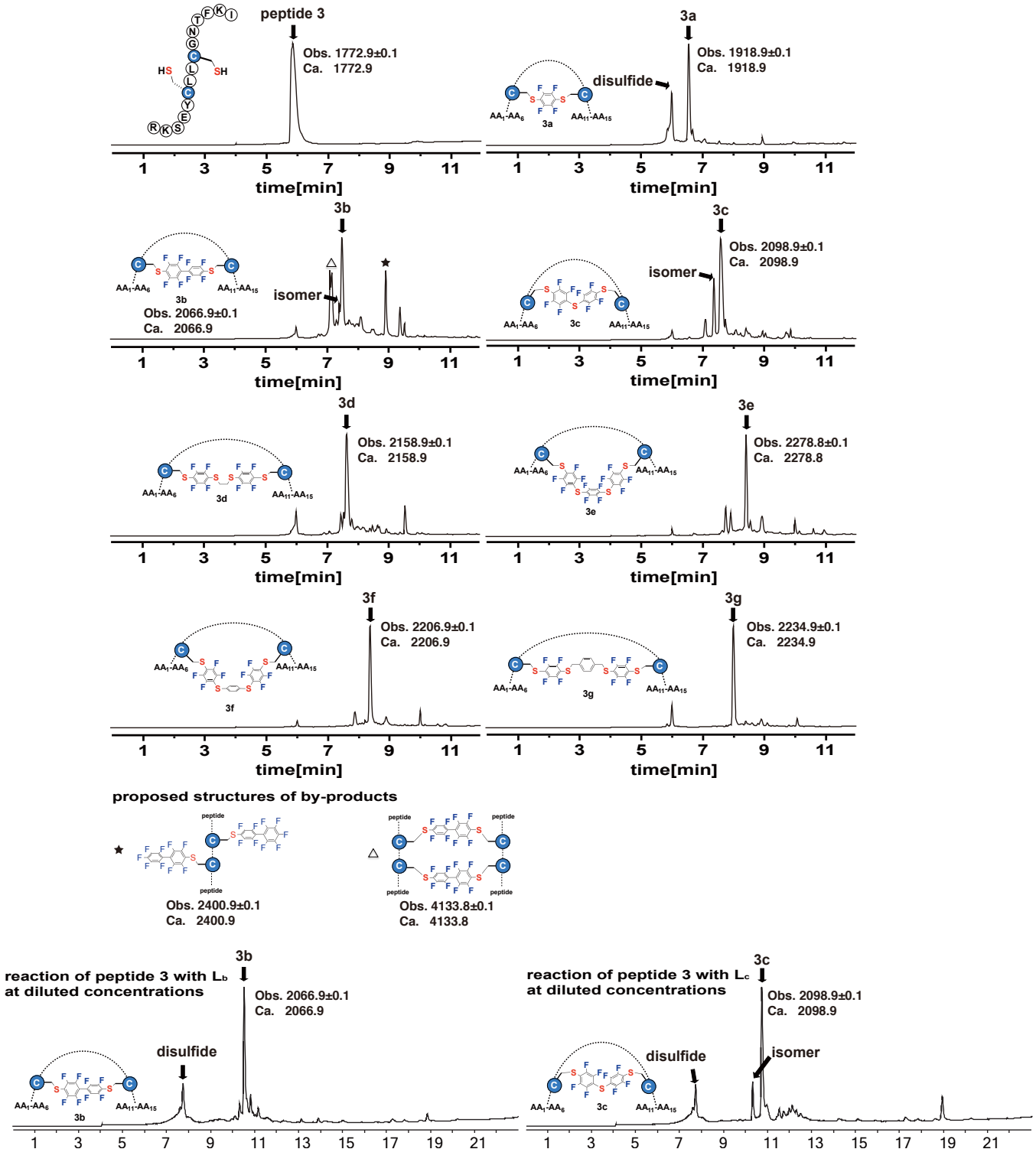


Figure S8. LC-MS chromatograms (total ion current) of purified peptide **3** and corresponding macrocyclization reactions with linkers $\mathbf{L_a}$-$\mathbf{L_g}$ analyzed *in situ*. Macrocyclization with linker $\mathbf{L_b}$ and $\mathbf{L_c}$ produce one isomer with same mass as the desired product. The peaks labeled as ★ are the dual-arylated by-products. The peaks labeled as Δ are by-products generated from crosslinking of two peptides with two linkers. Reaction conditions at diluted concentrations: 0.1 mM peptide, 0.1 mM linker $\mathbf{L_b}$ or $\mathbf{L_c}$, 20 mM Tris, DMF, room temperature, and 60 minutes. Mass (Da) shown was calculated at highest intensity of the TIC peaks.

