Electronic Supplementary Material (ESI) for ChemComm.

SUPPORTING INFORMATION

for

Sequence-Selective DNA Recognition and Enhanced Cellular Uptake by Peptide-Steroid Conjugates

Yara Ruiz García, Abhishek Iyer, Dorien Van Lysebetten, Y. Vladimir Pabon, Benoit Louage, Malgorzata Honcharenko, Bruno De Geest, C.I. Edvard Smith, Roger Strömberg and Annemieke Madder*

TABLE OF CONTENTS

1	MATERIALS & METHODS	
	1. I. Products	S2
	1. II. Peptide Synthesis	S2
	1. III. Analysis & Equipment	S2
2	SYNTHESIS OF SCAFFOLDS	
	2. I. Steroid template molecules	S3
	2. II. NMR spectra of steroid template molecules	S7
3	PEPTIDE SYNTHESIS	
	3. I. Design of Linear Peptides	S13
	3. II. Manual coupling protocols	S14
	3. III. Automated synthesis of linear peptides	S15
	3. IV. Analysis of linear peptides	S17
	3. V. Synthesis of compound **12**	S21
4	CuAAC CONJUGATION	
	4. I. Compounds for DNA binding	S22
	4. II. Compounds for cell penetration	S30
5	ELECTROPHORESIS MOBILITY SHIFT ASSAY	S33
6	DLS	S39
7	MTT ASSAY	S40
8	CONFOCAL MICROSCOPY	S41
9	FLOW CYTOMETRY	S42
10	REFERENCES	S43

1. Materials and Methods

1. I. Products

All organic solvents were purchased from commercial suppliers and used without further purification or drying. DMF extra dry (with molecular sieves, water < 50 ppm) was acquired from ACROS Organics. DMF an NMP (peptide synthesis grade) were purchased from Biosolve. HPLC grade quality Hexane and Chloroform were purchased from Fisher Scientific. Ethyl Acetate, Acetonitrile, Methanol, Diethyl Ether DIPEA, supplied as extra dry, redistilled, 99.5 % pure and Triisopropylsilane were purchased from Sigma Aldrich. Deuterated solvents $CDCl_3$ (99.8% atom D) and CD_3OD (99.8% atom D) were obtained from Sigma-Aldrich. Water with the Milli-Q grade standard was obtained in-house either from a Millipore ROs 5 purification system or a Sartorius Arium 611 DI. Rink-Amide ChemMatrix (100-200 µm, manufacturer's loading: 0.52-0.54 mmol/g) was obtained from Biotage. All reagents were acquired from commercial sources and used without prior purification.

Fmoc-Propargylglycine-OH, Fmoc-Azidohomoalanine-OH, PyBOP and HBTU coupling reagents were obtained from either Merck Novabiochem or IRIS Biotech GmbH, while HATU (purum ≥ 98.0 %) was acquired from Fluka. TFA was purchased from Iris Biotech GmbH. NHS-Fluorescein was purchased from Thermo Scientific. The Nα-Fmoc protected amino acids were purchased at Merck Novabiochem, IRIS Biotech GmbH and Fluka, or supplied by MultiSynTech GmbH. All chiral α-amino acids used in this paper possessed the L configuration. Throughout this work, residues with standard acid-sensitive side-chain PGs were used: Cys(Trt) [C], Asp(OtBu) [D], Glu(OtBu) [E], Lys(Boc) [K], Asn(Trt) [N], Gln(Trt) [Q], Arg(Pbf) [R], Ser(tBu) [S], Thr(tBu) [T].

All oligonucleotides used were commercially purchased from Eurogentec (HPLC purified using RP-cartridge-Gold, 200 nm scale) and were used as such.

1. II. Peptide Synthesis

Automated peptide syntheses were performed on a fully-automated SYRO Multiple Peptide Synthesizer robot, equipped with a vortexing unit for the 24-reactor block (MultiSynTech GmbH). Reactions were open to the atmosphere and were executed at ambient temperature.

1. III. Analysis & Equipment

Reversed-Phase HPLC analysis and purification was performed on an Agilent 1100 Series instrument with diode array detector (at 214, 254, 280, 310, 360 nm), equipped with a Phenomenex Luna C18(2) 100 Å column (250 x 4.6 mm, 5 µm, at 35 °C) or a Phenomenex Jupiter C4 300 Å column (250 x 4.6 mm, 5 µm, at 35 °C). Linear gradient elutions were performed by flushing 2 min with A followed by 0 to 100% buffer B in 15 minutes and finally by a 5 min flushing with B using a binary solvent system composed of buffer A: 0.1% TFA in H_2O and B: MeCN with a flow of 1.0 mL/min at 35°C. MALDI-TOF-MS data were acquired on an Applied Biosystems Voyager-DE STR Biospectrometry Workstation, equipped with a high performance nitrogen laser (337 nm). All spectra were recorded in the positive and reflector mode, with delayed extraction. LC-TIC-MS data (reversed phase) were recorded on an Agilent 1100 Series instrument with diode array detector (at 214, 254, 280, 310, 360 nm), equipped with a Phenomenex Kinetex C18 100 Å (150 x 4.6 mm, 5 µm, at 35 °C), hyphenated to an Agilent ESI-single quadrupole MS detector type VL. Mass detection operated in the positive mode. Linear gradient elutions were performed by flushing 0.5 min with A followed by 0 to 100% buffer B in 6 minutes and finally by a 2 min flushing with B using a binary solvent system composed of buffer A: 0.1% HCOOH in H_2O and B: MeCN with a flow of 1.0 mL/min at 35°C.. Pre and post flushing is included in the LC-MS and RP-HPLC

results. A solution of 4-5 mg α-cyano-4-hydroxycinnamic acid in 500µl MeCN, 490µl mQ, 10µl 1M ammoniumcitrate, 1µl TFA was used as matrix for MALDI-TOF-MS. Concentrations were determined by ^{1}H-NMR ERETIC (Electronic REference To access In vivo Concentrations)[1] with a high-resolution 500 MHz NMR-spectrometer.

NMR spectra were recorded at room temperature on a Bruker Avance-300 spectrometer at 300 MHz for ^{1}H NMR and on a Bruker Avance-500 spectrometer at 500 MHz for ^{1}H NMR and 125 MHz for ^{13}C spectra. Chemical shifts (δ units) are expressed in parts per million (ppm) relative to tetramethylsilane (TMS) and the internal solvent peak was used for calibration. Wherever peak multiplicities have been reported, the following abbreviations have been used: s, singlet; d, doublet; t, triplet; m, multiplet; br, broad. Coupling constants (J values) are expressed in Hertz (Hz). The Attached Proton Test (APT) technique was used to assign ^{13}C peaks (C, CH, CH_2, CH_3). All ^{1}H spectra of deoxycholic acid derived products contain a region of high signal overlap between 1.00 and 2.00 ppm. Therefore, the detailed analysis of this envelop is almost impossible and often not mentioned or reported as a series of multiplets. Fortunately, in the analysis of the synthesized scaffolds all necessary information was shifted downfield from this region.

2. Synthesis of Scaffolds

2. I. Steroid template molecules

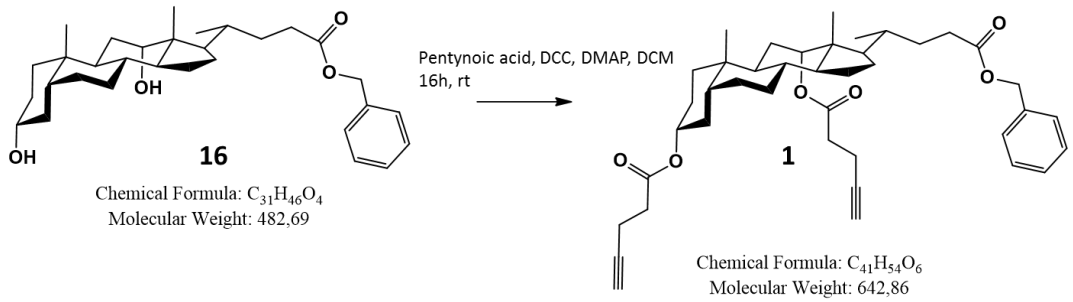
Benzyl 3α, 12α-dihydroxy-5β-cholan-24-oate. To a solution of deoxycholic acid (**15**) (3 g, 7.642 mmol) in DMF (7 mL), DBU (1.38 mL, 9.25 mmol) and benzyl bromide (1.1 mL, 9.25 mmol) were added and the mixture was stirred at 50 °C for 24 h. DMF was removed under vacuum, 10 % HCl (80 mL) was added to the residue and the product was extracted with CH_2Cl_2 (100 mL). The organic layer was washed with water (100 mL) and aq. $NaHCO_3$ (100 mL) and dried over anhyd. Na_2SO_4. The crude product obtained after removal of the solvent was purified using column chromatography on silica gel using 5 % EtOAc/$CHCl_3$ (Rf: 0.38) to yield 78 % of compound **16** as a white solid (2.81 g). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 7.36 − 7.38 (m, 5H), 5.15 (d, 1H, J = 14.13 Hz), 5.11 (d, 1H, J = 14.13 Hz), 3.92 (s, 1H), 3.60 (br, m, 1H), 2.45 – 2.36 (m, 1H), 2.30 – 2.20 (m, 1H), 0.90 (d, J = 6.28 Hz, 3H), 0.80 (s, 3H), 0.63 (s, 3H). ES-MS m/z (% rel. int.) calcd. 482.3 (100), found 541.3 (100) [M + OAc^-].

Chemical Formula: $C_{24}H_{40}O_4$
Molecular Weight: 392,57

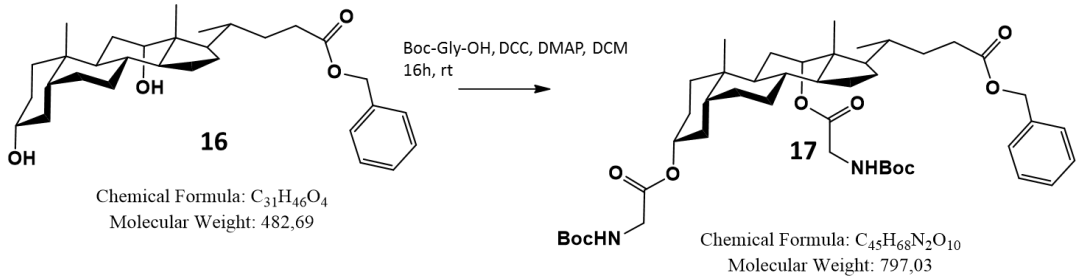
Chemical Formula: $C_{31}H_{46}O_4$
Molecular Weight: 482,69

Benzyl 3α, 12α-dipentynoate-5β-cholan-24-oate. Compound **16** (50 mg, 0.1037 mmol) was dissolved in 1.5 mL dry DCM. Then a solution of pentynoic acid (30.52 mg, 0.311 mmol), DCC (64.21 mg, 0.311 mmol) and DMAP (50.7 mg, 0.415 mmol) in dry DCM (2.5 mL) was added. The reaction mixture was stirred for 16h at room temperature. The precipitate formed was filtered off and the solvent was evaporated under vacuo. The residue was purified by column chromatography (silica gel EtOAc/Hexane 0.6:1; Rf: 0.60) to give compound **1** as a yellow oil (64.6 mg, 97% yield). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 7.38 − 7.36 (m, 5 H), 5.15 (s, H), 5.15 (d, 1H, J = 13.56 Hz), 5.11 (d, 1H, J = 13.75 Hz), 4.6 (br, m, 1 H), 2.55-2.45 (m, 4 H), 2,43-2.40 (m, 4 H), 2.35 – 2.33 (m, 1 H), 2.30 – 2.20 (m, 1 H), 2.21 (t, J = 12 Hz, 2 H), 0.85 (s, 3 H), 0.75 (d, J = 6.40 Hz, 3 H), 0.65 (s, 3 H). ^{13}C-NMR (125 MHz, $CDCl_3$): δ = 173.90 (COOR-C24), 171.2 (COOR-C3), 170.87 (COOR-C12), 136.03 (C - benzyl group), 128.53 (2CH-benzyl ring), 128.25 (2CH- benzyl ring), 128.20 (CH-benzyl ring), 82.69 (C alkyne), 82.67 (C alkyne), 76.36 (CH ester), 74.57 (CH ester), 69.37 (CH_2), 68.95 (CH2), 66.13 (CH_2 benzyl ester), 49.34 (CH), 47.61 (CH), 45.10 (C), 35.63 (CH), 34.80 (CH), 34.70 (CH), 34.69 (CH_2), 34.50 (CH), 34.45 (CH), 34.31 (CH), 34.05 (C), 33.91 (CH_2), 33.64 (CH_2), 32.16 (CH_2), 31.24 (CH_2), 30.80 (CH_2), 27.34 (CH_2), 27.10 (CH_2), 26.83 (CH_2),

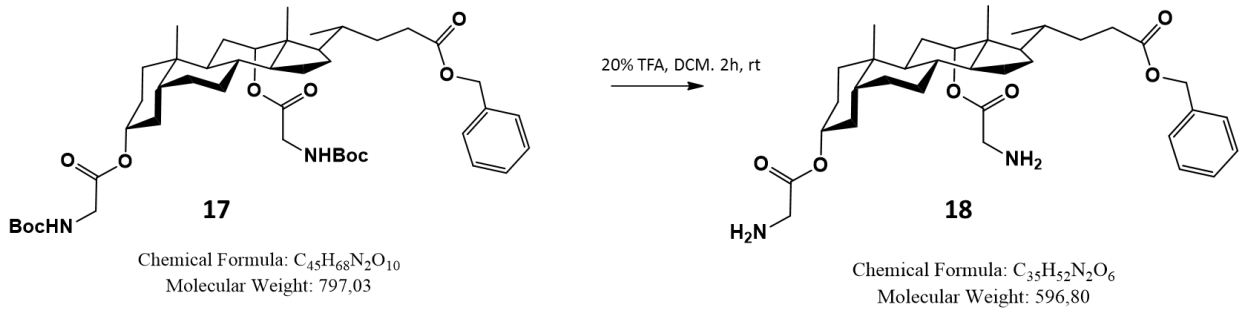
26.59 (CH_2), 23.44 (CH_2), 22.98 (19-CH_3), 17.50 (21 – CH_3), 14.42 (CH_2), 12.32 (18-CH_3). ES-MS m/z (% rel. int.) calcd. 642.4 (100), found 660.4 (100) [M+ $NH4^+$]. HR-MS (ES) m/z calcd. for $C_{41}H_{54}O_6$ 642.3920, found 660.4268 (M + $NH4^+$; Δ = 1.2 ppm).



Benzyl 3α, 12α-Gly-5β-cholan-24-oate. Compound **16** (50 mg, 0.1037 mmol) was dissolved in 1.5 mL dry DCM. Then a solution of Boc protected glycine (54.52 mg, 0.311 mmol), DCC (64.21 mg, 0.311 mmol) and DMAP (50.7 mg, 0.415 mmol) in dry DCM (2.5 mL) were added. The reaction mixture was stirred for 16 h at room temperature. The precipitate formed was filtered off and the solvent is evaporated under vacuo. The residue was purified by column chromatography (silica gel EtOAc/Hexane 0.6:1; Rf: 0.45) to give **17** as a yellow oil (76.8 mg, 93% yield). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 7.36 − 7.38 (m, 5 H), 5.2 (s, 1H), 5.15 (m, 2H), 4.6 (br, m, 1H), 3.87 (m, 4H), 2.35 – 2.33 (m, 1H), 2.30 – 2.20 (m, 1H), 2.21 (t, J = 12 Hz, 2H), 1.40 (s, 18H), 0.85 (s, 3H), 0.75 (d, J = 6.40 Hz, 3H), 0.65 (s, 3H). ES-MS m/z (% rel. int.) calcd. 796.5 (100), found 819.3 (100) [M+ Na^+].

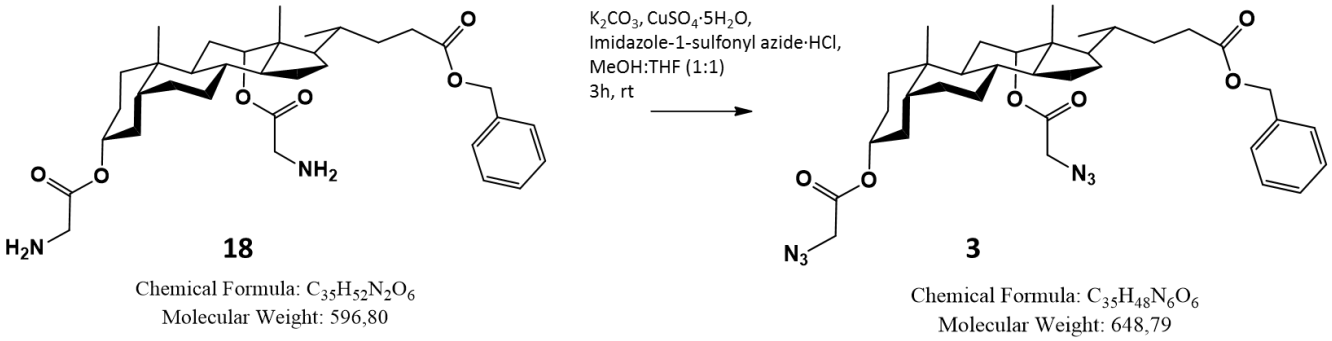


Compound 5 (Boc deprotection). Compound **17** (76.5 mg, 0.096 mmol) was treated with 10 mL of a solution of TFA (30%) in DCM for 2 h at room temperature. The solvent was evaporated and the white powder obtained was analyzed without further purification (Rf: 0.1, 56.7 mg, 99% yield). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 7.36 − 7.38 (m, 5H), 5.2 (s, 1H), 5.15 (m, 2H), 4.6 (br, m, 1H), 3.87 (m, 4 H), 2.35 – 2.33 (m, 1H), 2.30 – 2.20 (m, 1H), 2.21 (t, J = 12 Hz, 2H), 1.40 (s, 18H), 0.85 (s, 3H), 0.75 (d, J = 6.40 Hz, 3H), 0.65 (s, 3H). ES-MS m/z (% rel. int.) calcd. 596.4 (100), found 597.3 (100) [M+ H^+].



Compound 3. Diazotransfer reaction. Boc deprotected compound **18** (57 mg, 0.096 mmol) was dissolved in 2 mL MeOH/THF (1:1) and treated with K_2CO_3 (53 mg, 0.386 mmol) followed by $CuSO_4{\cdot}5H_2O$ (0.725 mg, 0.0029 mmol). To this mixture, imidazole-1-sulfonyl azide·HCl (61 mg, 0.290 mmol) dissolved in 1 mL MeOH/THF (1:1) was added to the

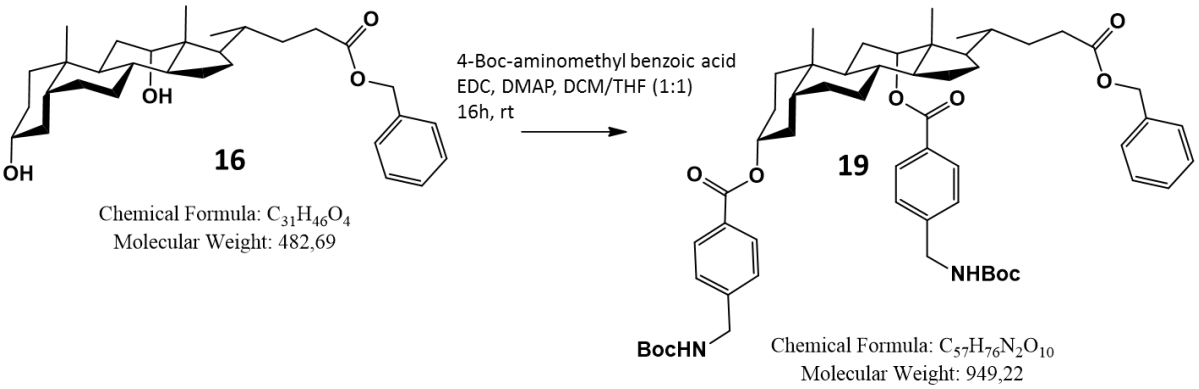
reaction mixture and stirred at room temperature for 3 h under argon. The reaction was quenched with 4.5 mL of water and 0.5 mL of AcOH was added. Then, the reaction mixture was extracted 2 times with hexane (5 mL). The organic phase was dried under $MgSO_4$. The solvent was evaporated and the residue purified by column chromatography (silica gel DCM/MeOH 9.5:0.5; Rf: 0.20) to get a yellowish oil (59.7 mg, 96% yield). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 7.36 − 7.38 (m, 5H), 5.2 (s, 1H), 5.15 (d, 1H, J = 13.75 Hz), 5.10 (d, 1H, J = 13.94 Hz), 4.6 (br, m, 1H), 3.98-3.90 (m, 4H), 2.35 – 2.33 (m, 1H), 2.30 – 2.20 (m, 1H), 2.21 (t, J = 12 Hz, 2H), 1.40 (s, 18H), 0.85 (s, 3H), 0.75 (d, J = 6.40 Hz, 3H), 0.65 (s, 3H). ^{13}C-NMR (125 MHz, CDCl3): δ = 173.85 (COOR-C24), 167.76 (COOR-C3), 167.47 (COOR-C12), 135.99 (C - benzyl group), 128.52 (2CH-benzyl ring), 128.25 (2CH- benzyl ring), 128.19 (CH-benzyl ring), 77.97 (CH-C3), 75.97 (CH-C12), 66.14 (CH_2 benzyl ester), 50.91 (CH_2 ester glycine C3), 50.51 (CH_2 ester glycine C12), 47.64 (CH), 45.22 (CH), 45.06 (C), 41.67 (CH), 35.53 (CH), 33.95 (C), 31.93 (CH2), 31.17 (CH_2), 30.70 (CH_2), 29.66 (CH_2), 27.24 (CH_2), 26.90 (CH_2), 26.70 (CH_2), 25.90 (CH_2), 25.77 (CH), 25.63 (CH), 25.40 (CH_2), 23.34 (CH_2), 22.85 (19-CH_3), 17.61 (21 – CH_3), 12.32 (18-CH_3). ES-MS m/z (% rel. int.) calcd. 648.4 (100), found 666.4 (100) [M+$NH4^+$]. HR-MS (ES) m/z calcd. for $C_{35}H_{48}N_6O_6$ 648.3635, found 666.3968 (M + $NH4^+$; Δ = 0.9 ppm).