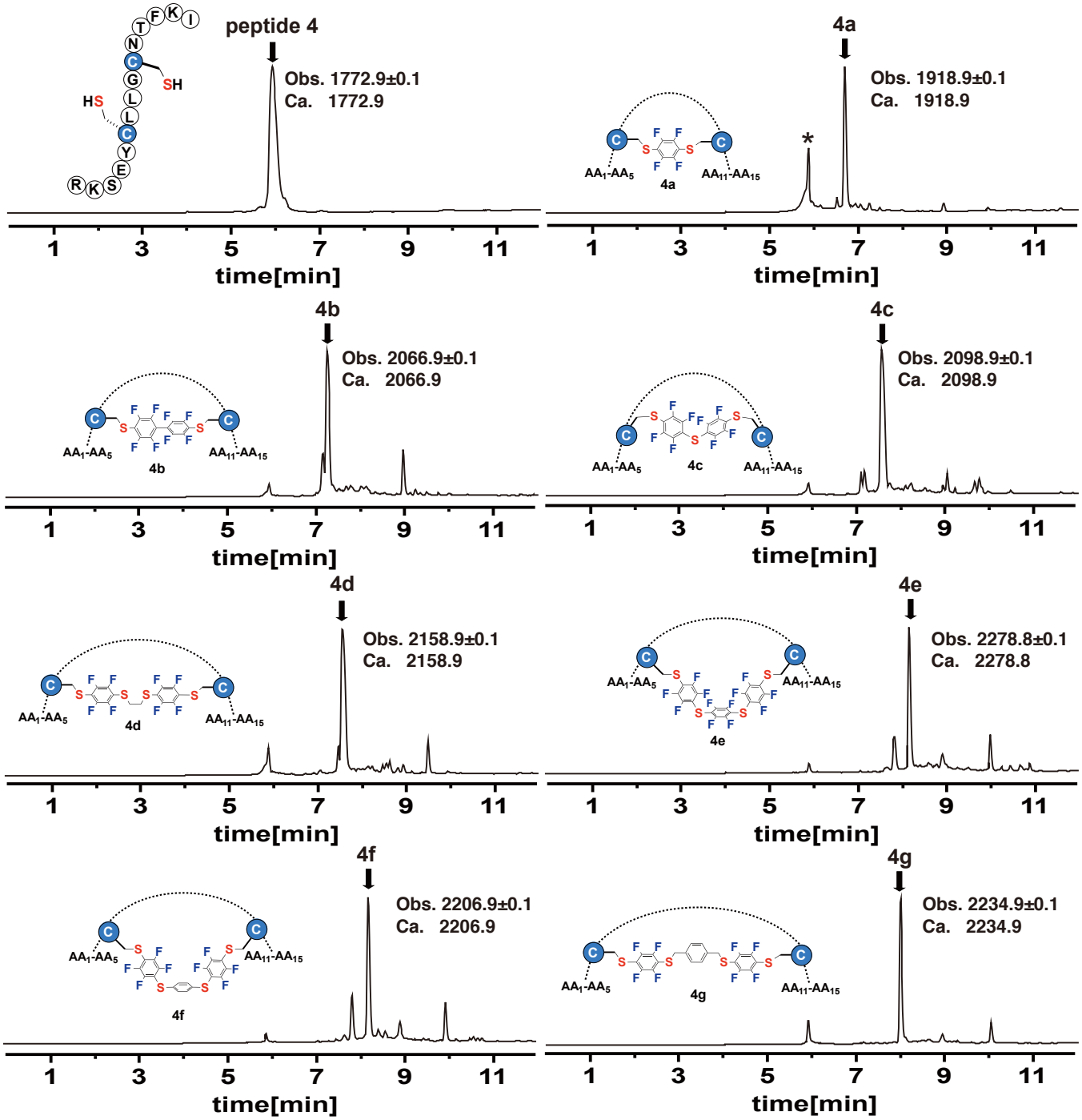


Figure S9. LC-MS chromatograms (total ion current) of purified peptide **4** and corresponding macrocyclization reactions with linkers $\mathbf{L_a}$-$\mathbf{L_g}$ analyzed *in situ*. The peak labeled as * is the oxidized disulfide by-product. Mass (Da) was calculated at highest intensity of the TIC peaks.

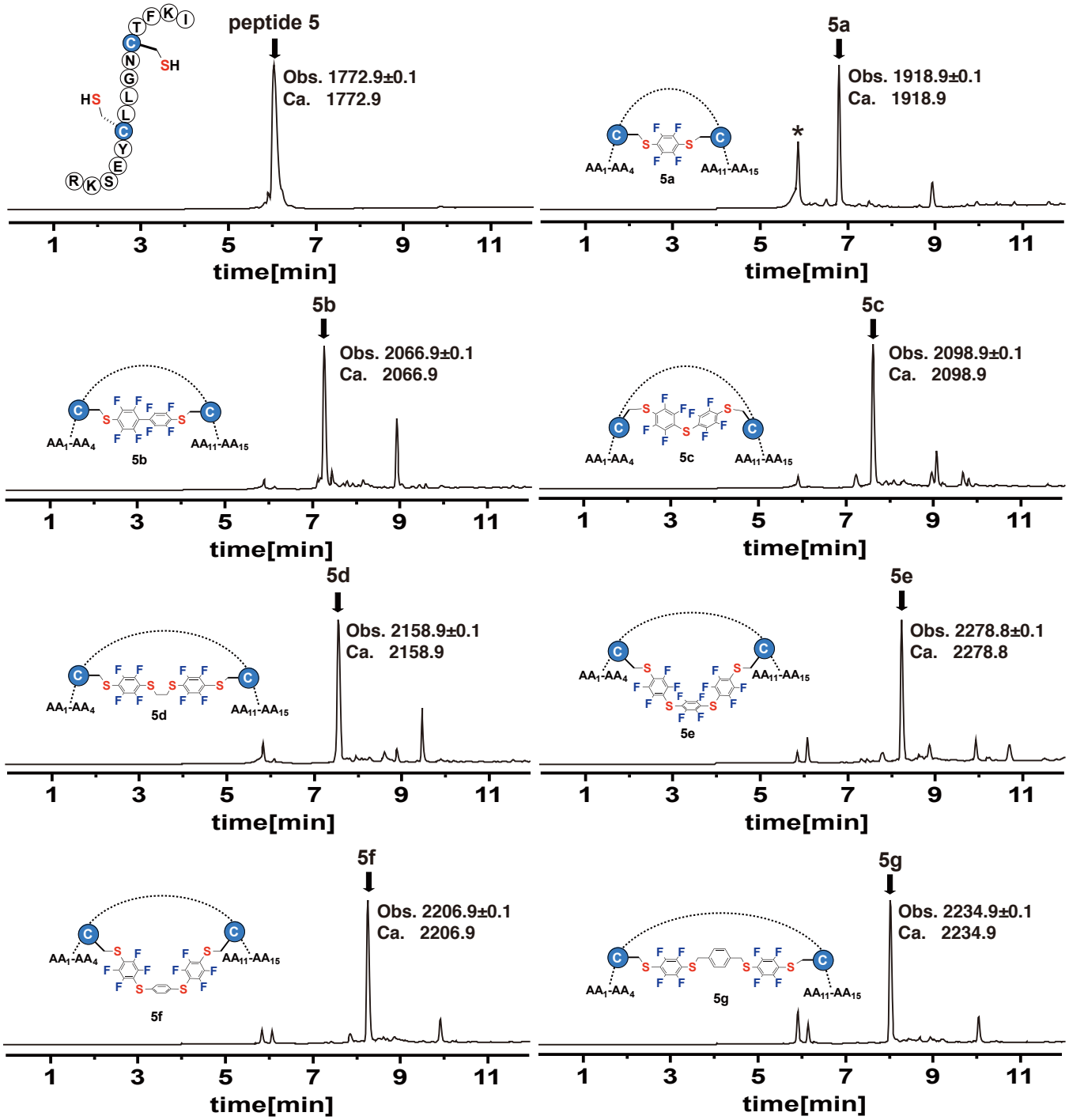


Figure S10. LC-MS chromatograms (total ion current) of purified peptide **5** and corresponding macrocyclization reactions with linkers $\mathbf{L_a}$-$\mathbf{L_g}$ analyzed *in situ*. The peak labeled as * is the oxidized disulfide by-product. Mass (Da) was calculated at highest intensity of the TIC peaks.

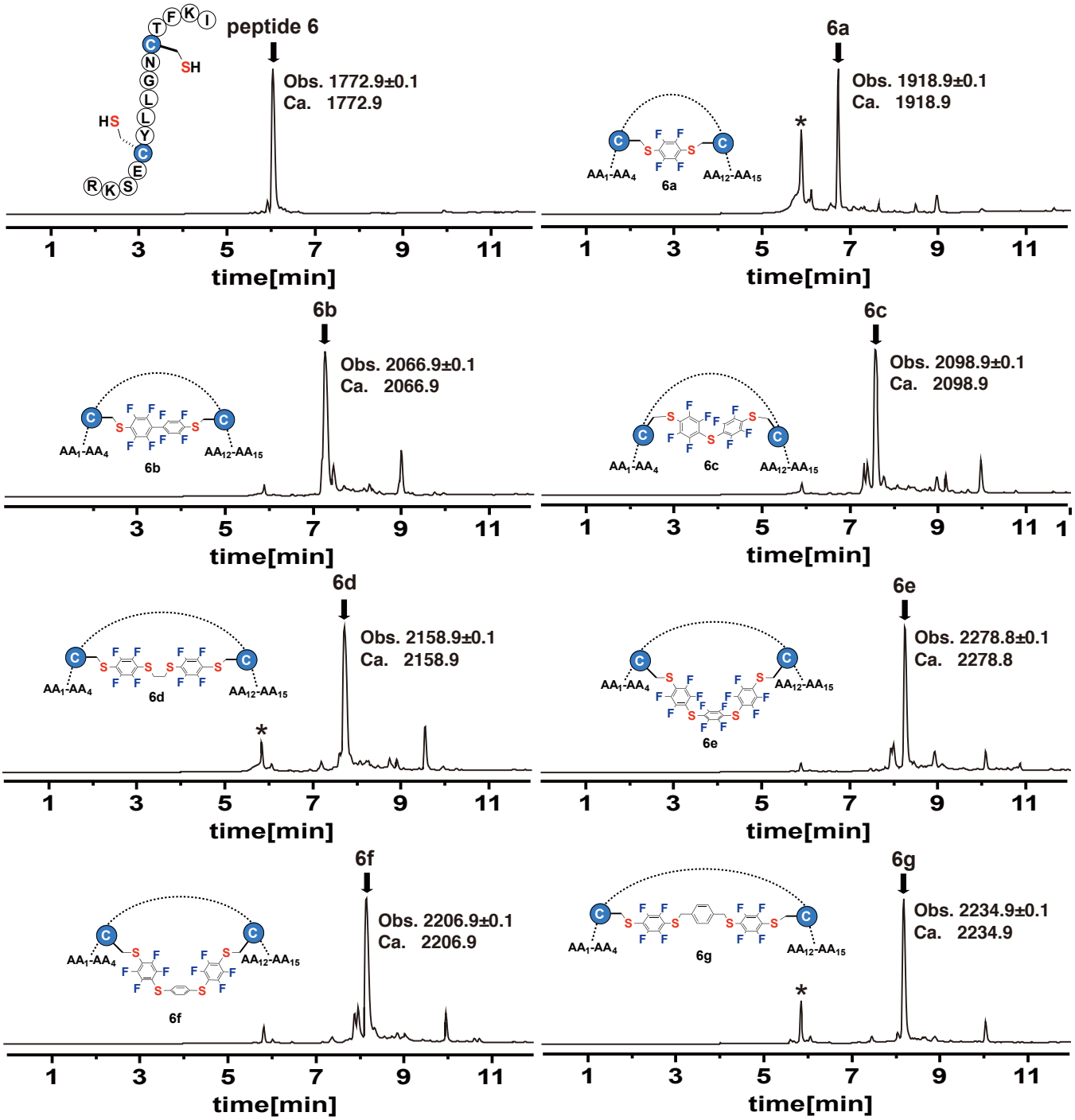


Figure S11. LC-MS chromatograms (total ion current) of purified peptide **6** and corresponding macrocyclization reactions with linkers $\mathbf{L_a}$-$\mathbf{L_g}$ analyzed *in situ*. The peaks labeled as * are the oxidized disulfide by-products. Mass (Da) was calculated at highest intensity of the TIC peaks.