Chemical Formula: $C_{35}H_{52}N_2O_6$
Molecular Weight: 596,80

Chemical Formula: $C_{35}H_{48}N_6O_6$
Molecular Weight: 648,79

Compound 19. (esterification with boc-aminomethylbenzoic acid). 4-Boc-aminomethyl benzoic acid (417.4 mg, 1.6596 mmol) was dissolved in 4 mL dry DCM/dry THF (1:1) at room temperature. EDC (0.294 mL, 1.6596 mmol) was added in portions and stirred for 30 min. DMAP (405.5 mg, 3.319 mmol) was added to the reaction mixture. Compound **15** (200 mg, 0.4149 mmol) dissolved in 6 mL dry DCM was added and the reaction was stirred overnight. The solvent was evaporated and the residue purified by column chromatography (silica gel EtOAc/Hexane 3:1; Rf: 0.57). 82 % yield (322.7 mg). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 8.1 (d, J = 7.75 Hz, 2H), 7.9 (d, J = 8.63 Hz, 2H), 7.45 (d, J = 8,2 Hz, 2H), 7.38 (d, J = 8 Hz, 2H), 7.36 − 7.38 (m, 5H), 5.3 (s, 1H), 5.15 (m, 2H), 4.7 (br, m, 1H), 4.3 (br m, 4H), 2.35 – 2.33 (m, 1H), 2.30 – 2.20 (m, 1H), 2.21 (t, J = 12 Hz, 2H), 2,00 (s, 18H), 0.85 (s, 3H), 0.75 (d, J = 6.40 Hz, 3H), 0.65 (s, 3H). ES-MS m/z (% rel. int.) calcd. 948.5 (100), found 1007.5 (100) [M + OAc^-].

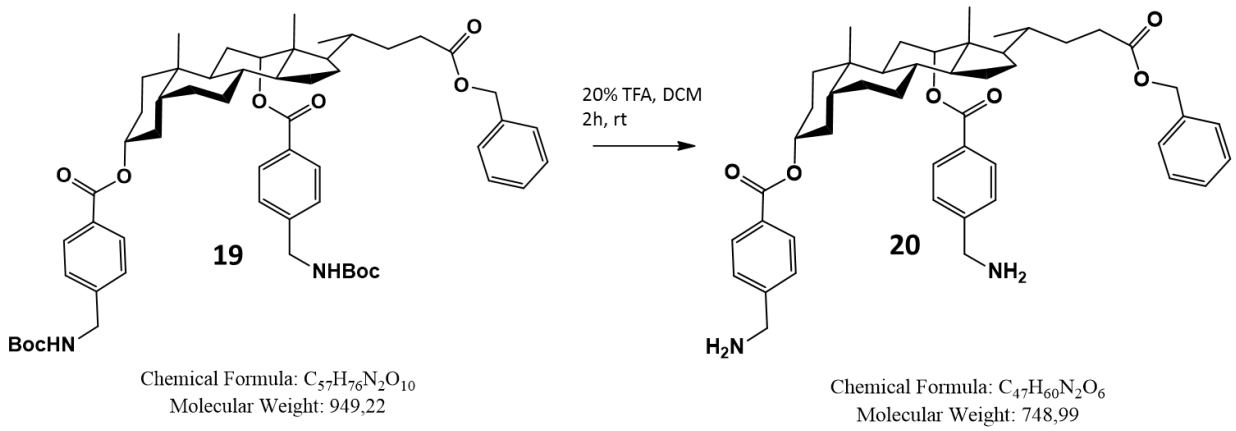


Chemical Formula: $C_{31}H_{46}O_4$
Molecular Weight: 482,69

Chemical Formula: $C_{57}H_{76}N_2O_{10}$
Molecular Weight: 949,22

Compound 20 (Boc deprotection). Compound **19** (55 mg, 0.058 mmol) was treated with 10 mL of a solution of TFA (30%) in DCM for 2 h at room temperature. The solvent was evaporated and the white powder obtained was analyzed without further purification (Rf: 0.12, 43 mg, 99% yield). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 8.1 (d, J = 7.75 Hz, 2H), 7.9 (d, J

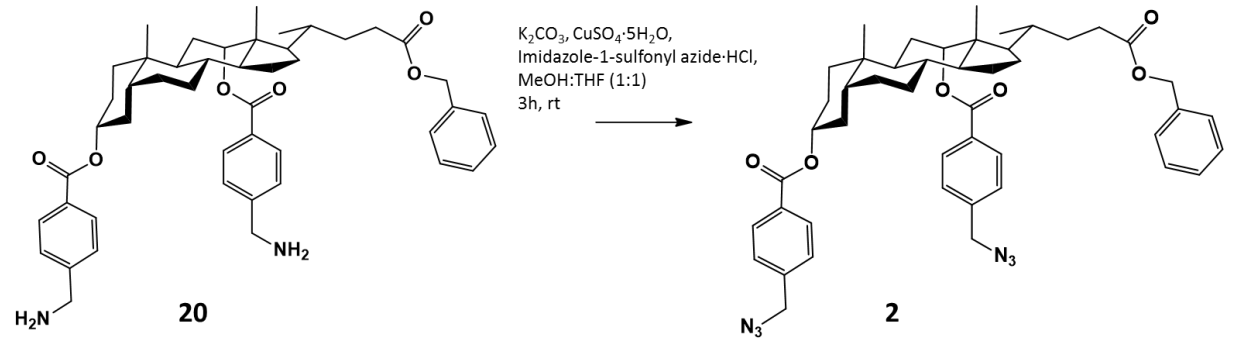
= 8.63 Hz, 2H), 7.45 (d, J = 8,2 Hz, 2H), 7.38 (d, J = 8 Hz, 2H), 7.36 − 7.38 (m, 5H), 5.3 (s, 1H), 5.15 (m, 2H), 4.7 (br, m, 1H), 4.3 (m, 2H), 4.1-3.9 (m, 2H), 2.35 – 2.33 (m, 1H), 2.30 – 2.20 (m, 1H), 2.21 (t, J = 12 Hz, 2H), 0.85 (s, 3H), 0.75 (d, J = 6.40 Hz, 3H), 0.65 (s, 3H). ES-MS m/z (% rel. int.) calcd. 748.5 (100), found 749.5 (100) [M + H^+].



Chemical Formula: $C_{57}H_{76}N_2O_{10}$
Molecular Weight: 949,22

Chemical Formula: $C_{47}H_{60}N_2O_6$
Molecular Weight: 748,99

Compound 2. Diazotransfer reaction. Compound **20** (50 mg, 0.067 mmol) was dissolved in 2 mL MeOH/THF (1:1) and treated with K_2CO_3 (36.9 mg, 0.267 mmol) followed by $CuSO_4{\cdot}5H_2O$ (0.5 mg, 0.002 mmol). To this mixture, imidazole-1-sulfonyl azide·HCl (54 mg, 0.2 mmol) dissolved in 1 mL MeOH/THF (1:1) was added to the reaction mixture and stirred at room temperature for 3h under argon. The reaction was quenched with 4.5 mL of water and 0.5 mL of AcOH was added. Then, the reaction was extracted 2 times with hexane (5 mL). The organic phase was dried under $MgSO_4$. The solvent was evaporated and the residue purified by column chromatography (silica gel EtOAc/Hexane 1:1; Rf: 0.65) to obtain a yellow oil (51.5 mg, 96% yield). ^{1}H-NMR (300 MHz, $CDCl_3$); δ 8.1 (d, J = 7.75 Hz, 2H), 7.9 (d, J = 8.63 Hz, 2H), 7.45 (d, J = 8,2 Hz, 2H), 7.38 (d, J = 8 Hz, 2H), 7.36 − 7.38 (m, 5H), 5.3 (s, 1H), 5.18 (d, 1H, J = 13.75 Hz), 5.15 (d, 1H, J = 13.75 Hz), 4.7 (br, m, 1 H), 4.40 (s, 2H), 4.38 (s, 2H), 2.35 – 2.33 (m, 1H), 2.30 – 2.20 (m, 1H), 2.21 (t, J = 12 Hz, 2H), 0.85 (s, 3H), 0.75 (d, J = 6.40 Hz, 3H), 0.65 (s, 3H). ^{13}C-NMR (125 MHz, CD_3OD): δ = 173.88 (COOR-C24), 165.47 (COOR-C3), 165.40 (COOR-C12), 140.45 (C), 140.30 (C), 140.13 (C), 136.04 (C-benzyl ester), 130.76 (C), 130.01 (2CH), 129.97 (2CH), 128.52 (2CH), 128.22 (2CH), 128.17 (CH), 127.95 (2CH), 127.79 (2CH), 76.67 (CH ester linker C3), 74.83 (CH ester linker C12), 66.09 (CH_2 benzyl ester), 54.26 (CH_2 ester linker C3), 54.23 (CH_2 ester linker C12), 50.14 (CH), 48.00 (CH), 45.51 (C), 41.77 (CH), 35.80 (CH), 35.74 (CH), 34.72 (CH), 34.05 (C), 32.29 (CH_2), 31.24 (CH_2), 30.75 (CH_2), 30.10 (CH_2), 27.37 (CH_2), 26.82 (CH_2), 26.47 (CH_2), 25.99 (CH_2), 25.89 (CH_2), 23.50 (CH_2), 23.05 (19-CH3), 17.46 (21-CH3), 12.55 (18-CH3). (ES-MS m/z (% rel. int.) calcd. 800.4 (100), found 818.4 (100) [M+ $NH4^+$]. HR-MS (ES) m/z calcd. for $C_{47}H_{56}N_6O_6$ 800.42613, found 818.4592 (M + $NH4^+$; Δ = 0.9 ppm).

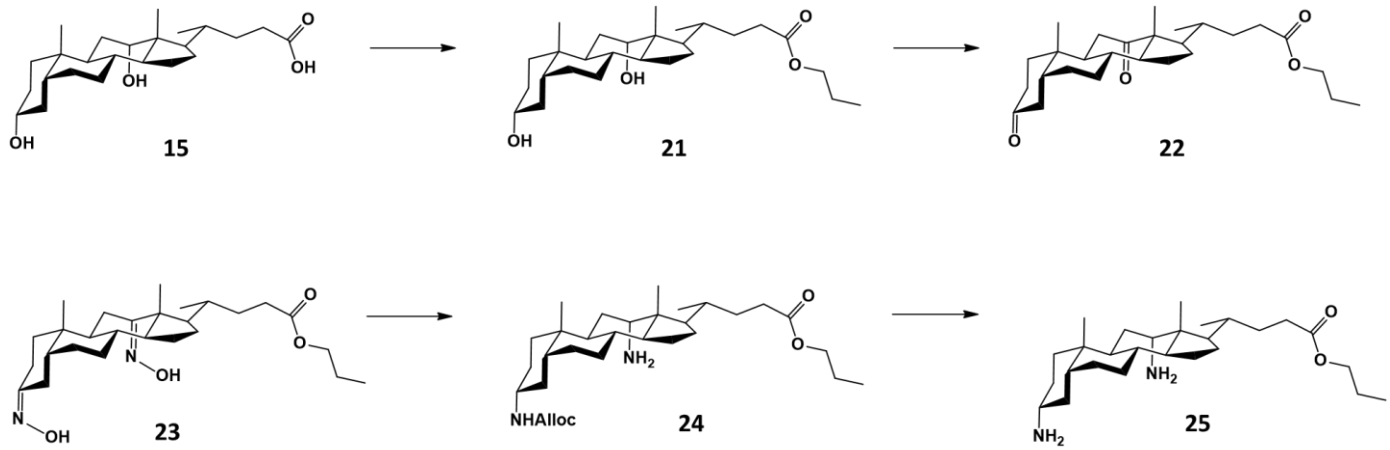


Chemical Formula: $C_{47}H_{60}N_2O_6$
Molecular Weight: 748,99

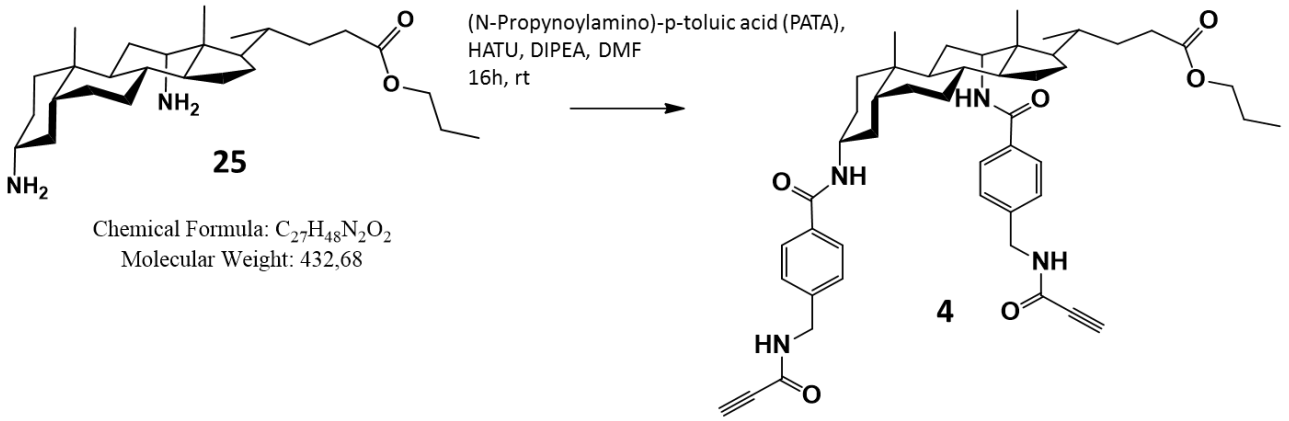
Chemical Formula: $C_{47}H_{56}N_6O_6$
Molecular Weight: 800,98

**** Diamino scaffold 26 was synthesized according to reference 2 and 3** [2,3]

**** (N-Propynoylamino)-p-toluic acid (PATA) was synthesized according as described** [4]



Compound 4. Coupling of PATA linker. Compound **25** (100 mg, 0.231 mmol) was dissolved in 3 mL dry DMF. Then, a solution of HATU (352 mg, 0.926 mmol) and DIPEA (0.161 mL, 0.926 mmol) in dry DMF was added and stirred for 10 min. Next, a solution of PATA (188 mg, 0.926 mmol) in dry DMF (4 mL) was added. The reaction mixture was stirred for 16 h at room temperature. The solvent was evaporated under vacuo. The residue was purified by column chromatography (silica gel EtOAc/Hexane 0.6:1; Rf: 0.17) to give **4** as a white powder (151 mg, 82% yield). ^{1}H-NMR (500 MHz, CD_3OD); δ 7.8 (d, J = 6.86 Hz, 2 H), 7.65 (d, J = 6.66 Hz, 2 H), 7.35 (m, 4H), 7.2 (t, 2 H), 4.40 (m, 4 H), 4,00 (t, J = 6.5 Hz, 2 H), 3.65 (s, 1 H), 3.60 (s, 1 H) 2.35–2.10 (m, 2 H), 1,2 (s, 3 H), 0.90 (s, 3 H), 0.85 (d, J = 6.40 Hz, 3 H), 0.80 (s, 3 H). ^{13}C-NMR (125 MHz, CD_3OD): δ = 175.93 (COOR-C24), 169.69 (CONH-C3), 169.16 (CONH-C12), 154.65 (2CONH linker), 142.85 (C), 142.76 (C), 135.49 (C), 135.01 (C), 131.03 (CH), 130.78 (CH), 129.83 (CH), 128.82 (CH), 128.69 (CH), 128.60 (CH), 128.54 (CH), 122.14 (CH), 78.09 (C), 76.37 (C), 67.04 (CH_2- Pr ester), 55.80 (CH-alkyne), 54.75 (CH-alkyne), 51.91 (CH), 50.62 (CH), 45.91 (C), 44.08 (CH), 43.85 (CH_2), 43.76 (CH_2), 37.32 (CH), 37.00 (CH_2), 36.02 (CH), 35.41 (C), 33.10 (CH_2), 32.10 (CH_2), 32.03 (CH_2), 28.50 (CH_2), 28.30 (CH_2), 28.16 (CH_2), 27.80 (CH_2), 26.80 (CH_2), 24.83 (CH_2), 23.89 (19-CH3), 23.01 (CH_2), 18.69 (CH), 17.89 (CH), 17.26 (21-CH3), 14.25 (CH), 13.14 (18-CH3), 10.74 (CH_3). (ESI-MS m/z (% rel. int.) calcd. 802.4 (100), found 803.4 (100) [M + H^+]. HR-MS (ES) m/z calcd. for $C_{49}H_{62}N_4O_6$ 802.4669, found 803.4734 (M + H^+; Δ = 1 ppm).

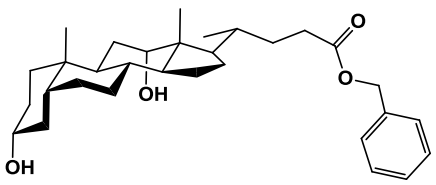


Chemical Formula: $C_{27}H_{48}N_2O_2$
Molecular Weight: 432,68

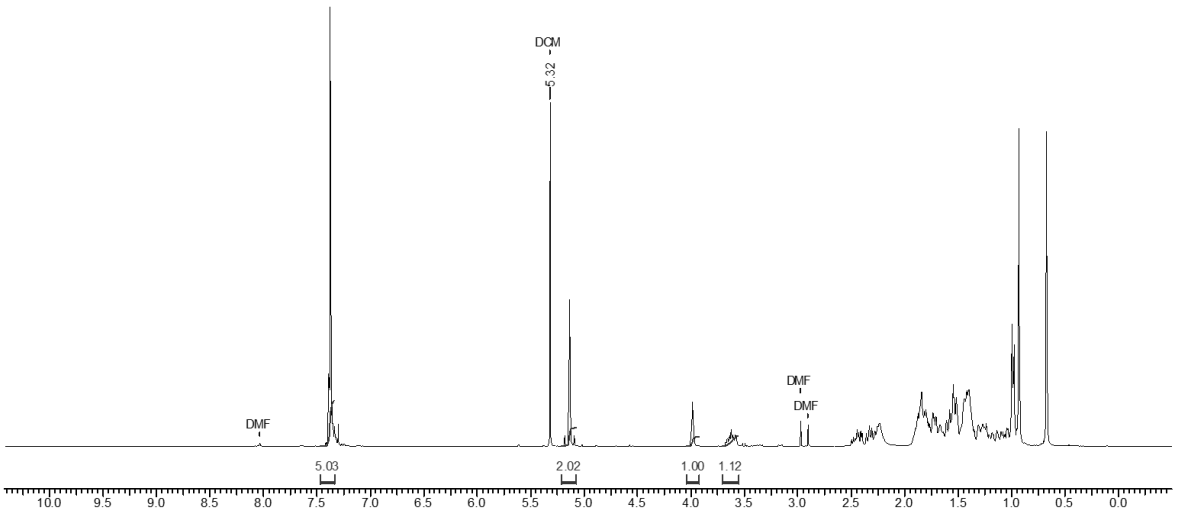
Chemical Formula: $C_{49}H_{62}N_4O_6$
Molecular Weight: 803,04

2. II. NMR Spectra for steroid template molecules:

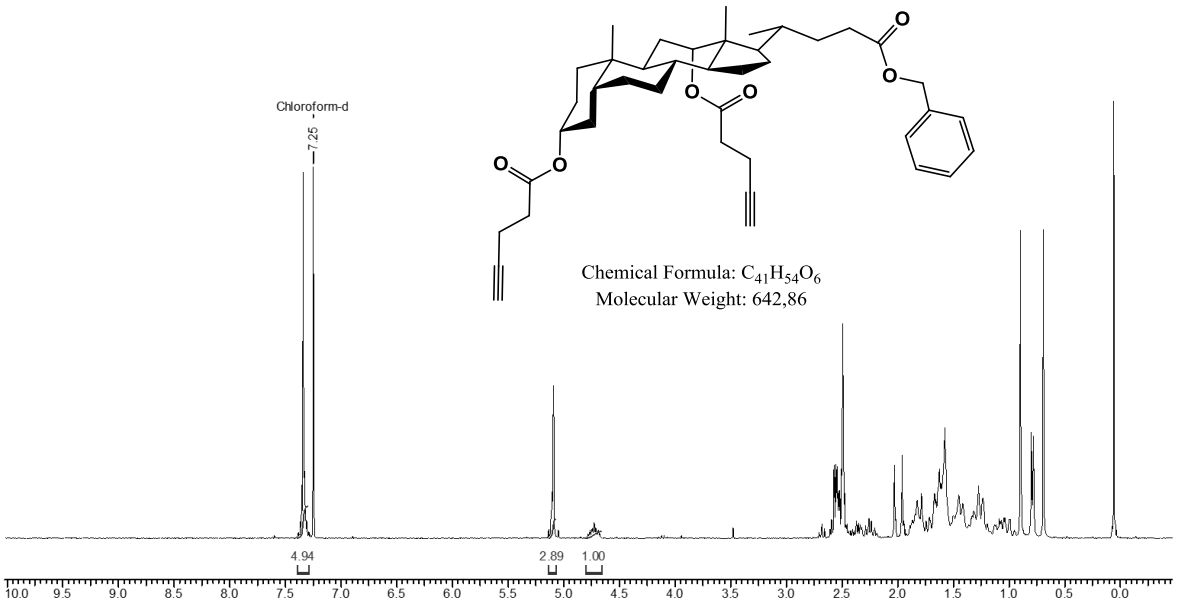
Compound 16:

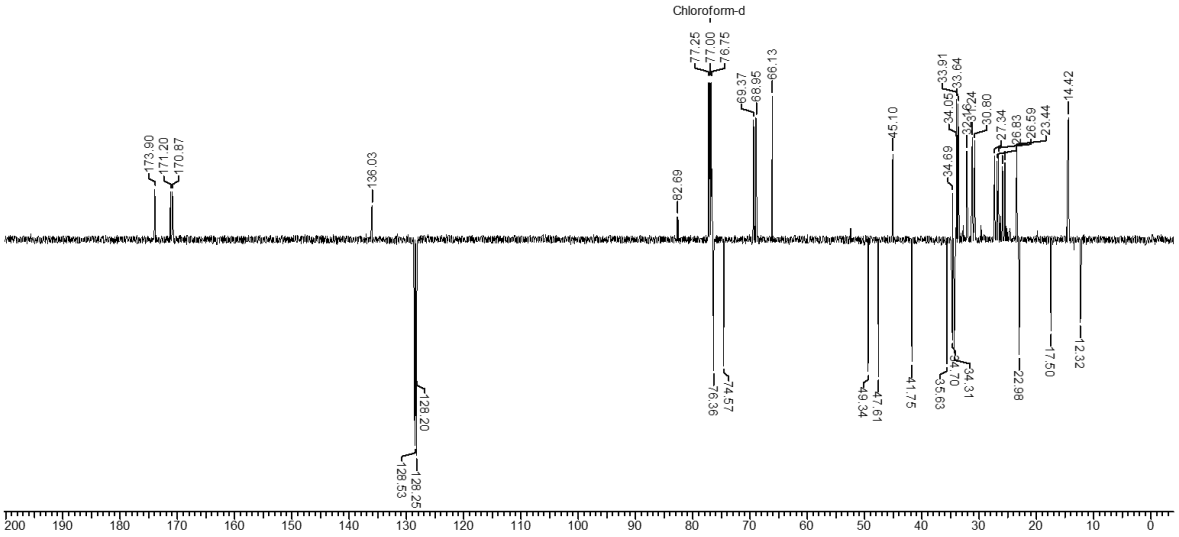


Chemical Formula: $C_{31}H_{46}O_4$
Molecular Weight: 482,69

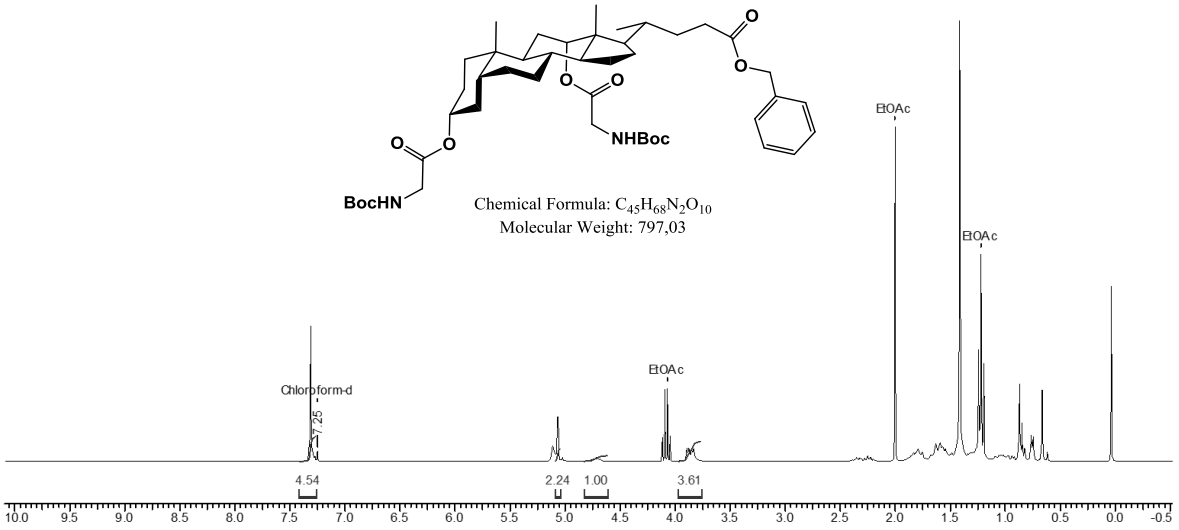

DCM
5.32
DMF
DMF
DMF
5.03
2.02
1.00
1.12

Compound 1:


Chloroform-d
7.25
Chemical Formula: C41H54O6
Molecular Weight: 642,86
4.94
2.89
1.00

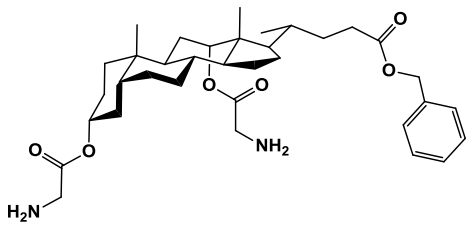


Compound 17:

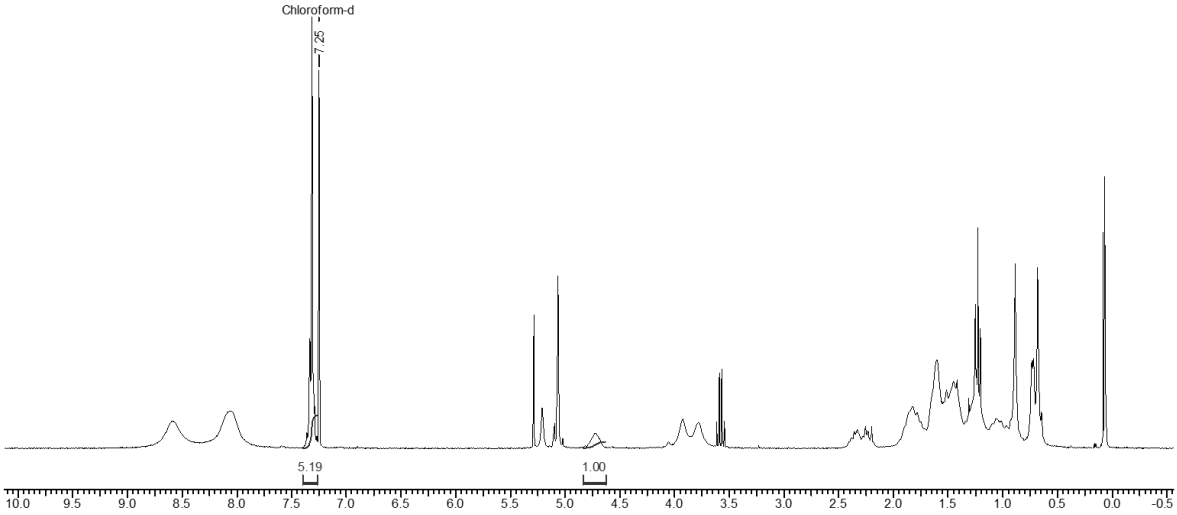


Chemical Formula: $C_{45}H_{68}N_2O_{10}$
Molecular Weight: 797,03

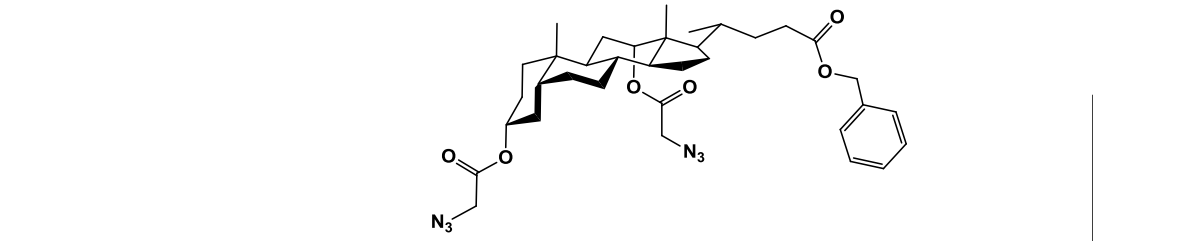
Compound 18:



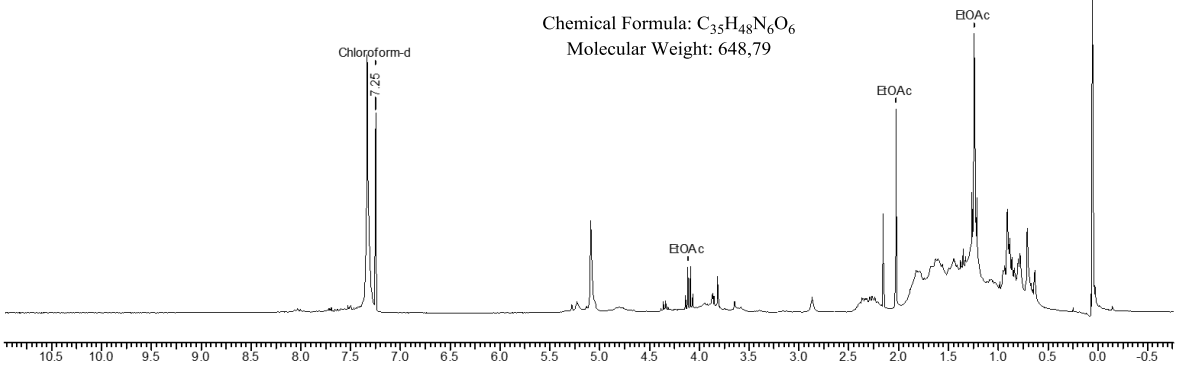
Chemical Formula: $C_{35}H_{52}N_2O_6$
Molecular Weight: 596,80

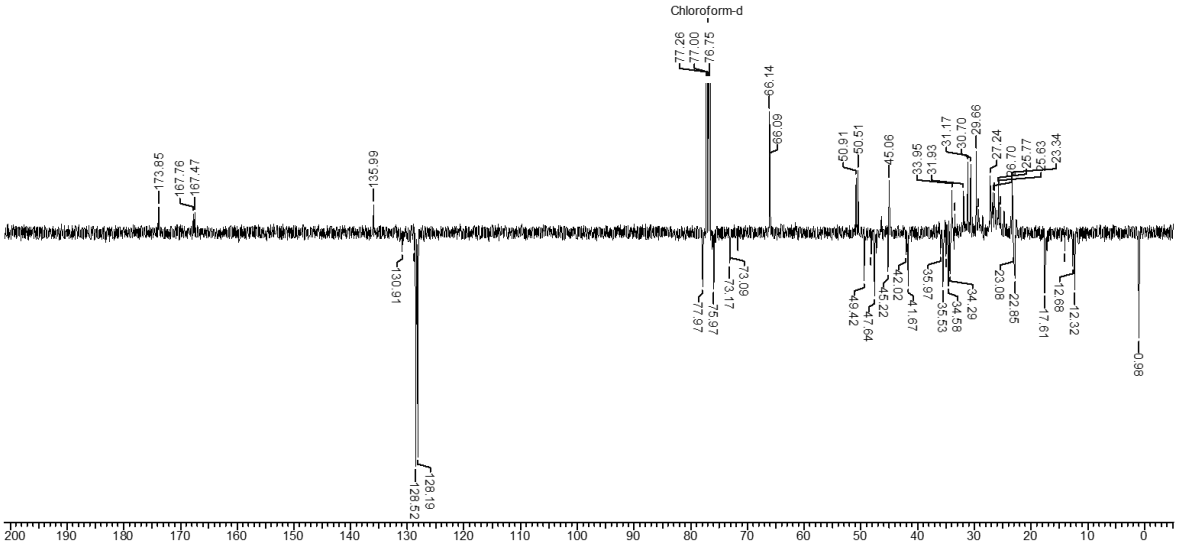


Compound 3:

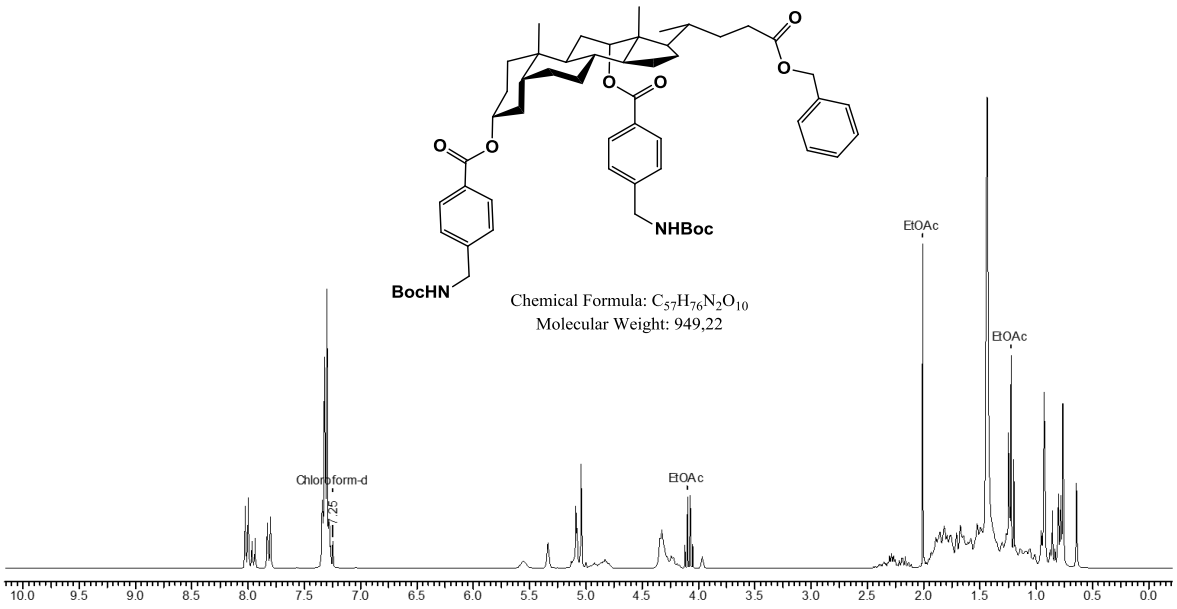


Chemical Formula: $C_{35}H_{48}N_6O_6$
Molecular Weight: 648,79

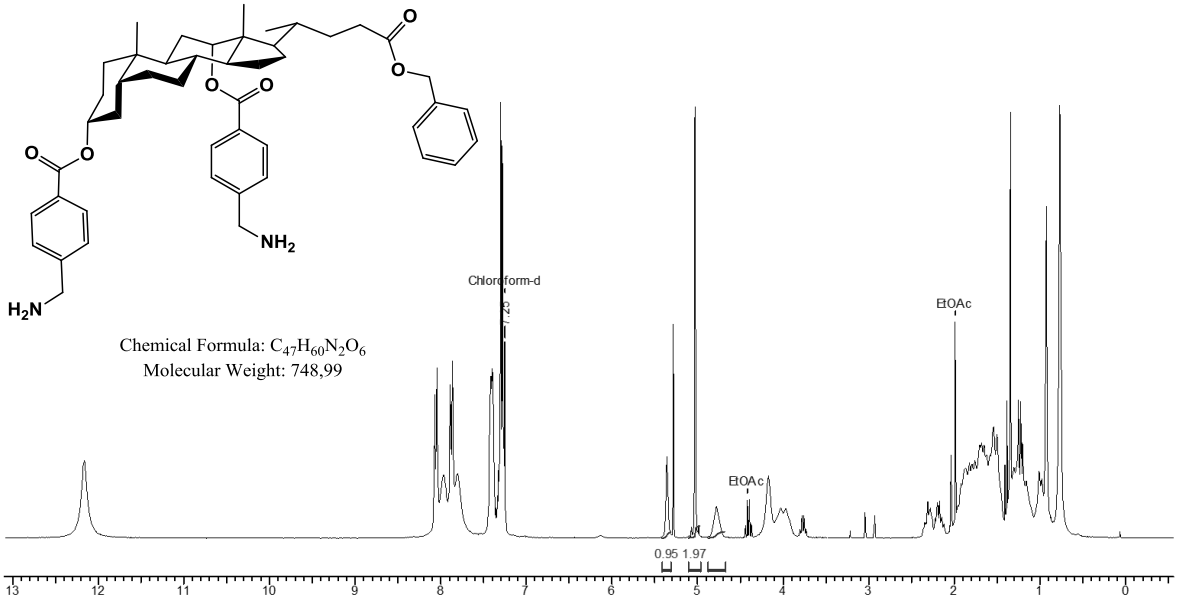




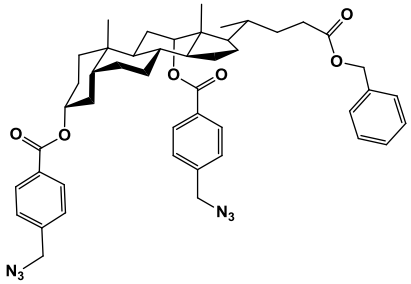
Compound 19:

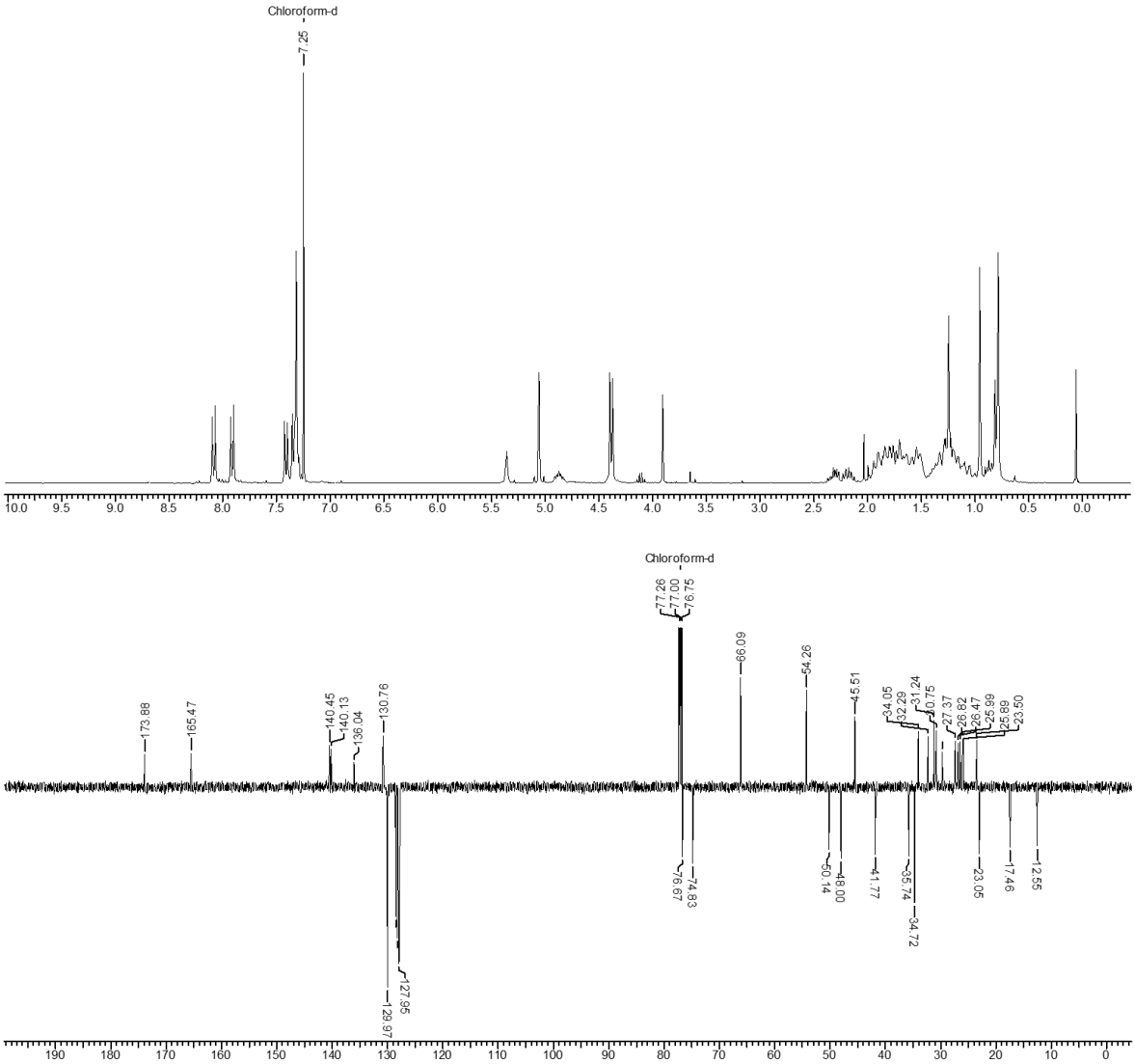


Compound 20:

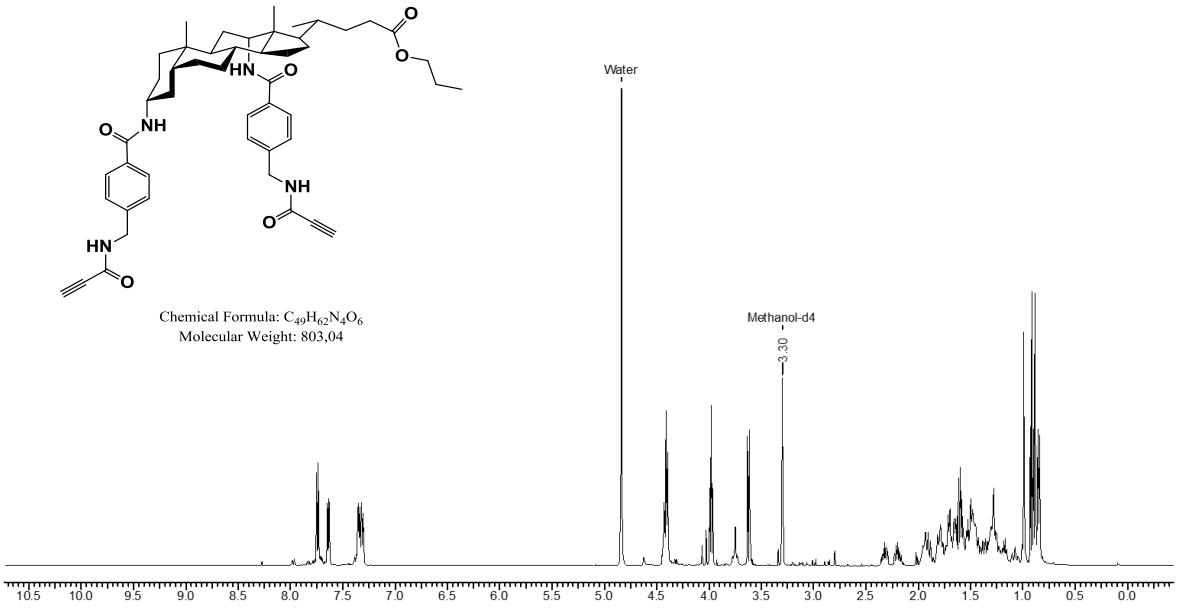


Compound 2:

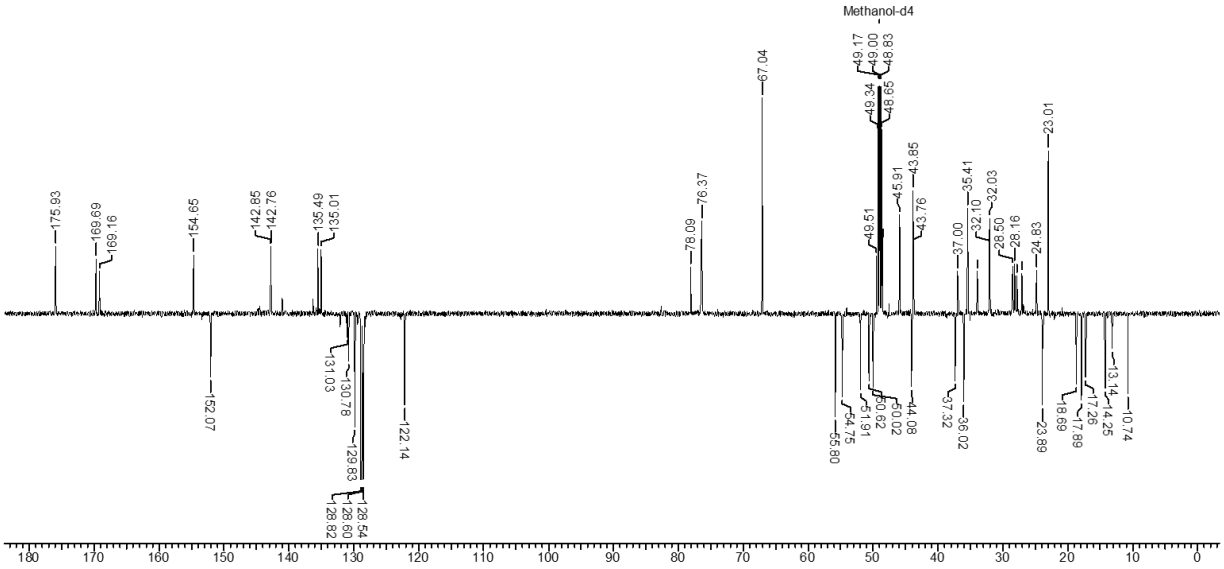




Compound 4:



Chemical Formula: $C_{49}H_{62}N_4O_6$
Molecular Weight: 803,04



3. Peptide synthesis:

3. I. Design of linear peptides:

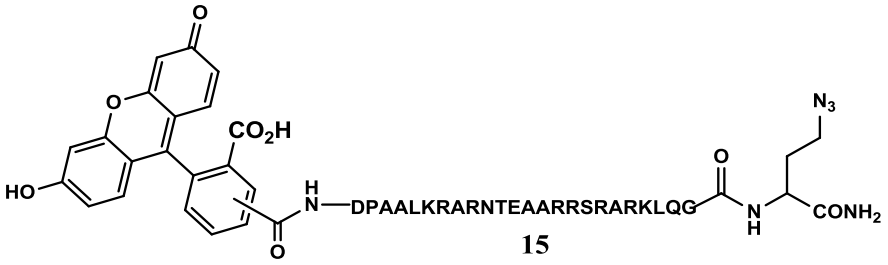
The following five linear peptides were synthesized in context of this study:

6

Chemical Formula: $C_{123}H_{212}N_{52}O_{35}$
Exact Mass: 2977.64
Molecular Weight: 2979.38

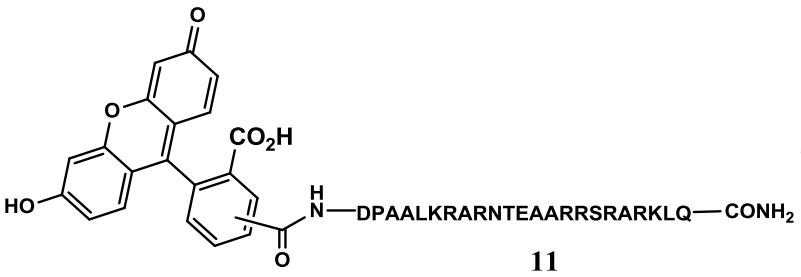
5

Chemical Formula: $C_{124}H_{211}N_{49}O_{35}$
Exact Mass: 2946.62
Molecular Weight: 2948.36



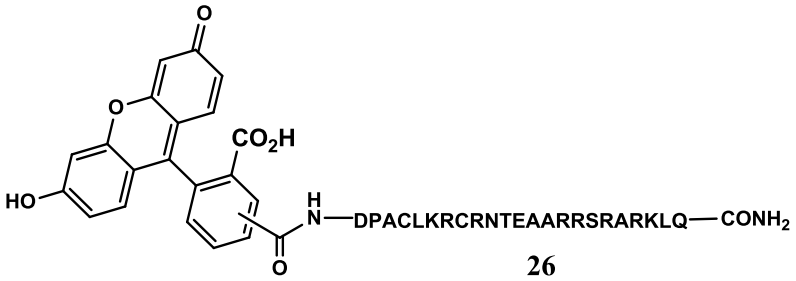
15

Chemical Formula: $C_{135}H_{215}N_{51}O_{39}$
Exact Mass: 3174.64
Molecular Weight: 3176.52



Chemical Formula: $C_{129}H_{206}N_{46}O_{37}$
Exact Mass: 2991.57
Molecular Weight: 2993.35

11

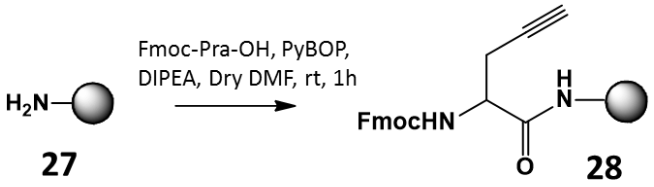


Chemical Formula: $C_{129}H_{206}N_{46}O_{37}S_2$
Exact Mass: 3055.51
Molecular Weight: 3057.47

26

3. II. Manual coupling protocols:

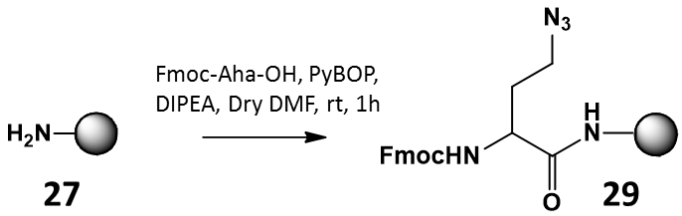
Immobilization of Fmoc-Pra-OH on Rink amide ChemMatrix resin



27 **28**

To a suspension of resin **27** (50 mg, 0.54 mmol/g, 0.027 mmol) in DMF (dry, 10 mL/g resin), were added Fmoc-Pra-OH (27.2 mg, 0.081 mmol, 4 eq.), PyBOP (42.1 mg, 0.081 mmol, 4 eq.) and DIPEA (0.028 mL, 0.162 mmol, 8 eq.). The mixture was shaken at room temperature for 1 h. After the reaction, the resin was washed with DMF, ACN and MeOH. The loading was 100%. The resin was then capped with acetic anhydride (0.016 mL, 0.162 mmol, 8 eq.) and DIPEA (0.028 mL, 0.162 mmol, 8 eq.) in dry DMF (2 mL) twice for 30 min.

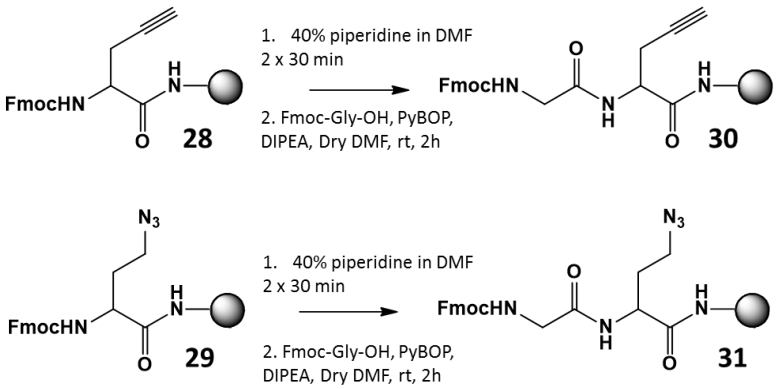
Immobilization of Fmoc-Aha-OH on Rink amide ChemMatrix resin



27 **29**

To a suspension of resin **27** (50 mg, 0.54 mmol/g, 0.027 mmol) in DMF (dry, 10 mL/g resin), were added Fmoc-Aha-OH (29.7 mg, 0.081 mmol, 4 eq.), PyBOP (42.1 mg, 0.081 mmol, 4 eq.) and DIPEA (0.028 mL, 0.162 mmol, 8 eq.). The mixture was shaken at room temperature for 1 h. After the reaction, the resin was washed with DMF, ACN and MeOH and dried in vacuum for 1h. The loading was calculated to be 0.412 mmol/g, indicating a coupling yield of 92%. The resin was then capped with acetic anhydride (0.016 mL, 0.162 mmol) and DIPEA (0.028 mL, 0.162 mmol) in dry DMF (2 mL) 2 times for 30 min.

Fmoc deprotection and coupling of Fmoc-Gly-OH



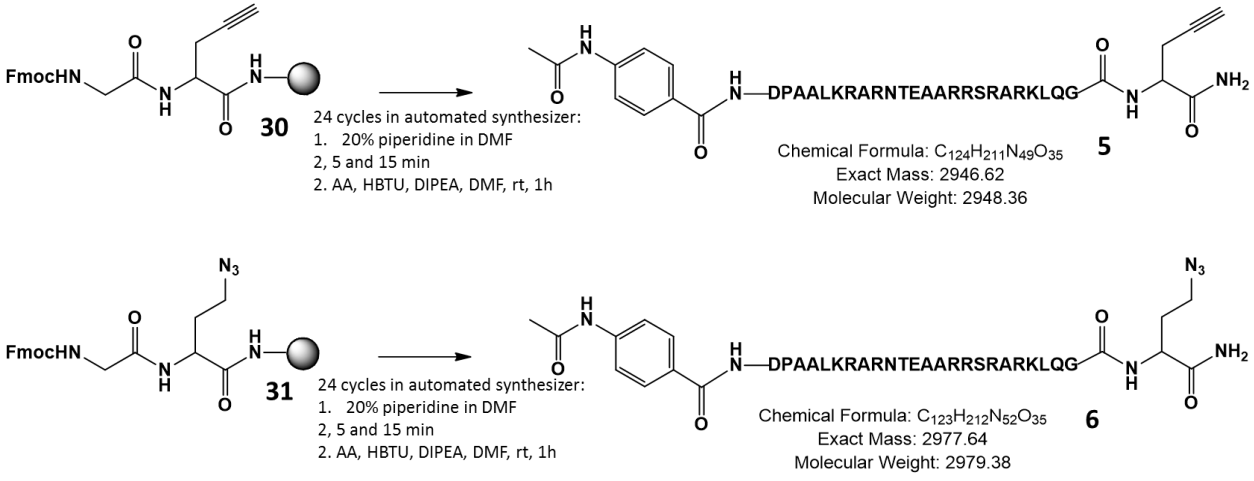
Fmoc deprotection of compounds **28** & **29**. After an initial DMF washing step, resin 17 (0.022 mmol) was treated twice successively for 30 min with a piperidine solution in DMF (40 % v/v, 2 mL) at ambient temperature, applying intermediate filtration under reduced pressure and washing with 3 x DMF, 3 x ACN, 3 x DCM and 3 x DMF.

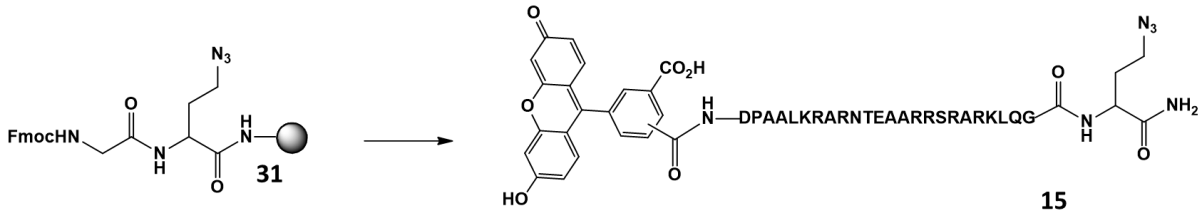
Coupling of Fmoc-Gly-OH to **28***.* Fmoc-Gly-OH (19.5 mg, 0.065 mmol, 0.5 M), PyBOP (34.15 mg, 0.065 mmol, 0.5 M) and DIPEA (0.012 mL, 0.065 mmol, 2 M) were added to a suspension of resin in dry DMF (2 mL). The mixture was shaken at room temperature for 2 h. After the reaction, the resin was washed again with DMF/MeOH/DCM/Et_2O/DMF.

Coupling of Fmoc-Gly-OH to **29***.* Fmoc-Gly-OH (22.3 mg, 0.075 mmol, 0.5 M), PyBOP (39 mg, 0.075 mmol, 0.5 M) and DIPEA (0.013 mL, 0.075 mmol, 2 M) were added to a suspension of resin in dry DMF (2 mL). The mixture was shaken at room temperature for 2 h. After the reaction, the resin was washed again with DMF/MeOH/DCM/Et_2O/DMF.

Coupling of Fluorescein: 3 eq. of NHS-fluorescein predissolved in 0.5 mL DMF. 6 eq. of DIPEA and 3 eq. of HOBt predissolved in 0.5 mL DMF and are added successively to the resin. After shaking for 15 h, the reaction mixture was filtered off and the resin washed with 3x DMF, 3x MeOH, 3x DMF and 3x Et_2O.

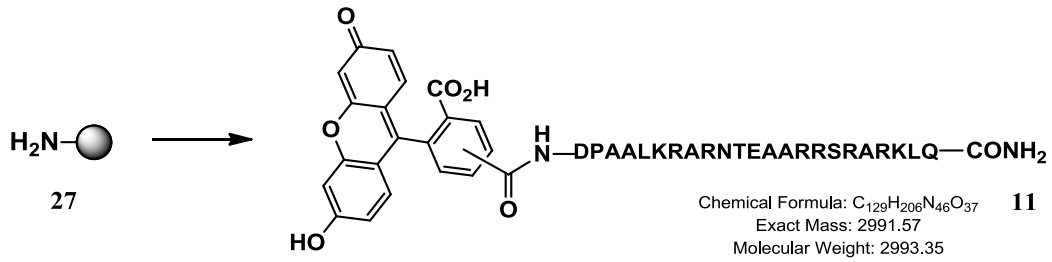
3. III. Automated synthesis of linear peptides:





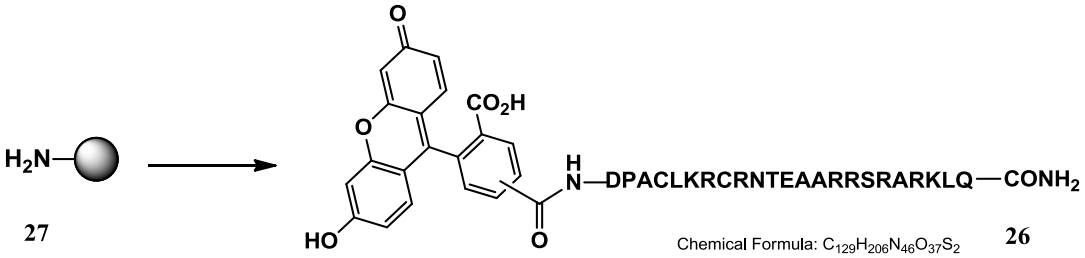
24 cycles in automated synthesizer:
1. 20% piperidine in DMF
2, 5 and 15 min
2. AA, HBTU, DIPEA, DMF, rt, 1h

Chemical Formula: $C_{135}H_{215}N_{51}O_{39}$
Exact Mass: 3174.64
Molecular Weight: 3176.47



Chemical Formula: $C_{129}H_{206}N_{46}O_{37}$
Exact Mass: 2991.57
Molecular Weight: 2993.35

24 cycles in an automated synthesizer:
1. 20% Piperdine in DMF, 2, 5, 15 min
2. AA, HBTU, DIPEA, DMF, 1 h, rt.
3. Fluorescein, HOBt, DIPEA in DMF 15h r.t.



Chemical Formula: $C_{129}H_{206}N_{46}O_{37}S_2$
Exact Mass: 3055.51
Molecular Weight: 3057.47

24 cycles in an automated synthesizer:
1. 20% Piperdine in DMF, 2, 5, 15 min
2. AA, HBTU, DIPEA, DMF, 1 h, rt.
3. Fluorescein, HOBt, DIPEA in DMF 15h r.t.

Automated solid phase peptide synthesis was carried out on a Syro automated peptide synthesizer from Biotage using standard Fmoc/tBu chemistry with HBTU as coupling reagent and 20-40% piperidine in NMP for Fmoc deprotection. The resin is preswollen in NMP for 10 min and then filtered off. The following protocols were used for Fmoc deprotection and coupling:

Fmoc deprotection: Fmoc deprotection is performed by adding a solution of 20-40% piperidine in NMP to the resin and shaking for 2 min, 5 min & 15 min durations. After each addition and shaking cycle, the resin is filtered off and washed with NMP (6 x 45 s).
Coupling: 10 equivalents of a 0.5 M solution of the Nα protected amino acid in NMP/DMF, 10 equivalents of a 0.5 M solution of HBTU in DMF and 10 equivalents of a 2.0 M solution of DIPEA in NMP are added to the resin. The reaction mixture is shaken for 1 hour. The resin is filtered off and washed with NMP (9 x 2 min).
Capping: 4-Acetamidobenzoic acid used for capping was coupled using the protocol in *coupling* above. In case of fluorescein a manual coupling was performed as described in *coupling of fluorescein* above.

The peptides were simultaneously cleaved from the resin and deprotected with a cocktail of TFA/TIS/water = 95: 2.5: 2.5 for 4h (4 mL/50 mg). After cleavage the resin was washed with neat TFA and was precipitated in cold ether (-20°C) by centrifugation at 0-4°C. The peptides were analyzed by RP-HPLC/LC-MS and/or MALDI.

3. IV. Analysis of linear peptides.

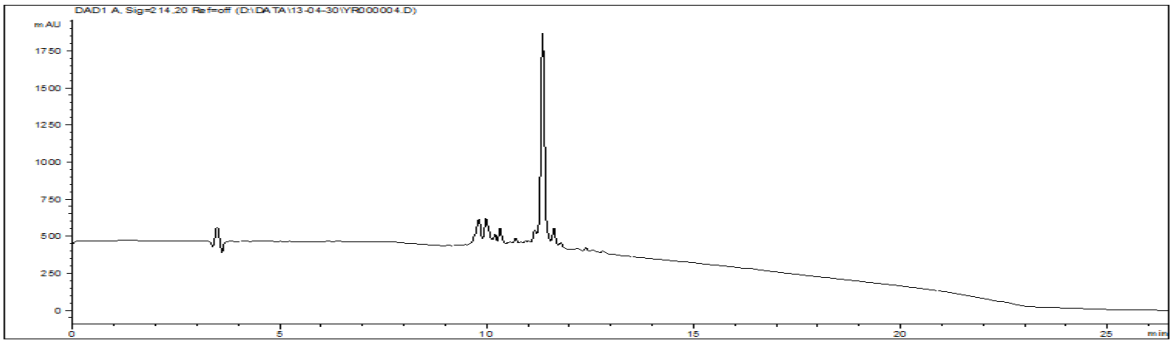


Figure 1: RP-HPLC Chromatogram of crude compound **5** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes)

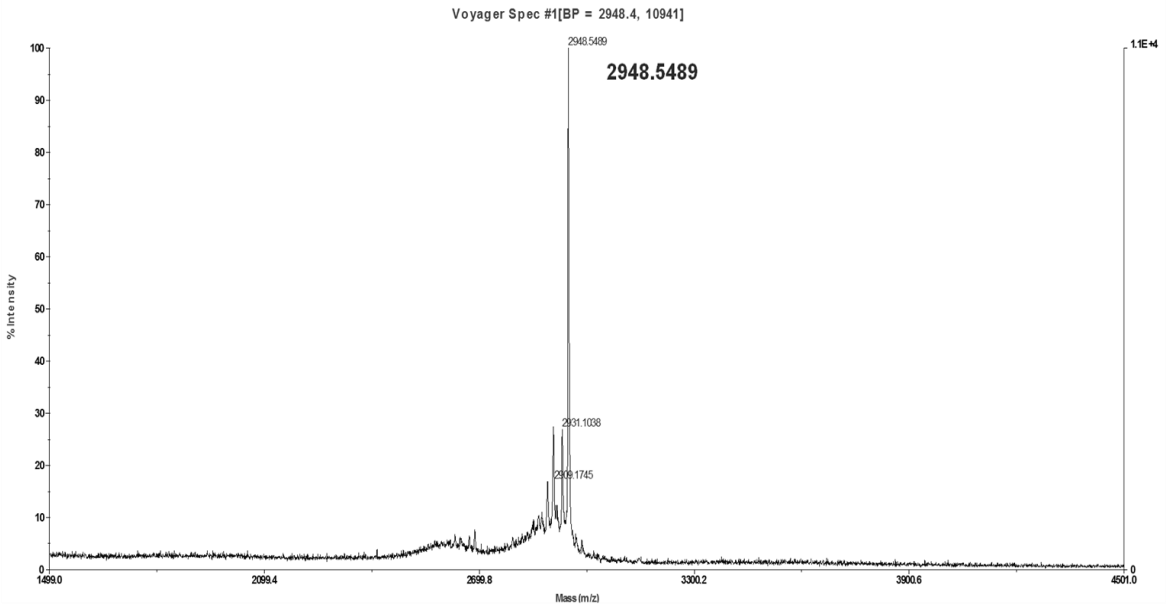


Figure 2: MALDI-TOF Spectrum of crude compound **5**. Calculated m/z (100%) = 2947.63; m/z (73.9%) = 2948.63. Found: 2948.5489 = M + H^+.