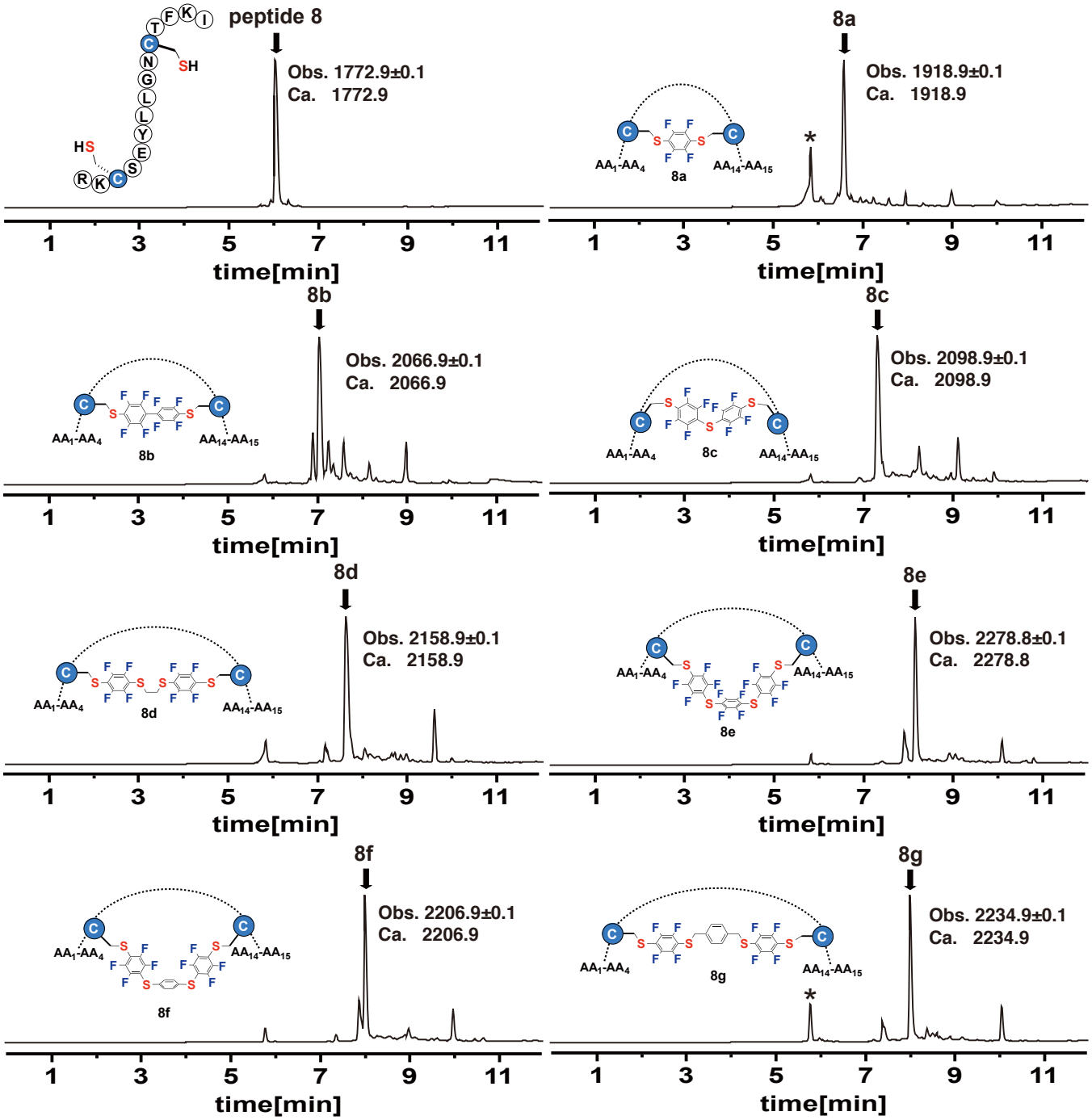


Figure S12. LC-MS chromatograms (total ion current) of purified peptide **8** and corresponding macrocyclization reactions with linkers $\mathbf{L_a}$-$\mathbf{L_g}$ analyzed *in situ*. The peaks labeled as * are the oxidized disulfide by-products. Mass (Da) was calculated at highest intensity of the TIC peaks.

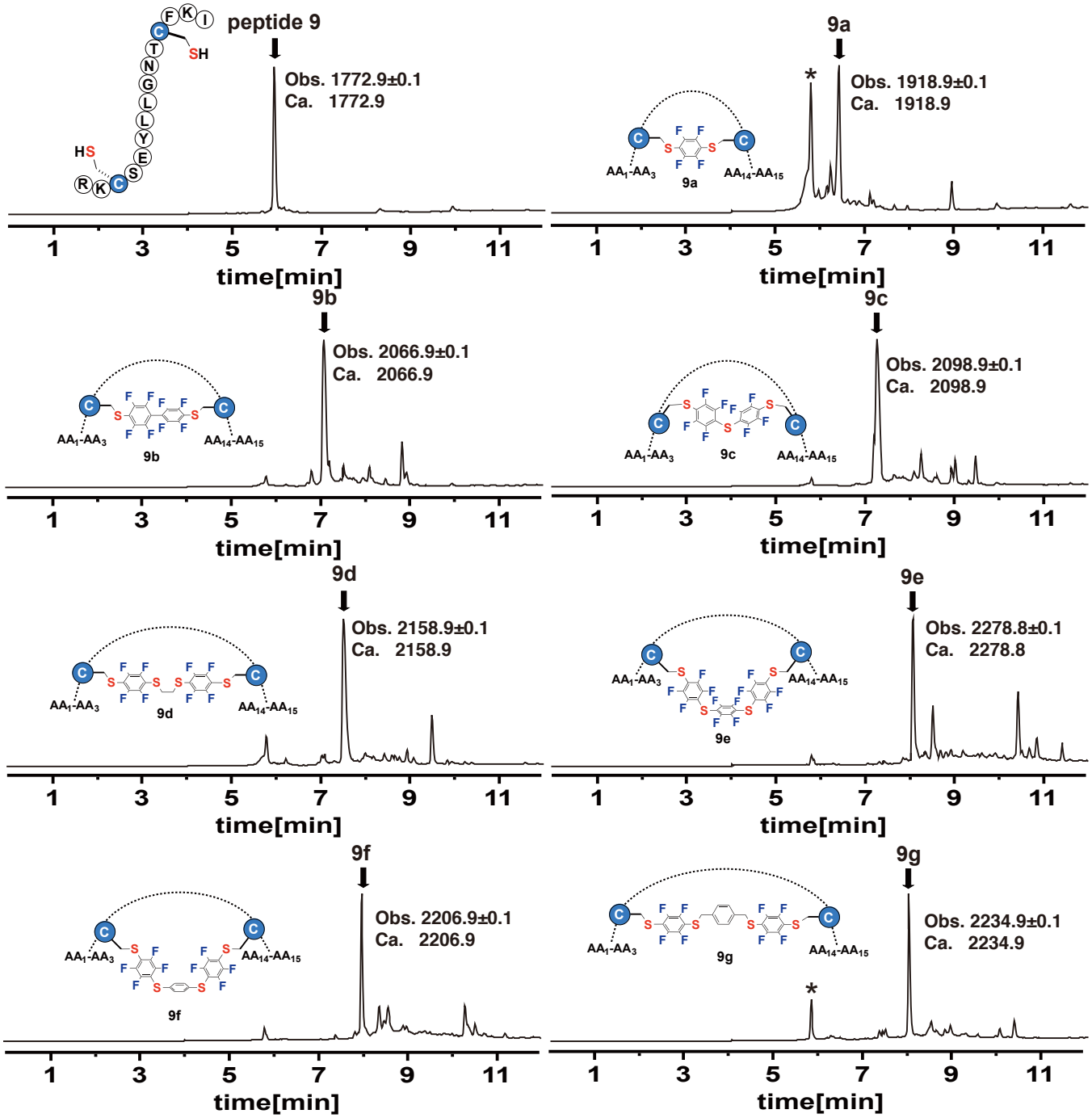


Figure S13. LC-MS chromatograms (total ion current) of purified peptide **9** and corresponding macrocyclization reactions with linkers $\mathbf{L_a}$-$\mathbf{L_g}$ analyzed *in situ*. The peaks labeled as * are the oxidized disulfide by-products. Mass (Da) was calculated at highest intensity of the TIC peaks.