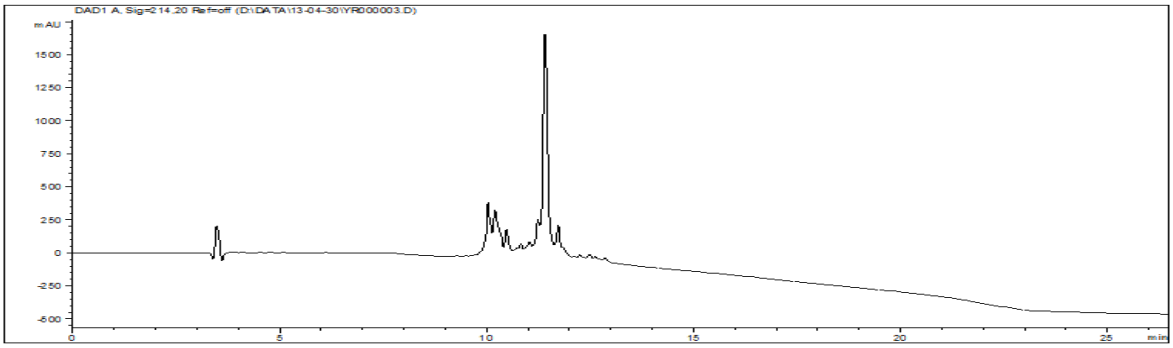


Figure 3: RP-HPLC Chromatogram of crude compound **6** (C4, 300Å column using a gradient from 0 to 100 % CH3CN in 15 minutes)

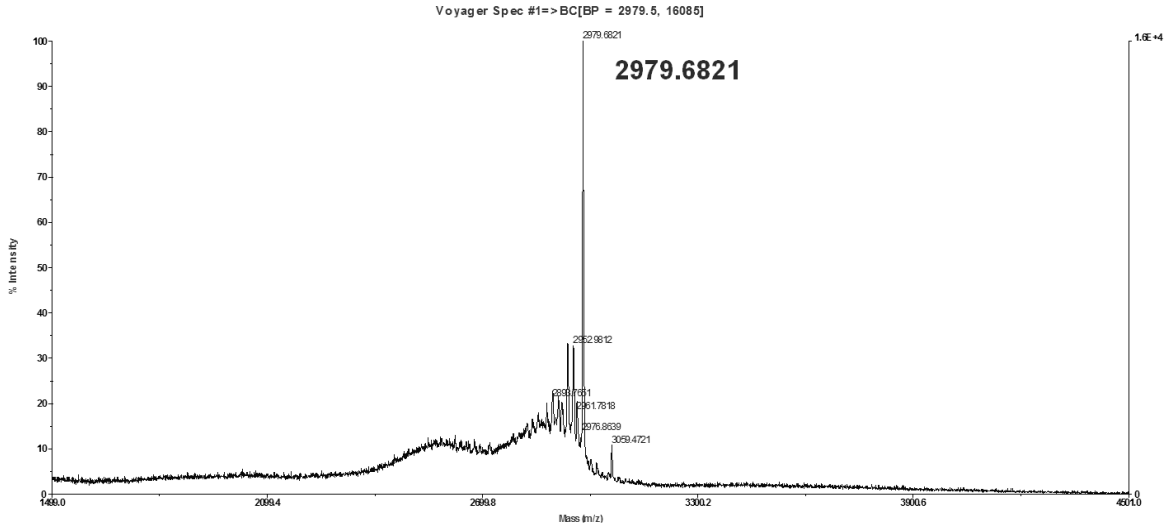


Figure 4: MALDI-TOF Spectrum of crude compound **6**. Calculated m/z (100%) = 2978.64; m/z (65.1%) = 2979.65. Found: 2979.6821 = M + H^+.

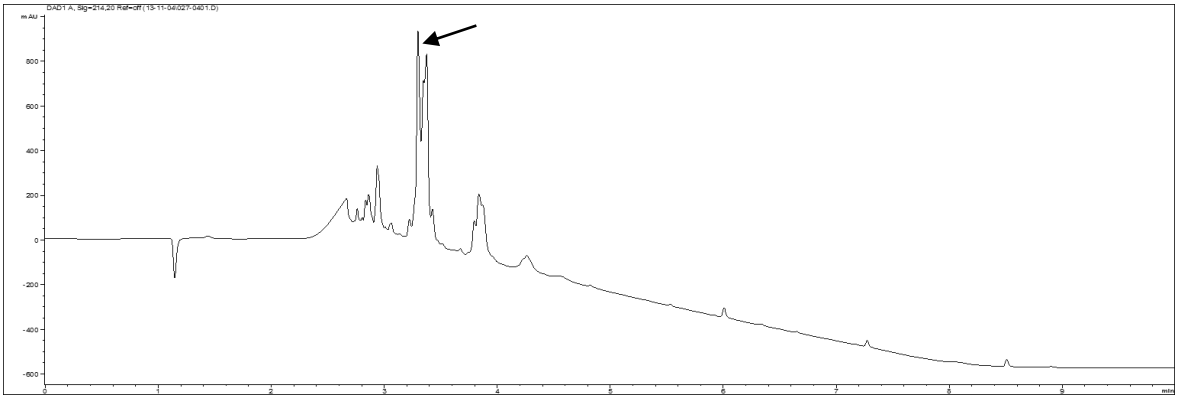


Figure 5: RP-HPLC of compound **15** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C) (double peak due to the isomers of fluorescein)

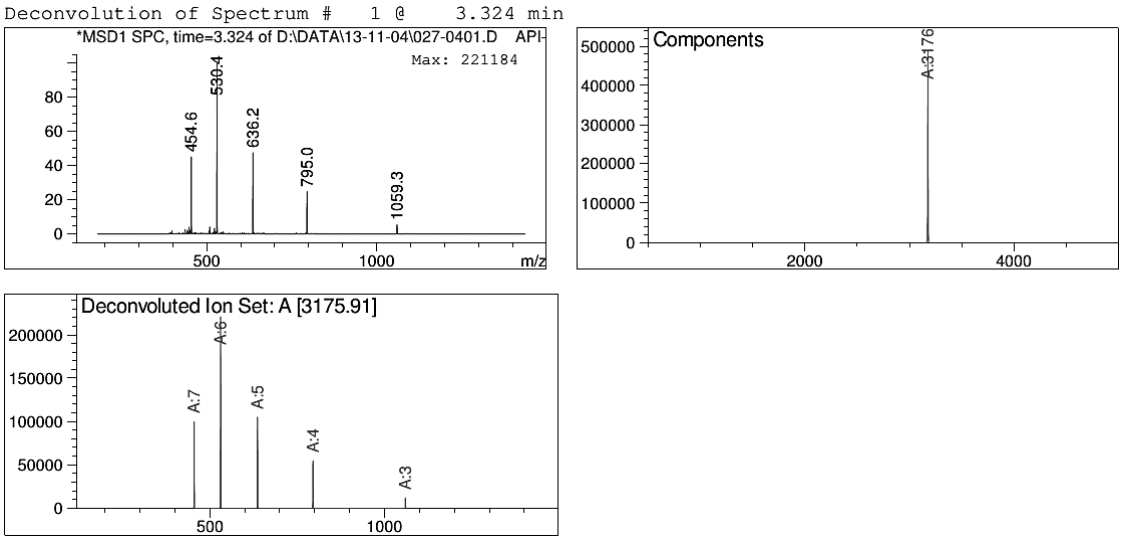


Figure 6: ESI-MS from LC-MS at t = 3.324 min for crude compound **15**. E.M calcd. for $C_{135}H_{215}N_{51}O_{39}$ = 3174.64 and deconvoluted mass found 3175.91.

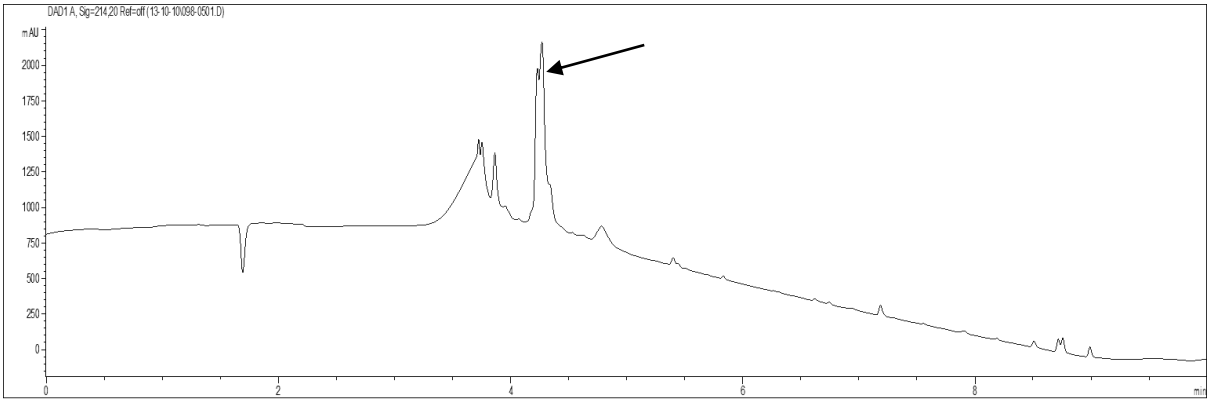


Figure 7: RP-HPLC of crude compound **11** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C) (double peak due to the isomers of fluorescein)

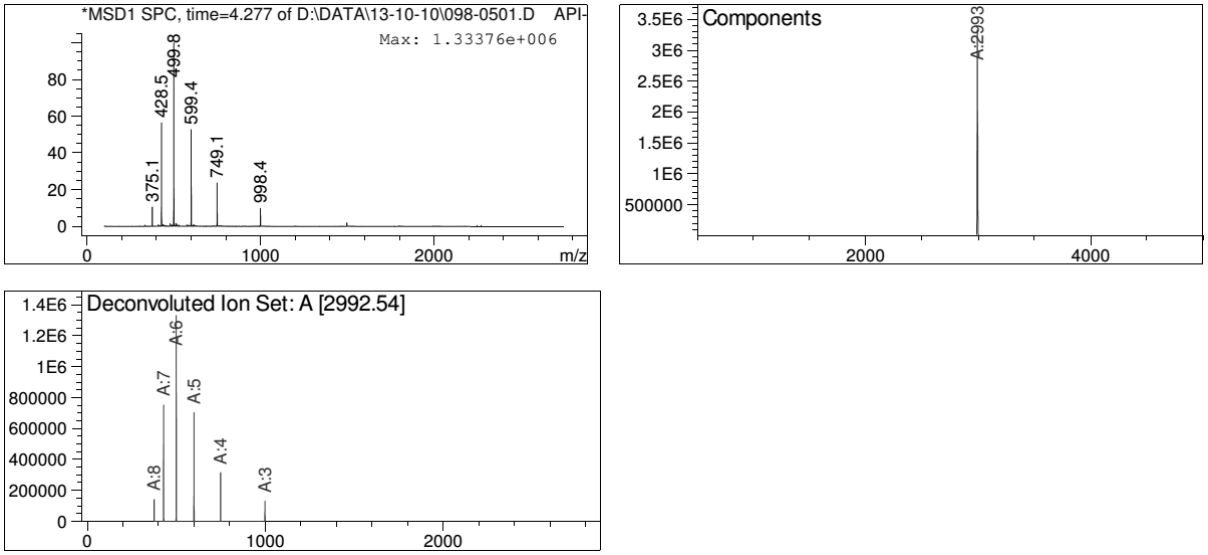


Figure 8: ESI-MS from LC-MS at t = 4.277 min for crude compound **11**. E.M calcd. for $C_{129}H_{206}N_{46}O_{37}$ = 2991.57 and deconvoluted mass found 2992.54.

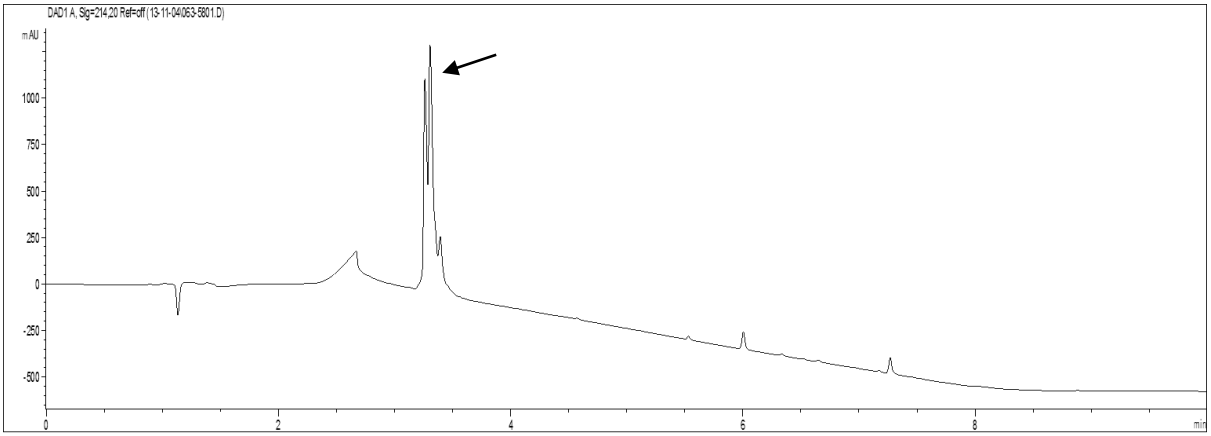


Figure 9: RP-HPLC trace of HPLC purified compound **11** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C) (double peak due to the isomers of fluorescein)

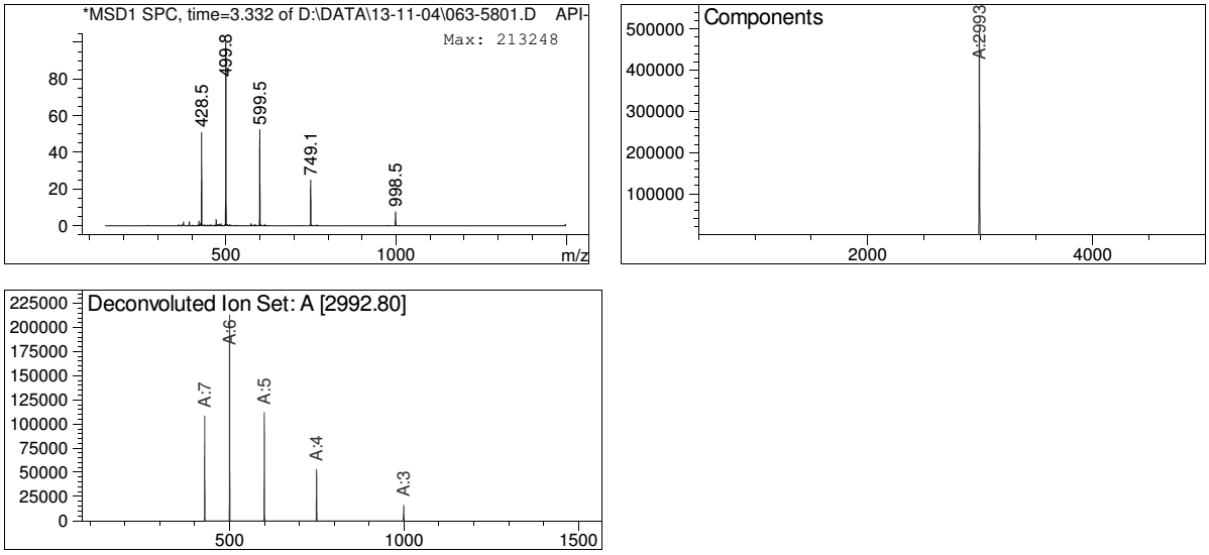


Figure 10: ESI-MS from LC-MS at t = 3.332 min for HPLC purified compound **11**. E.M calcd. for $C_{129}H_{206}N_{46}O_{37}$ = 2991.57 and deconvoluted mass found 2992.80.

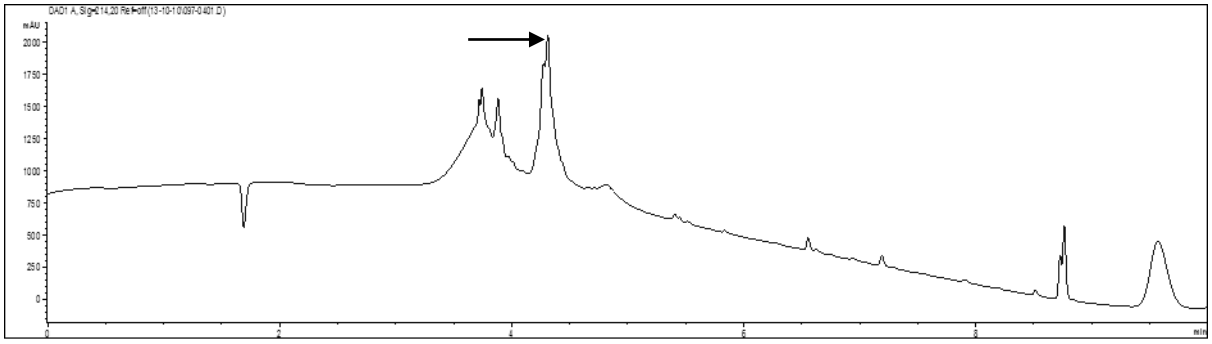


Figure 11: RP-HPLC trace of HPLC crude compound **26** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C) (double peak due to the isomers of fluorescein)

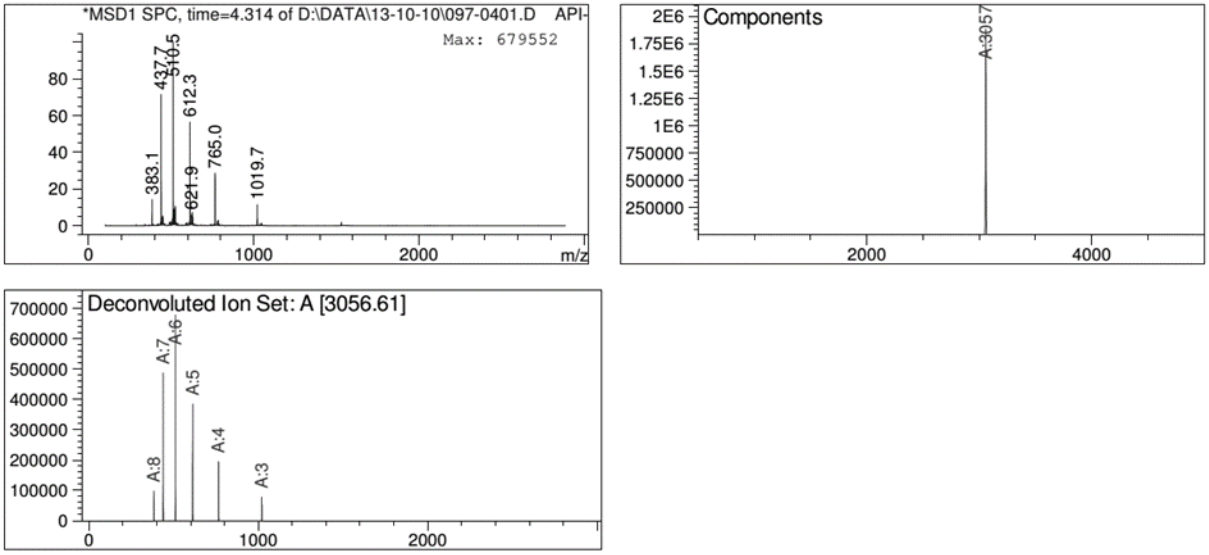
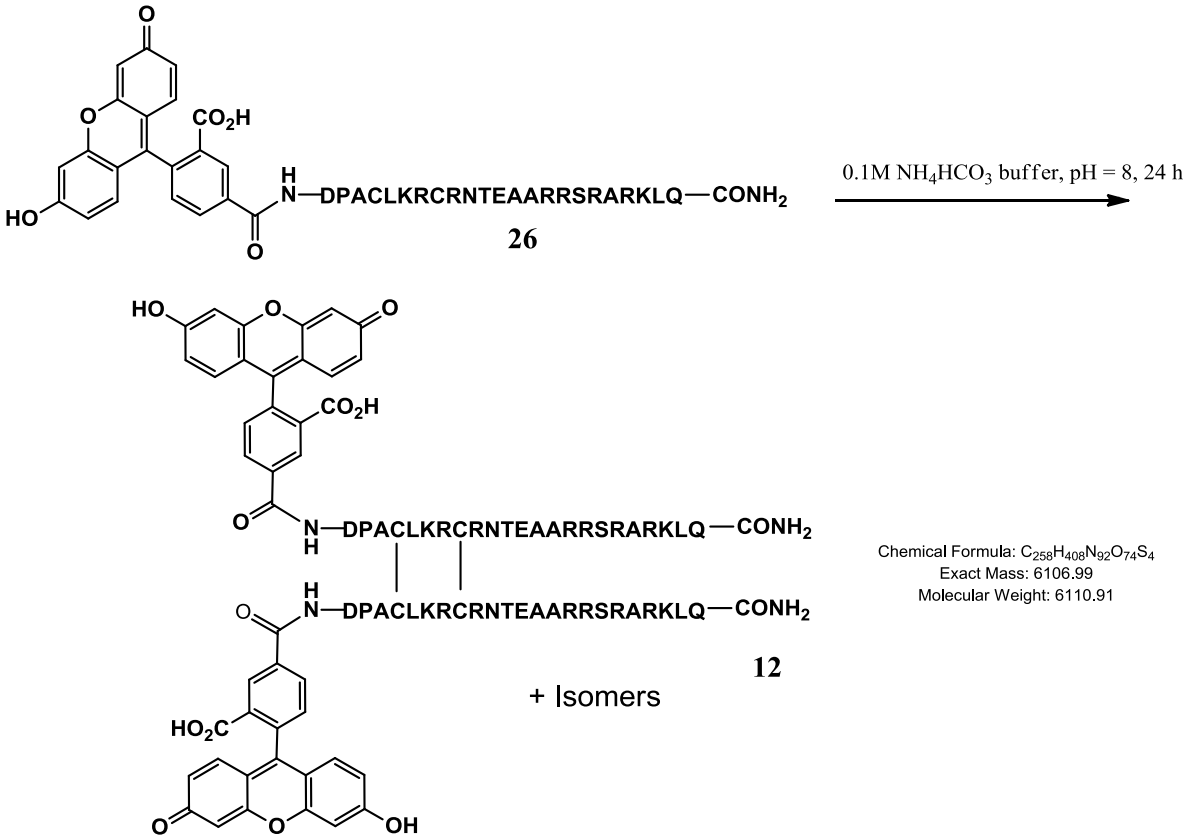


Figure 12: ESI-MS from LC-MS at t = 4.314 min for crude compound **26**. E.M calcd. for $C_{129}H_{206}N_{46}O_{37}S_2$ = 3055.51 and deconvoluted mass found 3056.61.