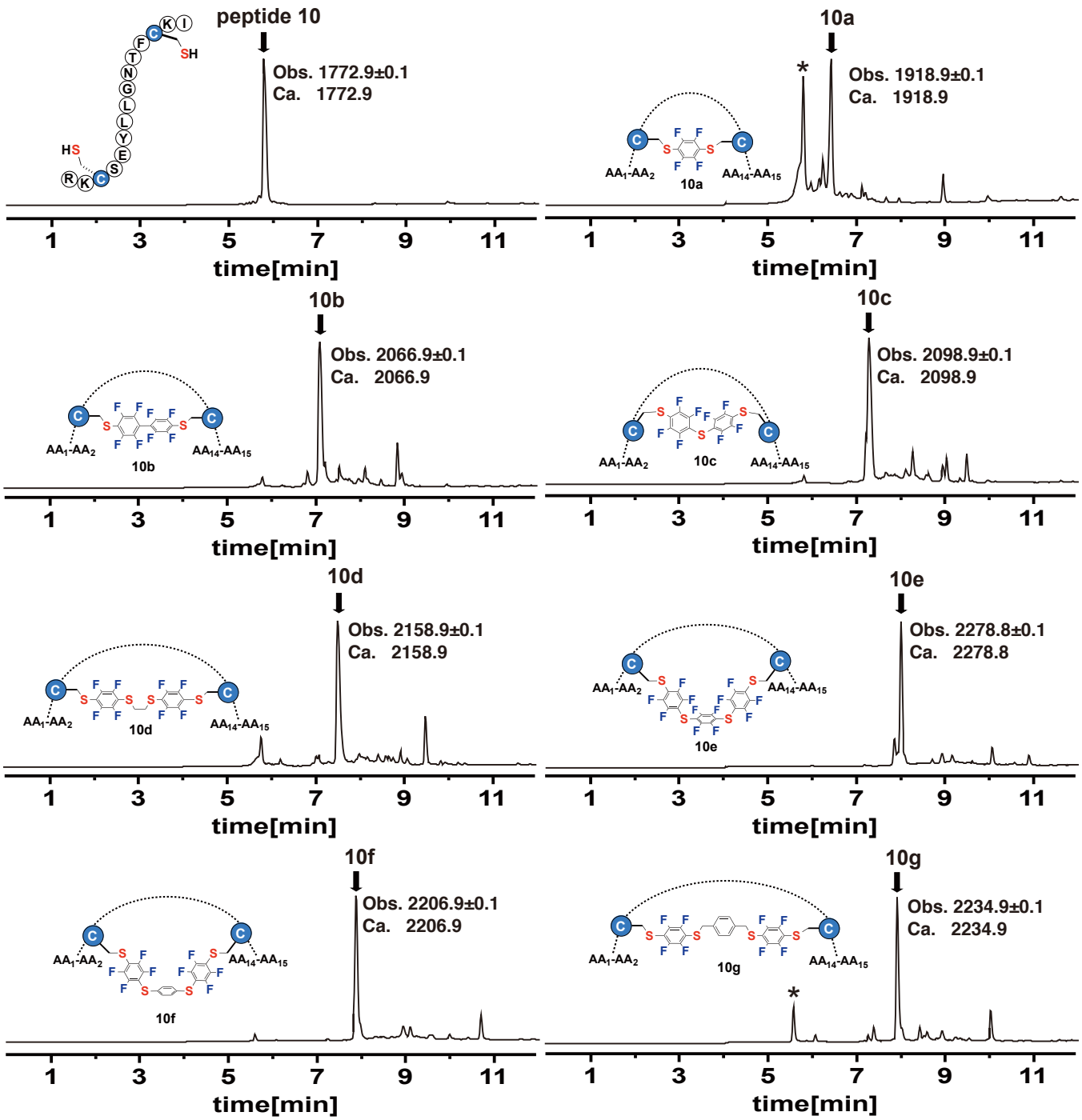


Figure S14. LC-MS chromatograms (total ion current) of purified peptide **10** and corresponding macrocyclization reactions with linkers $\mathbf{L_a}$-$\mathbf{L_g}$ analyzed *in situ*. The peaks labeled as * are the oxidized disulfide by-products. Mass (Da) was calculated at highest intensity of the TIC peaks.

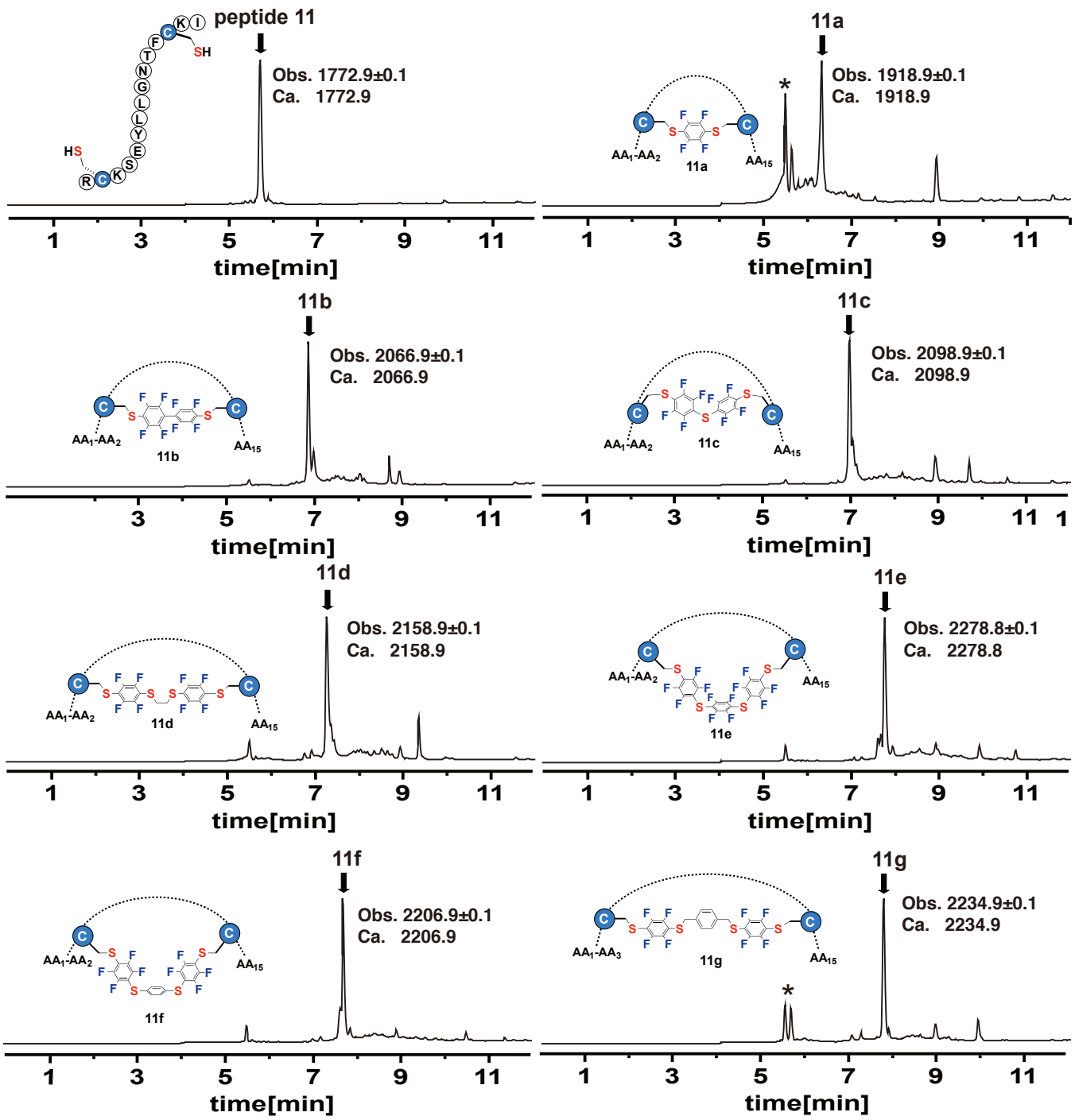


Figure S15. LC-MS chromatograms (total ion current) of purified peptide **11** and corresponding macrocyclization reactions with linkers $\mathbf{L_a}$-$\mathbf{L_g}$ analyzed *in situ*. The peaks labeled as * are the oxidized disulfide by-products. Mass (Da) was calculated at highest intensity of the TIC peaks.

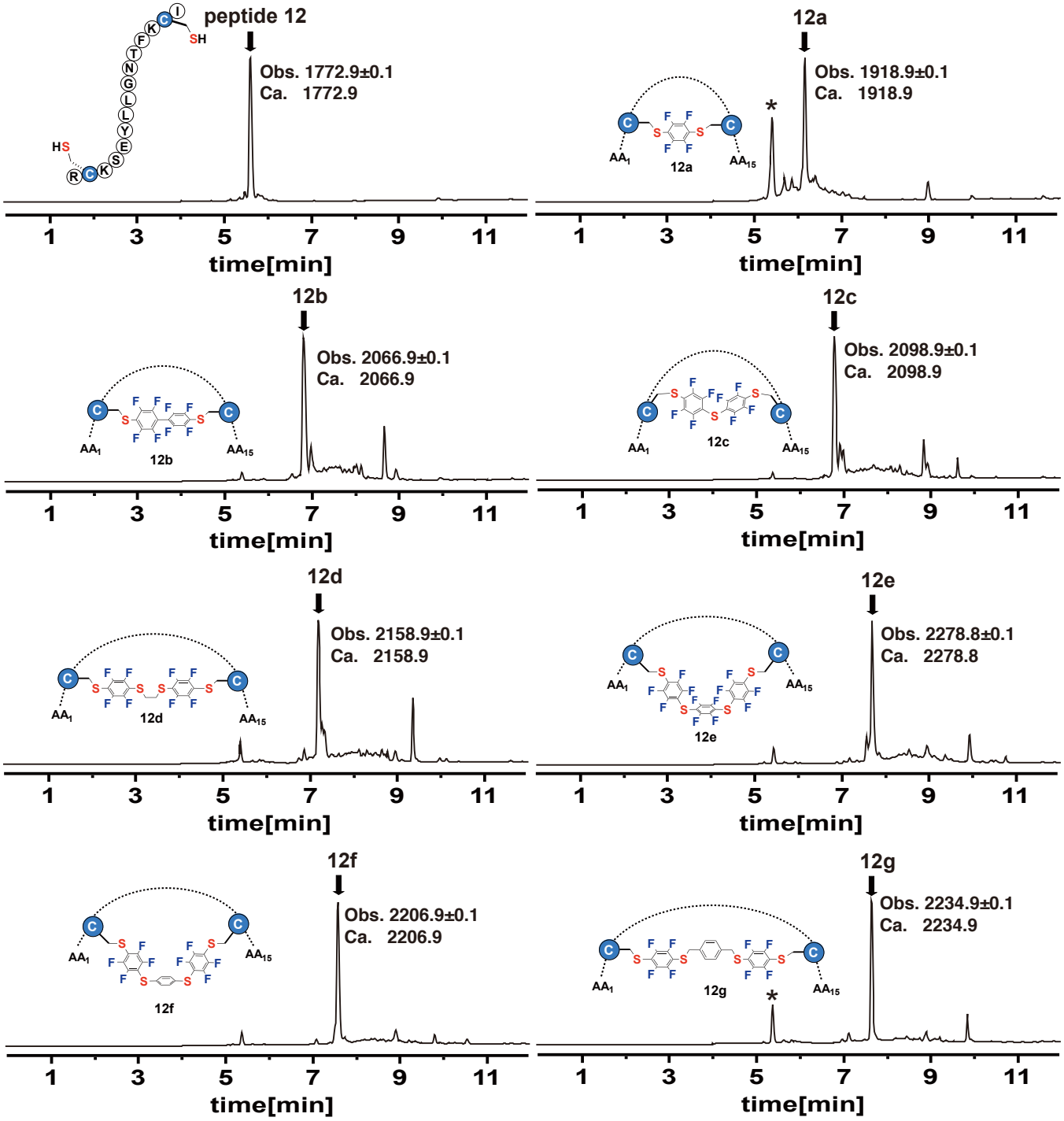


Figure S16. LC-MS chromatograms (total ion current) of purified peptide **12** and corresponding macrocyclization reactions with linkers $\mathbf{L_a}$-$\mathbf{L_g}$ analyzed *in situ*. The peaks labeled as * are the oxidized disulfide by-products. Mass (Da) was calculated at highest intensity of the TIC peaks.