3. V. Synthesis of compound 12



Peptide **26** (5 mg, 1.63 µmol) was dissolved in 0.1 M NH_4HCO_3 buffer (1 mL) and stirred vigorously for 24h. The reaction was then quenched with TFA. A mixture of peptides containing 2 peptide strands but with single or double disulfide bonds was isolated using RP-HPLC. No higher mass products were observed.

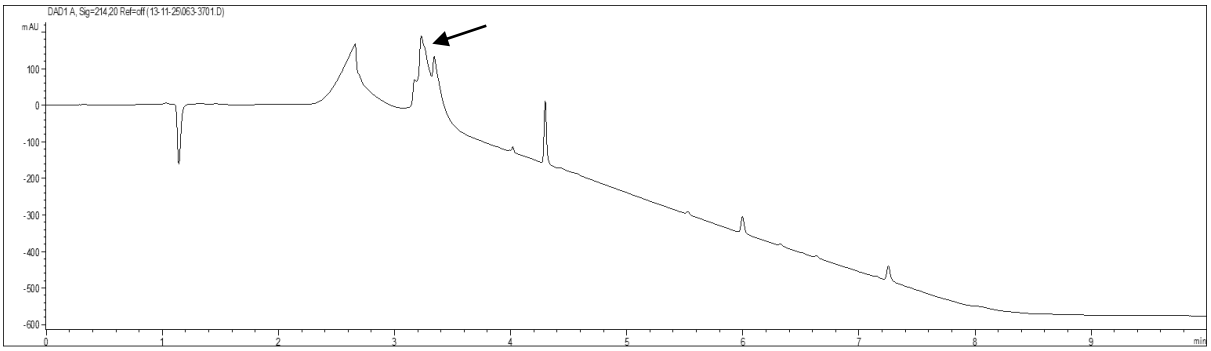


Figure 13: RP-HPLC trace of compound **12** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C) (mixture of isomers)

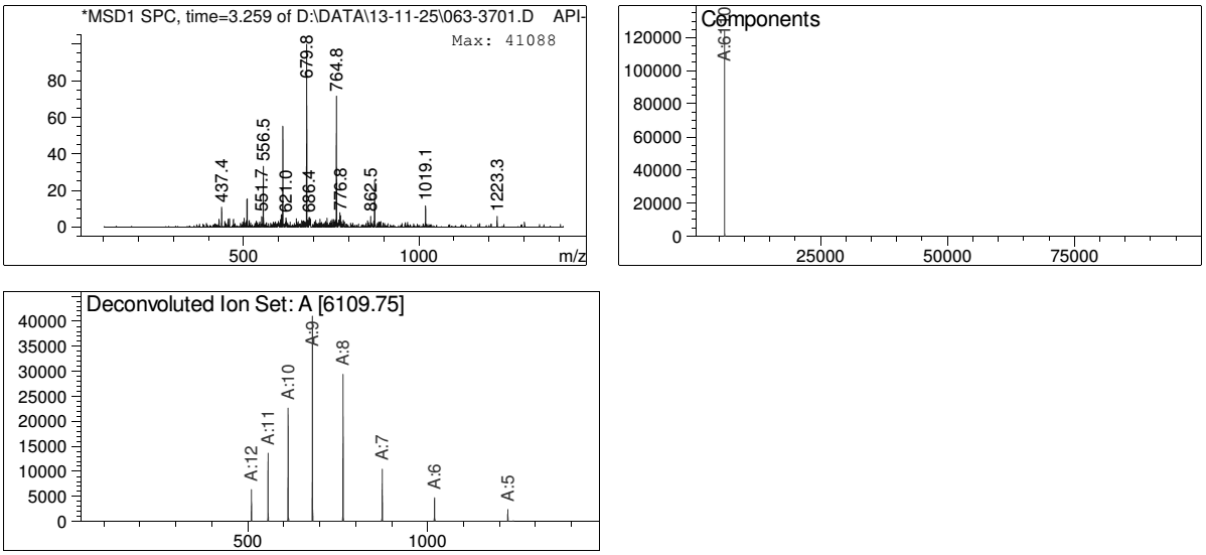


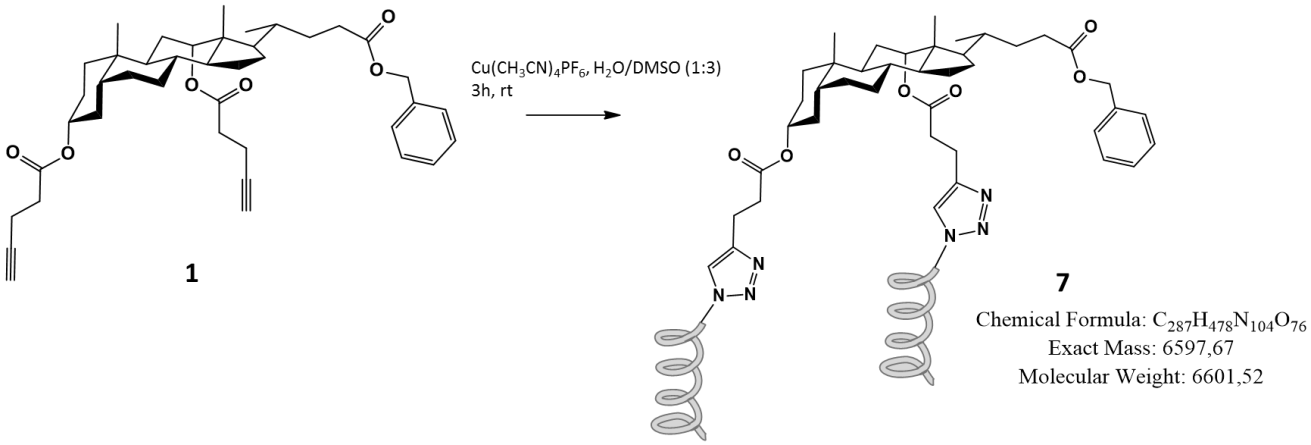
Figure 14: ESI-MS from LC-MS at t = 3.259 min for purified compound **12**. E.M calcd. for $C_{258}H_{408}N_{92}O_{74}S_4$ = 6106.99 and $C_{258}H_{410}N_{92}O_{74}S_4$ = 6109.00 and deconvoluted mass found 6109.75.

4. CuAAC Conjugation

Scaffold-Peptide Conjugation Protocol via CuAAC

4. I. COMPOUNDS FOR DNA BINDING

Pentynoic acid (7): Scaffold **1** (3.2 mg, 5 µmol) was dissolved in 0.2 mL dry DMSO in a 5 mL round bottom flask. Peptide **22** (3.7 mg, 1.2 µmol) was dissolved in 0.1 mL miliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (4.63 mg, 12 µmol) was dissolved in 75 µL dry DMSO and added it to the reaction mixture. The reaction was stirred for 3 h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC. Fractions were lyophilized and analyzed by RP-HPLC and MALDI-TOF.



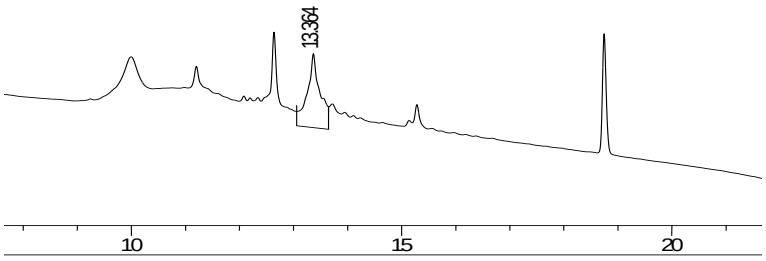


Figure 15: RP-HPLC trace of HPLC crude compound **7** (0-100% ACN in 15 min on Jupiter c4 300 Å, 250 x 2.1 mm, 2.6 µm, at 35 °C)

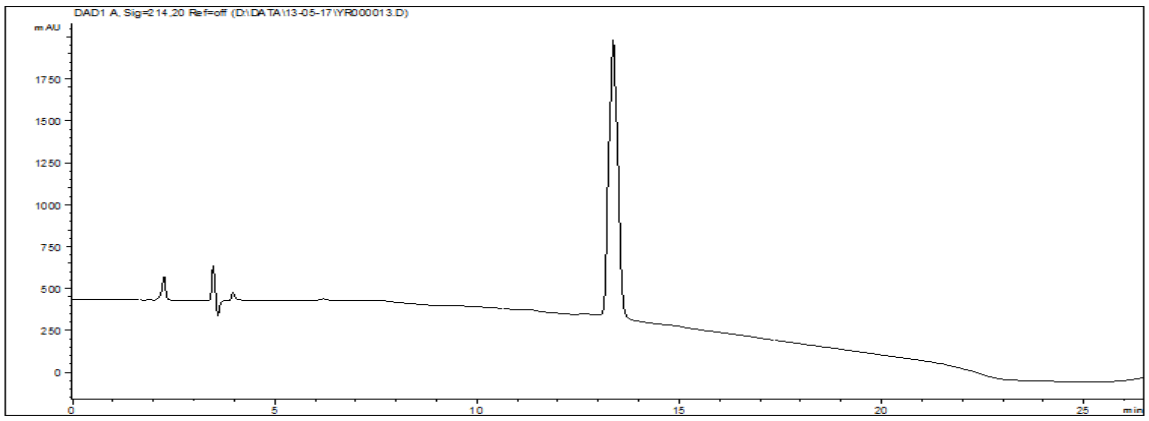


Figure 16: RP-HPLC Chromatogram of pure compound **7** (C4, 300Å column using a gradient from 0 to 100 % CH3CN in 15 minutes)

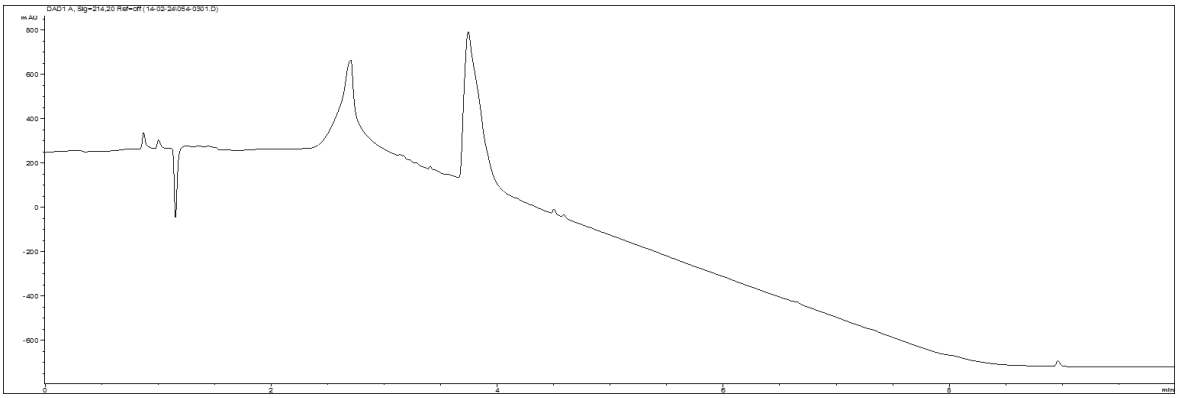


Figure 17: RP-HPLC of compound **7** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

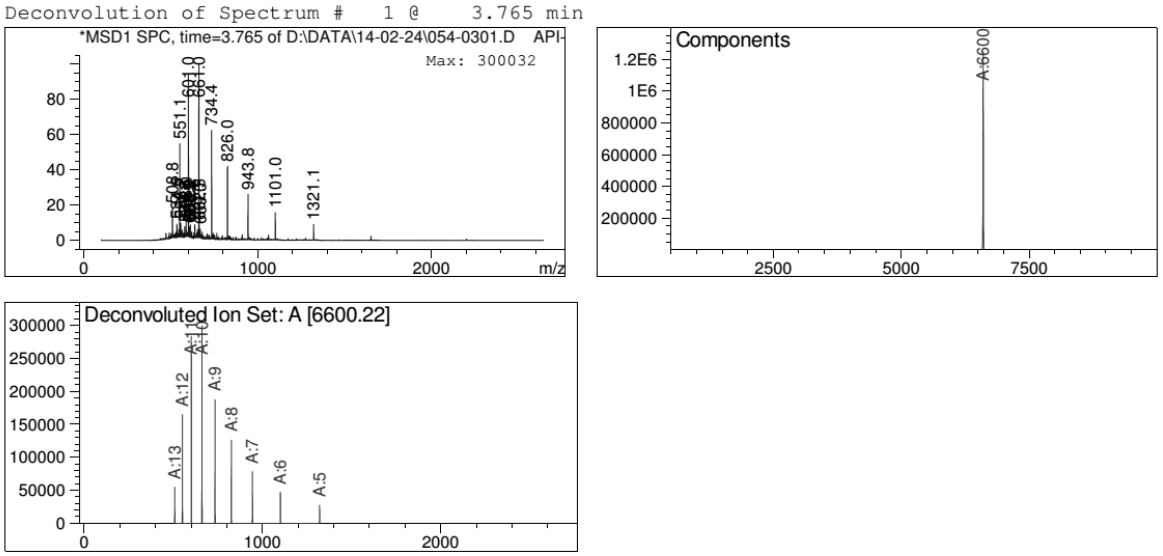
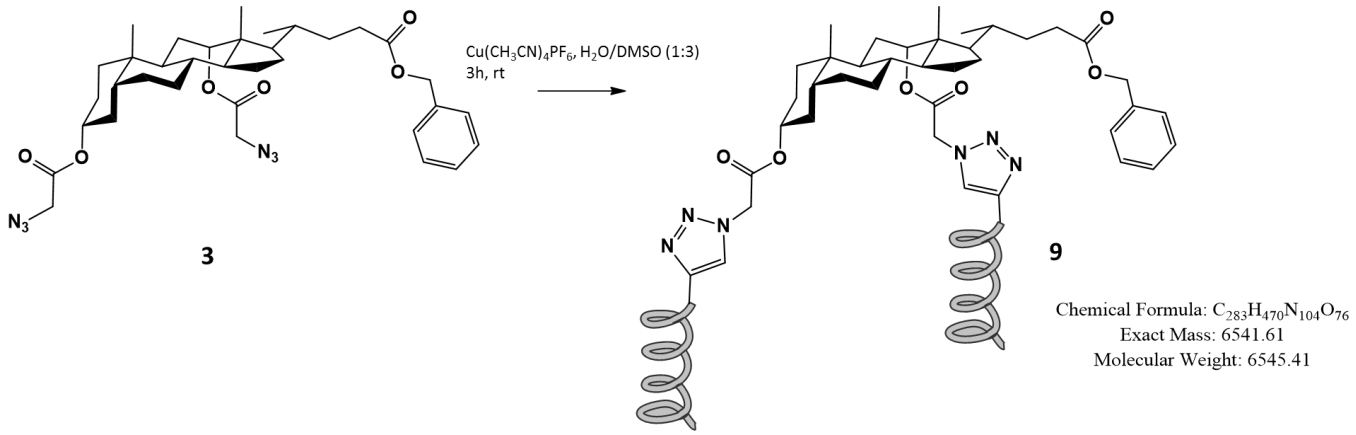


Figure 18: ESI-MS spectrum from LC-MS at t = 3.765 min of pure compound **7**. EM calcd. for $C_{287}H_{478}N_{104}O_{76}$ = 6597.67 and deconvoluted mass found 6600.2

Gly (9): Scaffold **3** (3.2 mg, 5 µmol) was dissolved in 0.2 mL dry DMSO in a 5 mL round bottom flask. Peptide **5** (3.7 mg, 1.2 µmol) was dissolved in 0.1 mL miliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (4.68 mg, 13 µmol) was dissolved in 75 µL dry DMSO and added it to the reaction mixture. The reaction was stirred for 3 h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC. Fractions were lyophilized and analyzed by RP-HPLC and MALDI-TOF.



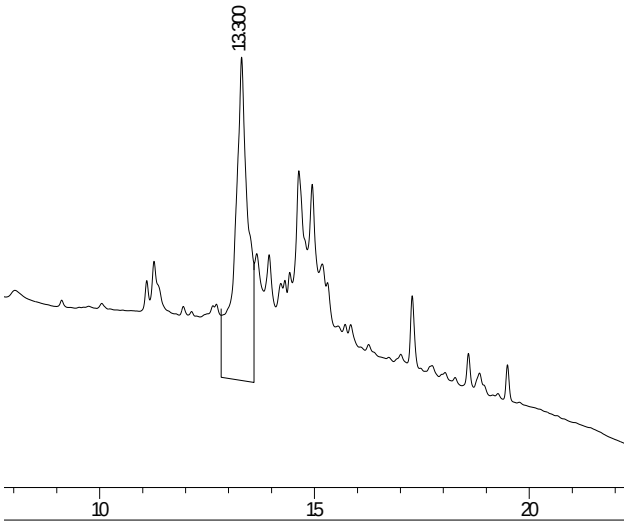


Figure 19: RP-HPLC trace of HPLC crude compound **9** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

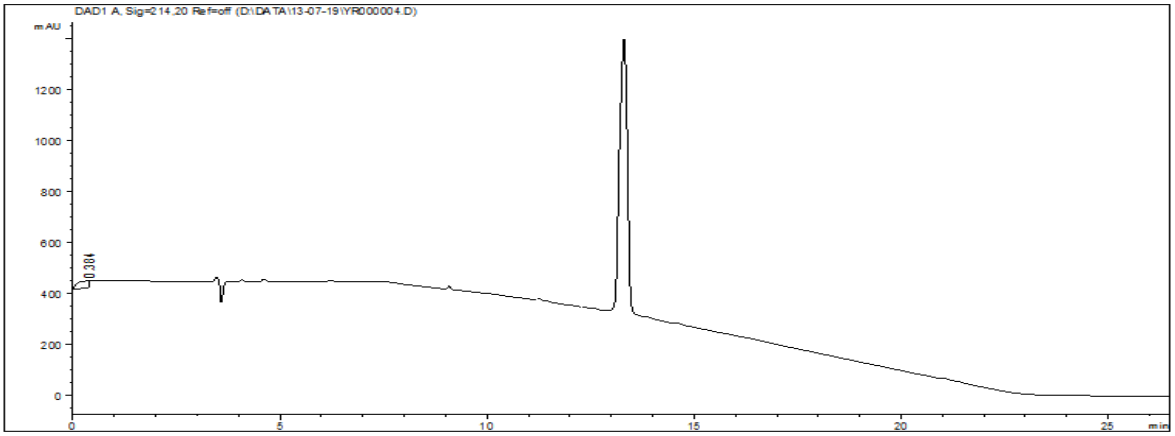


Figure 20: RP-HPLC Chromatogram of pure compound **9** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes)

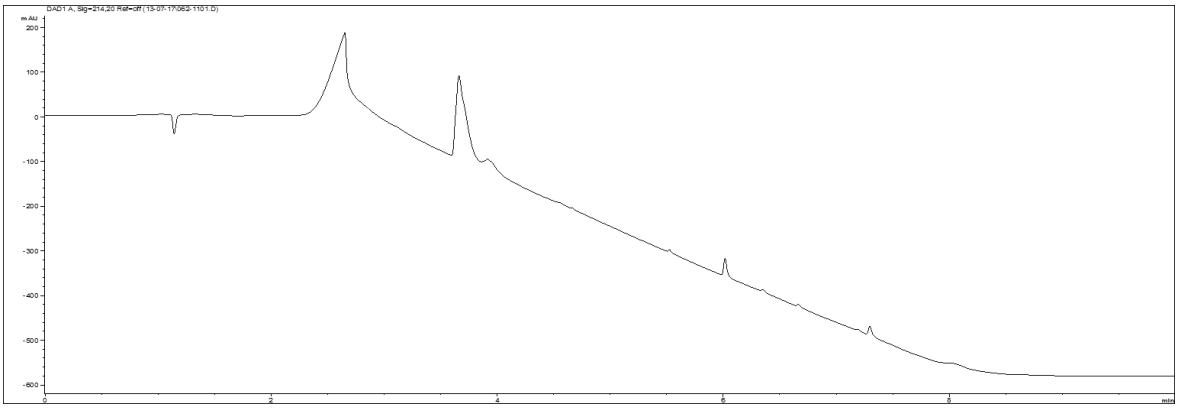


Figure 21: RP-HPLC trace of compound **9** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