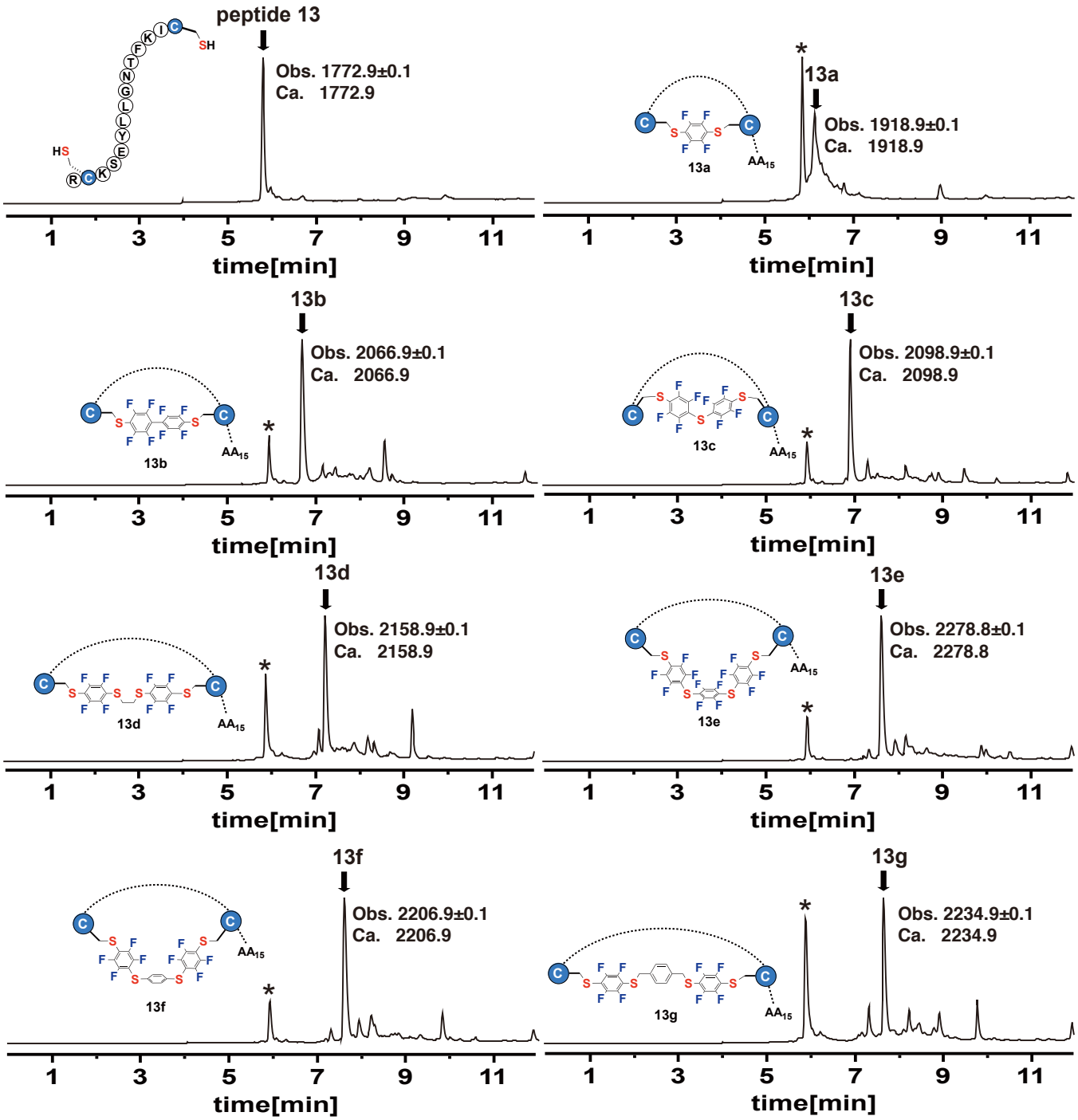


Figure S17. LC-MS chromatograms (total ion current) of purified peptide **13** and corresponding macrocyclization reactions with linkers $\mathbf{L_a}$-$\mathbf{L_g}$ analyzed *in situ*. The peaks labeled as * are the oxidized disulfide by-products. Mass (Da) was calculated at highest intensity of the TIC peaks.

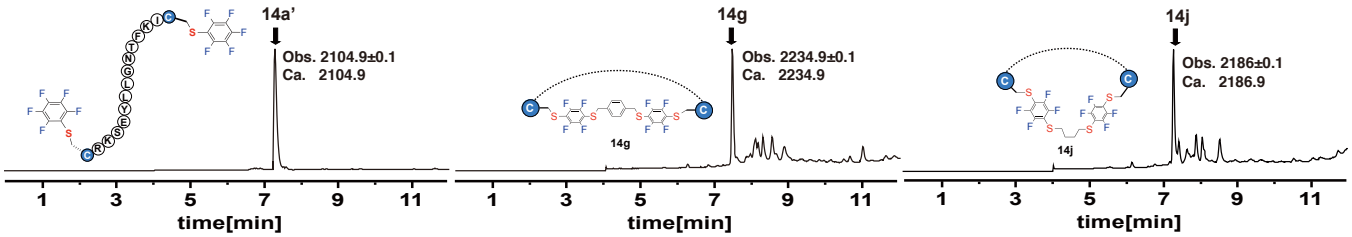


Figure S18. LC-MS chromatograms (total ion current) of purified peptide **14a'**, and corresponding macrocyclization with dithiol linkers analyzed *in situ*. Mass (Da) shown was calculated at the highest intensity of the TIC peaks.

References:

1. Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. *Int. J. Peptide Protein Res.* **1992**, *40*, 180.
2. M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *J. Cheminformatics* **2012**, *4*, 17.
3. Gaussian 09, Revision D.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.
4. L. Johnston, M. E. Peach, *J. Fluor. Chem.* **1978**, *12*, 41—47.