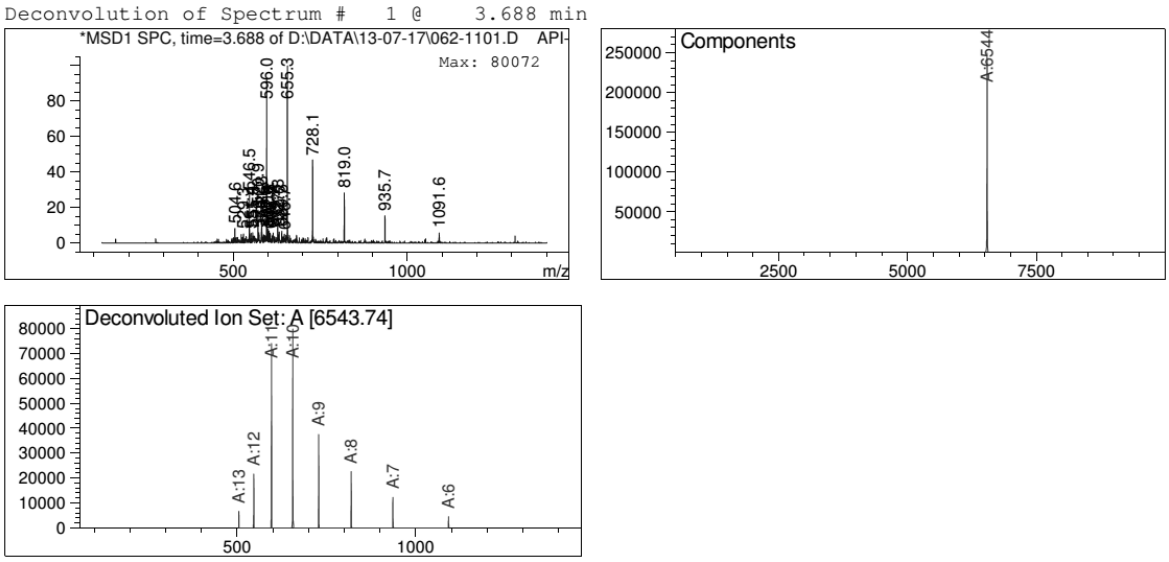
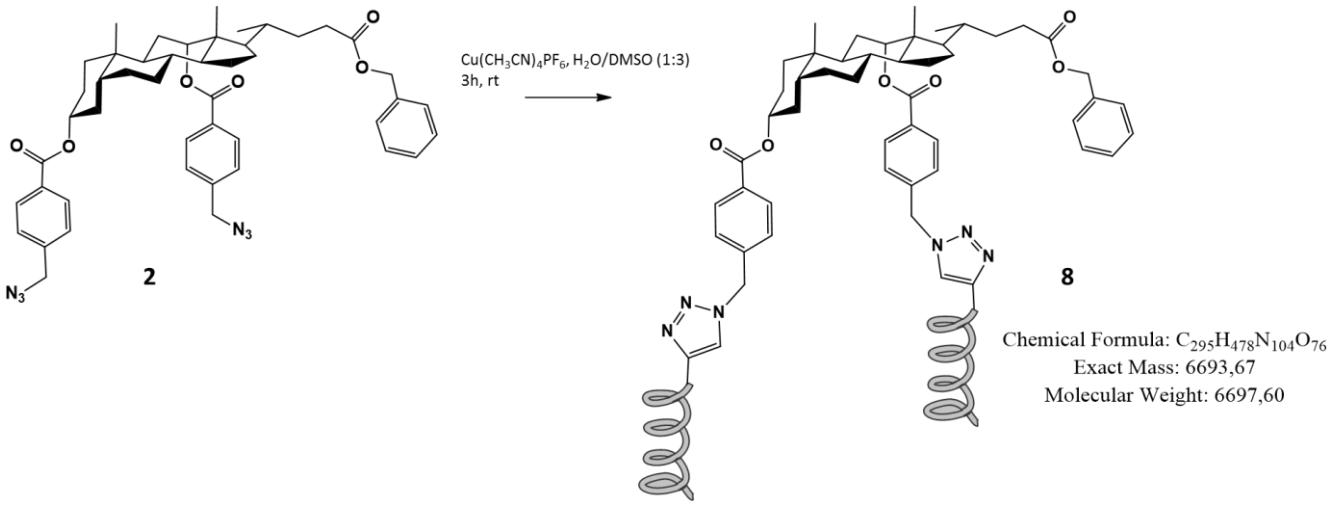


Figure 22: ESI-MS spectrum from LC-MS at t = 3.688 min of compound **9**. EM cacld. For $C_{283}H_{470}N_{104}O_{76}$ = 6541.61 and deconvoluted mass found 6543.74.

Azidobenzoic acid (8): Scaffold **2** (4 mg, 5 µmol) was dissolved in 0.2 mL dry DMSO in a 5 mL round bottom flask. Peptide **5** (3.7 mg, 1.2 µmol) was dissolved in 0.1 mL miliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (4.68 mg, 13 µmol) was dissolved in 75 µL dry DMSO and added it to the reaction mixture. The reaction was stirred for 3 h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC. Fractions were lyophilized and analyzed by RP-HPLC and MALDI-TOF.



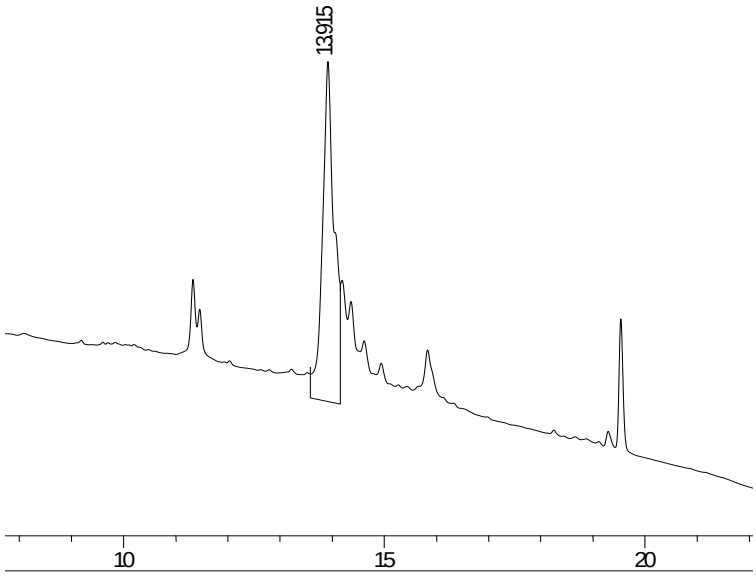


Figure 23: RP-HPLC trace of HPLC crude compound **8** eluting at 13.9 min and excess of scaffold at 19.6 min (0-100% ACN in 15 min on Jupiter c4 300 Å, 250 x 2.1 mm, 2.6 µm, at 35 °C)

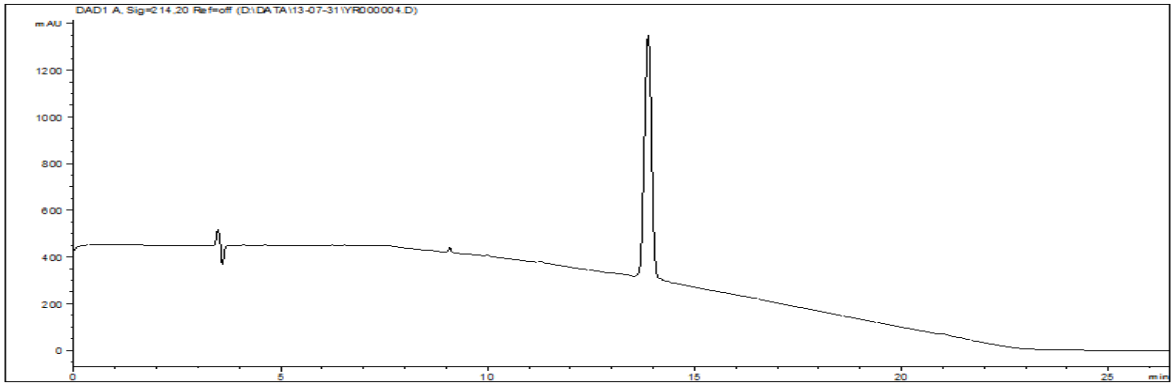


Figure 24: RP-HPLC Chromatogram of pure compound **8** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes)

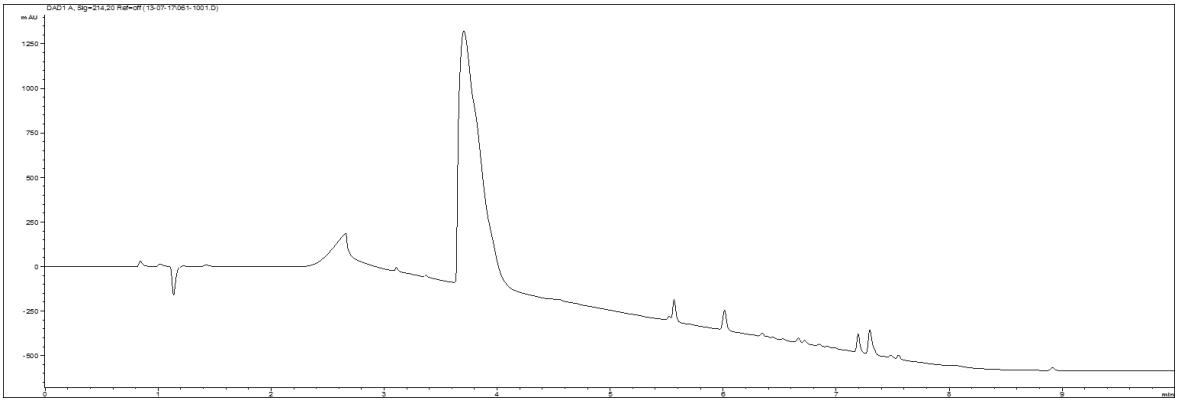


Figure 25: RP-HPLC trace of compound **8** (0-100% ACN in 6 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

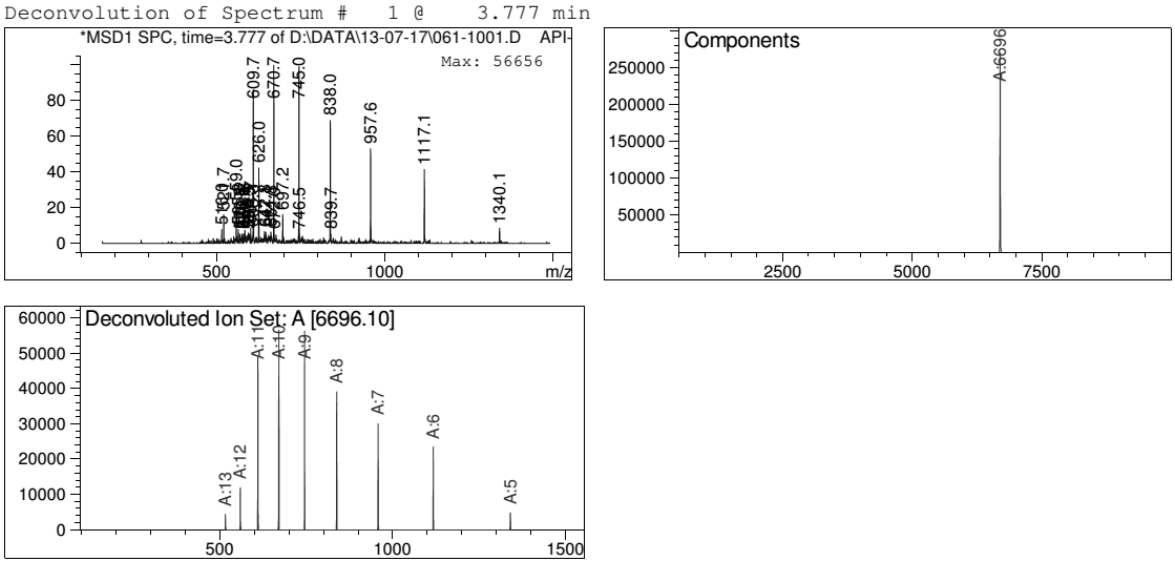
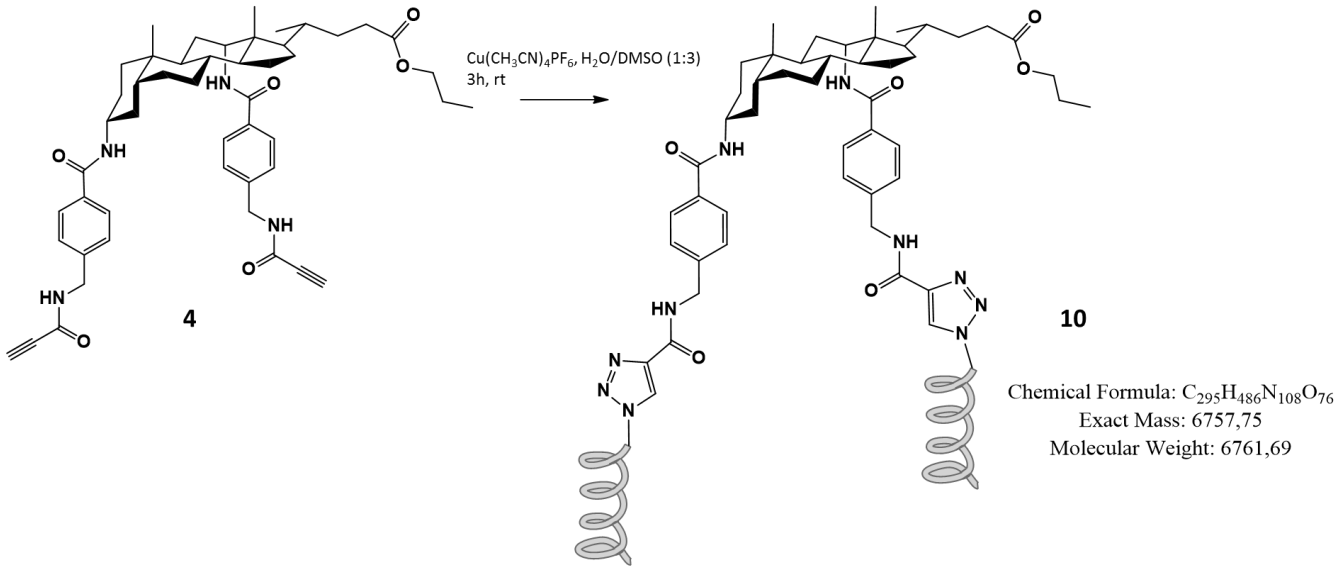


Figure 26: ESI-MS from LC-MS at t = 3.777 min of compound **8**. EM calcd. for $C_{295}H_{478}N_{104}O_{76}$ = 6693.97 and deconvoluted mass found 6696.10

PATA linker (10): Scaffold **4** (4 mg, 5 µmol) was dissolved in 0.2 mL dry DMSO in a 5 mL round bottom flask. Peptide **6** (3.7 mg, 1.2 µmol) was dissolved in 0.1 mL miliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (4.63 mg, 12 µmol) was dissolved in 75 µL dry DMSO and added it to the reaction mixture. The reaction was stirred for 3 h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC. Fractions were lyophilized and analyzed by RP-HPLC and LC-MS.



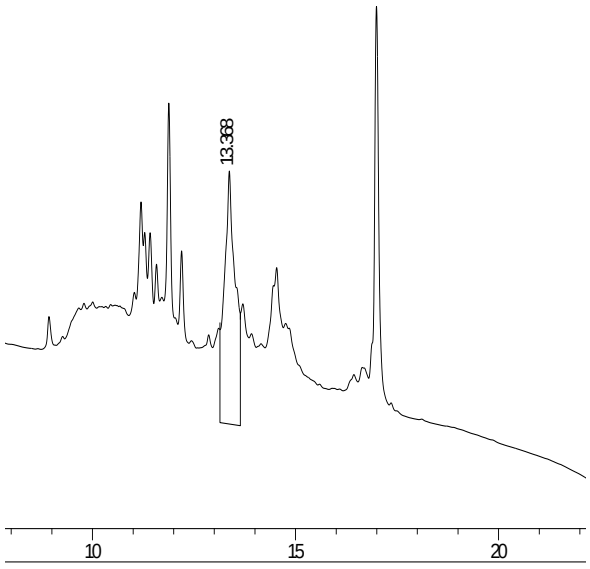


Figure 27: RP-HPLC trace of HPLC crude compound **10** eluting at 13.3 min and excess of scaffold at 16.9 min (0-100% ACN in 15 min on Jupiter c4 300 Å, 250 x 2.1 mm, 2.6 µm, at 35 °C)

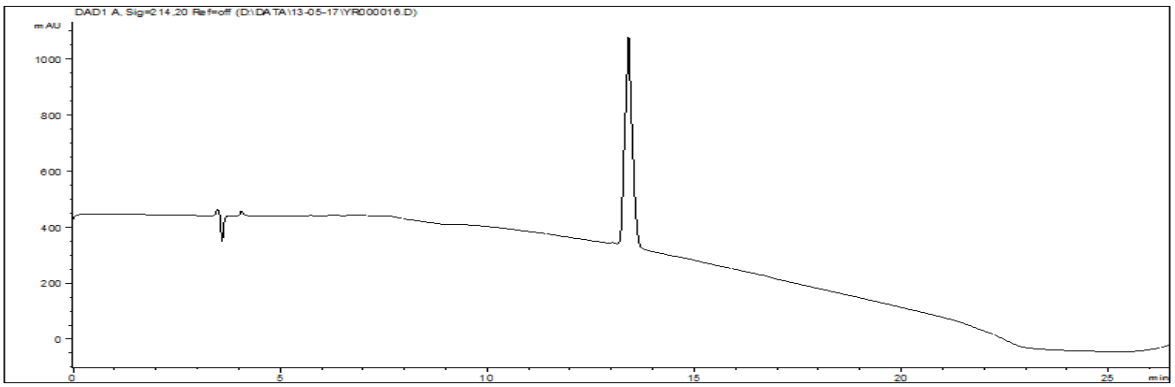


Figure 28: RP-HPLC Chromatogram of pure compound **10** (C4, 300Å column using a gradient from 0 to 100 % CH_3CN in 15 minutes)

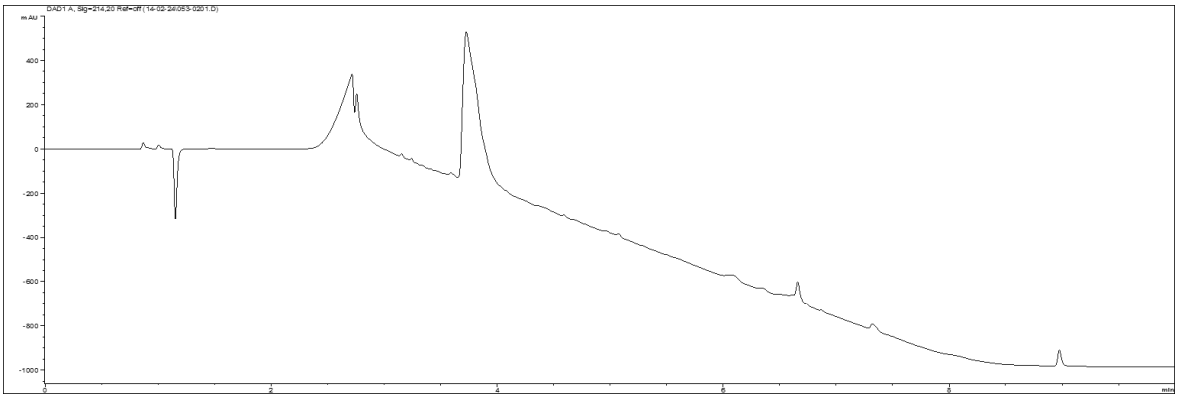


Figure 29: RP-HPLC of compound **10** (0-100% ACN in 15 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

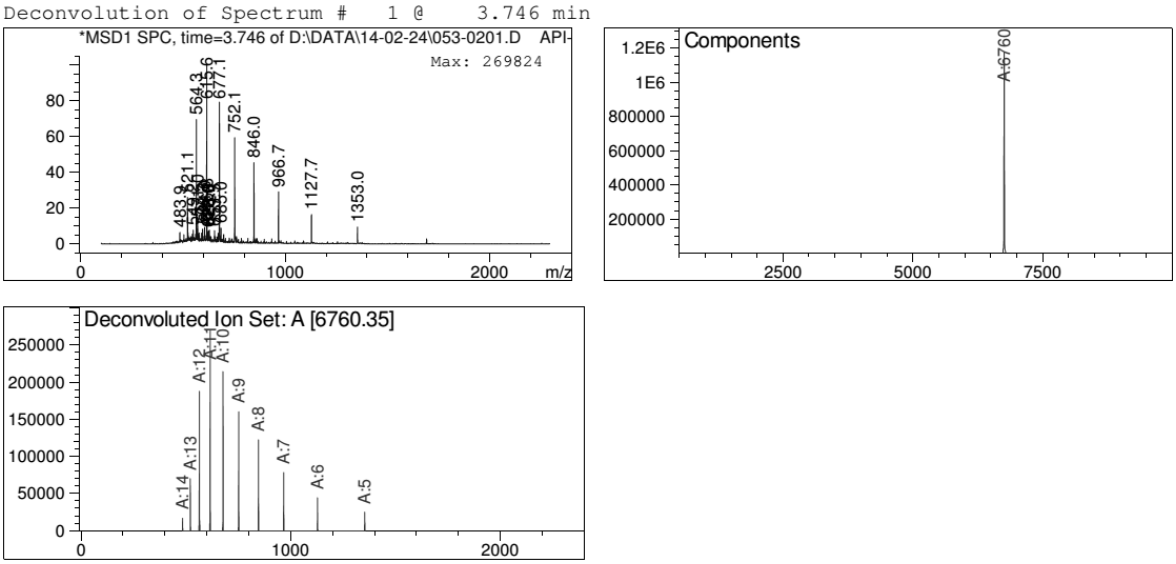
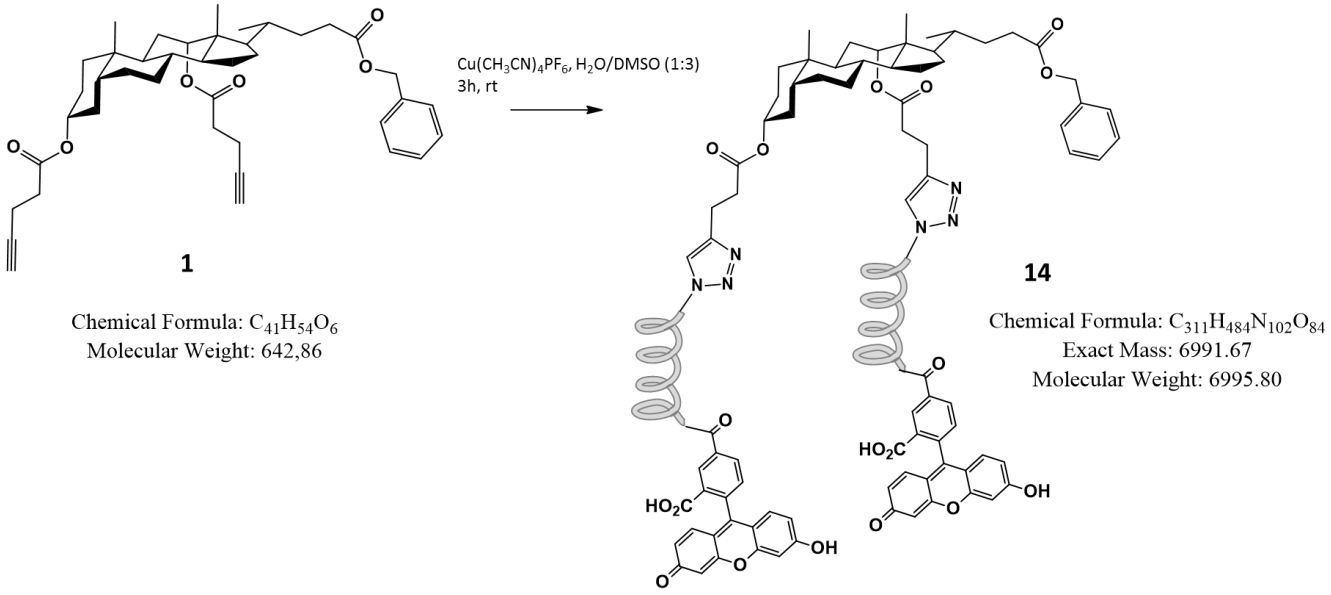


Figure 30: ESI-MS spectrum from LC-MS at t = 3.746 min of compound **10**. EM calcd. for $C_{295}H_{486}N_{108}O_{76}$ = 6757.75 and deconvoluted mass found 6760.35

4. II. COMPOUNDS FOR CELL PENETRATION

Synthesis of 14: Scaffold **1** (0.27 mg, 0.4 µmol) was dissolved in 0.2 mL dry DMSO in a 5 mL round bottom flask. Peptide **11** (4 mg, 1.2 µmol) was dissolved in 0.1 mL miliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (2.34 mg, 6.3 µmol) was dissolved in 75 µL dry DMSO and added it to the reaction mixture. The reaction was stirred for 3 h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC. Fractions were lyophilized and analyzed by RP-HPLC and MALDI-TOF.



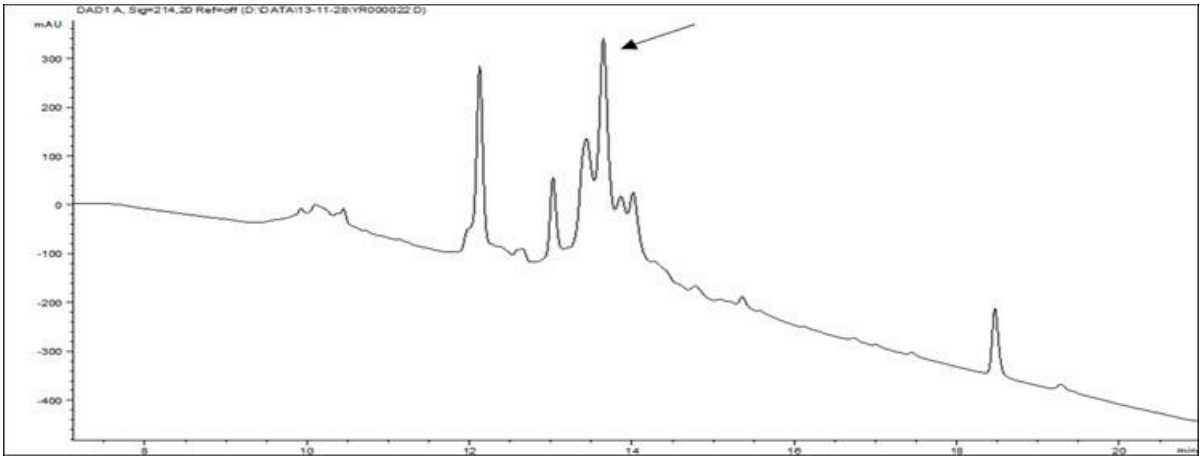


Figure 31: RP-HPLC trace of HPLC crude compound **14** (0-100% ACN in 15 min on Jupiter c4 300 Å, 250 x 2.1 mm, 2.6 µm, at 35 °C)

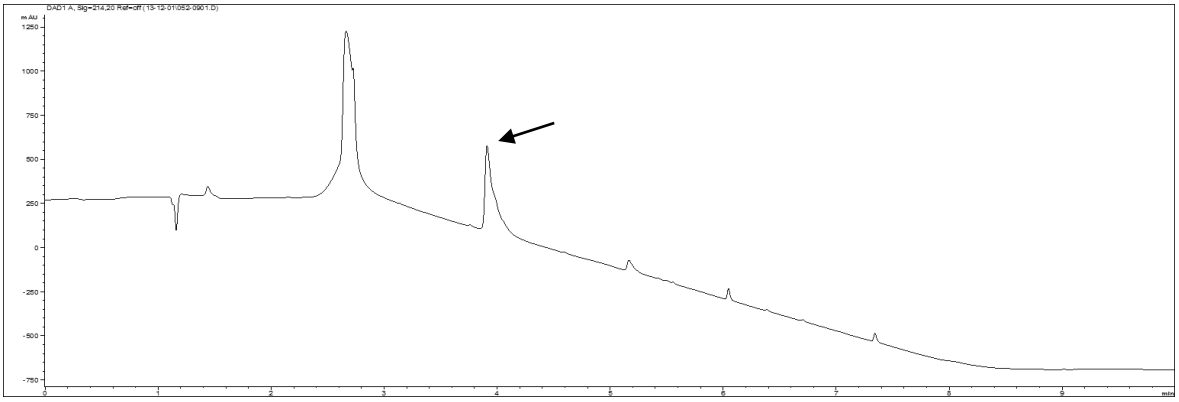


Figure 32: RP-HPLC trace of compound **14** (0-100% ACN in 15 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

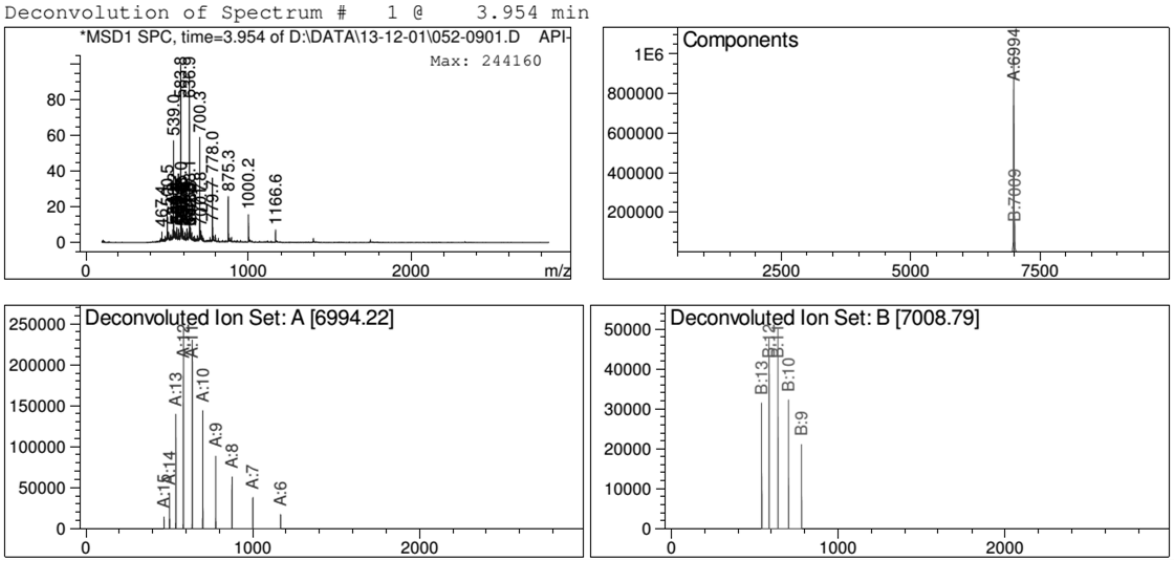


Figure 33: ESI-MS from LC-MS at t = 3.954 min for pure compound **14**. E.M calcd. for $C_{311}H_{484}N_{102}O_{84}$ = 6991.67 and E.M. + NH_4^+ = 7009.67 and deconvoluted mass found 6994.22 and 7008.79 respectively.

Synthesis of 13: Scaffold **1** (3.11 mg, 5 µmol) was dissolved in 0.2 mL dry DMSO in a 5 mL round bottom flask. Peptide **11** (3.85 mg, 1.2 µmol) was dissolved in 0.1 mL miliQ water and added it to the reactor. $Cu(CH_3CN)_4PF_6$ (4.51 mg, 12

µmol) was dissolved in 75 µL dry DMSO and added it to the reaction mixture. The reaction was stirred for 3 h at room temperature under argon. The reaction was monitored by RP-HPLC and purified by fraction collection in RP-HPLC. Fractions were lyophilized and analyzed by RP-HPLC and MALDI-TOF.

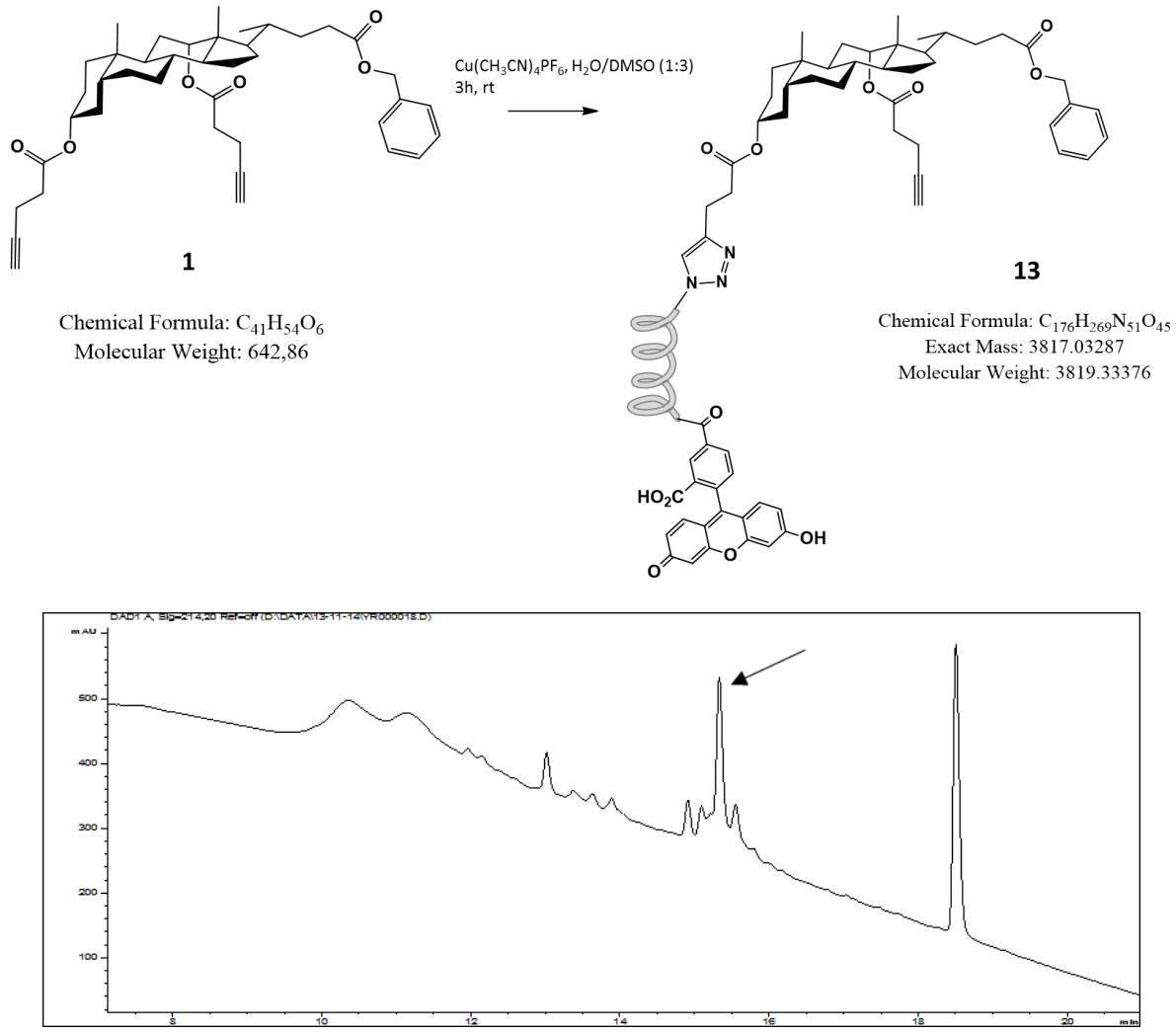


Figure 34: RP-HPLC trace of HPLC crude compound **13** (0-100% ACN in 15 min on Jupiter c4 300 Å, 250 x 2.1 mm, 2.6 µm, at 35 °C)

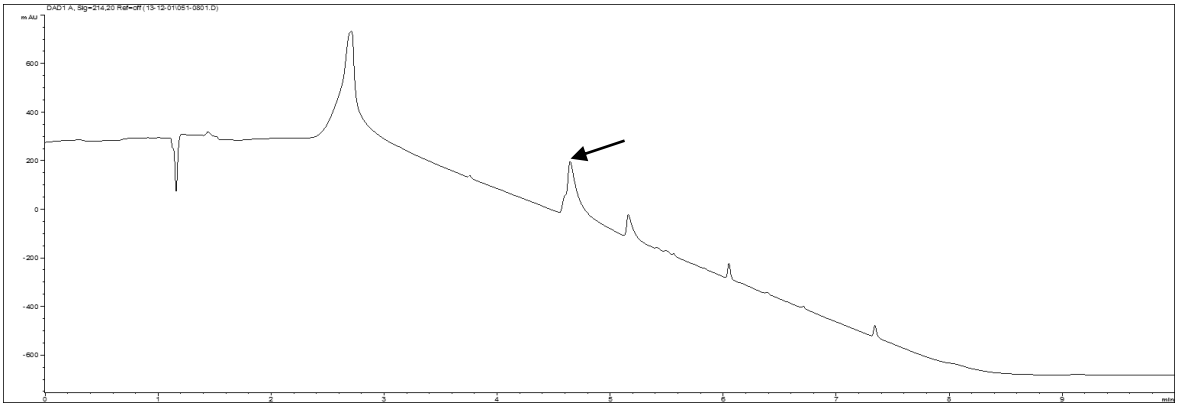


Figure 35: RP-HPLC trace of compound **13** (0-100% ACN in 15 min on Kinetex C18 100 Å, 150 x 2.1 mm, 2.6 µm, at 35 °C)

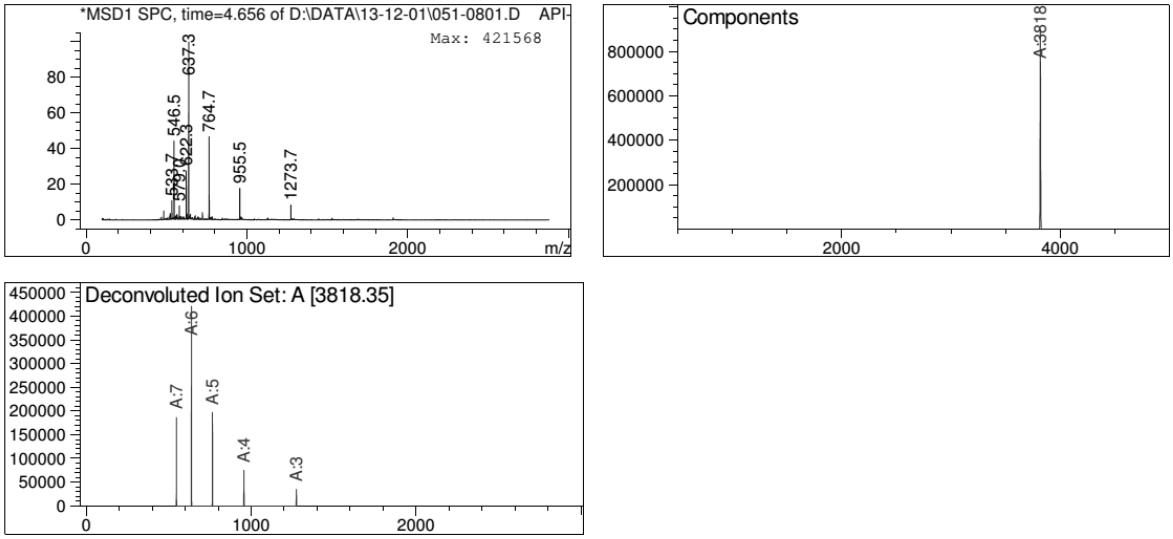


Figure 36: ESI-MS from LC-MS at t = 4.656 min for pure compound **13**. E.M calcd. for $C_{176}H_{269}N_{51}O_{45}$ = 3817.03 and deconvoluted mass found 3818.35.

5. Electrophoretic Mobility Shift Assay

Sample preparation:

The following stock solutions were prepared (fresh each time, except for DNA and peptide):

DNA: 1.67 µM prepared from CRE (5' – CGG ATG ACG TCA TTT TTT TTC – 3') & CRE complement (5' – GAA AAA AAA TGA CGT CAT CCG – 3') and Random (5' – GCG CGA GAA GGA AAG AAA GCC GG – 3') & complement (5' – CCG GCT TTC TTT CCT TCT CGC GC – 3') DNA solutions (commercially obtained) by diluting with 20 µL 0.5 M Tris, pH = 8, 40 µL 2.5 M NaCl, 40 µL 0.025 M EDTA and then adding milliQ water such that the total volume is 1 mL. The DNA was annealed by heating in a Thermomixer from room temperature to 95°C and maintaining the temperature at 95°C in total time of 24 min. The machine was then turned off and the sample was allowed to cool down slowly.

Loading buffer: 20 µL Tris 1 M, pH = 7.6, 20 µL KCl 0.2 M, 20 µL $MgCl_2$ 0.1 M, 40 µL EDTA 0.025 M.

Sucrose: 30% sucrose in mQ (300 mg/mL)

Peptides: 10 µL stock solutions (10x) were prepared in MiliQ water (0, 1.67, 5.01, 6.68, 7.51, 8.35, 10.02, 11.69, 13.36, 16.7 µM)

Loading mixture: The loading mixture comprised of: 10 µL mQ, 4 µL sucrose, 2 µL loading buffer, 2 µL DNA, 2 µL peptide. The loading mixture was prepared only 1-2 h prior to running of gels and kept on ice as soon as ready.

Preparation of Gels (for 2 Gels):

In a clean falcon tube the following reagents were added (in given order): 15.595 mL mQ, 0.4 mL TBE, 4.005 mL of 40% acrylamide solution, 200 µL APS (10% w/w in mQ). The solution was mixed by sonication to remove any air bubbles and cooled to 0°C (1 h under ice). 20 µL of TEMED was then added to the mixture and was again mixed properly before pouring it gently along parallel glass plates. The glass plates were tapped gently to ensure removal of all air bubbles and the markers were squeezed between the plates to ensure uniform width of each well. Sufficient time was given for polymerization (~1 h).

Gel Electrophoresis:

A pre-run of the gels was performed prior to loading them. Care was taken to see that the gels were properly immersed in 0.2x TBE buffer (non-denaturing gel, without urea) and the loading wells were free from any air bubbles. Instrument settings: 150 V, 100 mA, 19 W for 30 mins at 4°C. The wells were washed after the pre-run. 5 µL of the loading mixture was then loaded onto the wells. Instrument settings: 150 V, 100 mA, 19 W for 45 mins at 4°C.

Staining of gels:

After the run, the gels were removed from the glass and were stained using 100 mL of 0.2x TBE buffer + 10 µL Sybr Gold (commercially purchased stock solution 10,000X in DMSO). The gels were then washed twice with mQ and gently placed under a UV lamp (dark room) to observe the gel pattern.

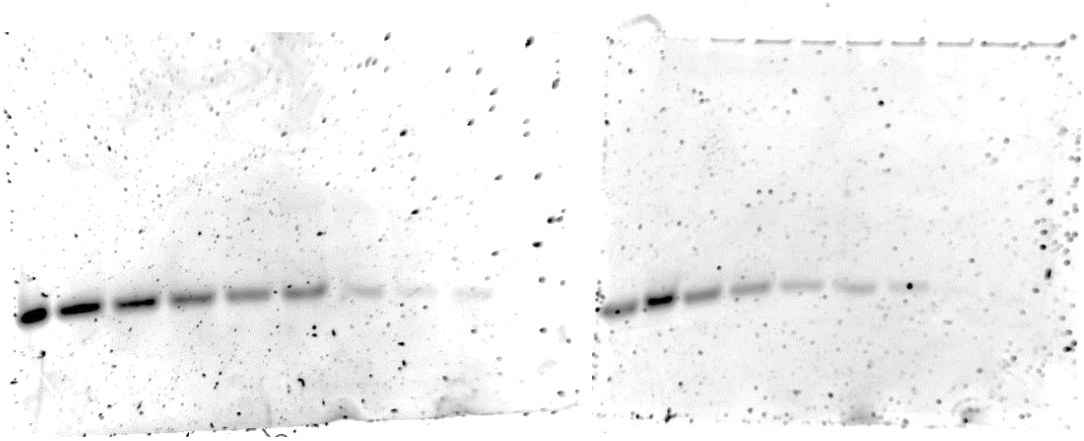


Figure 37 (left). EMSA titration of the dipodal peptidosteroid **7** to the Random dsDNA. Lane 1-10 contain respectively 0, 0.167, 0.501, 0.668, 0.751, 0.835, 1.002, 1.169, 1.336, 1.67 µM of the mimic, stained with SybrGold. (right). EMSA titration of the dipodal peptidosteroid **9** to the Random dsDNA. Lane 1-10 contain respectively 0, 0.167, 0.501, 0.668, 0.751, 0.835, 1.002, 1.169, 1.336, 1.67 µM of the mimic, stained with SybrGold.

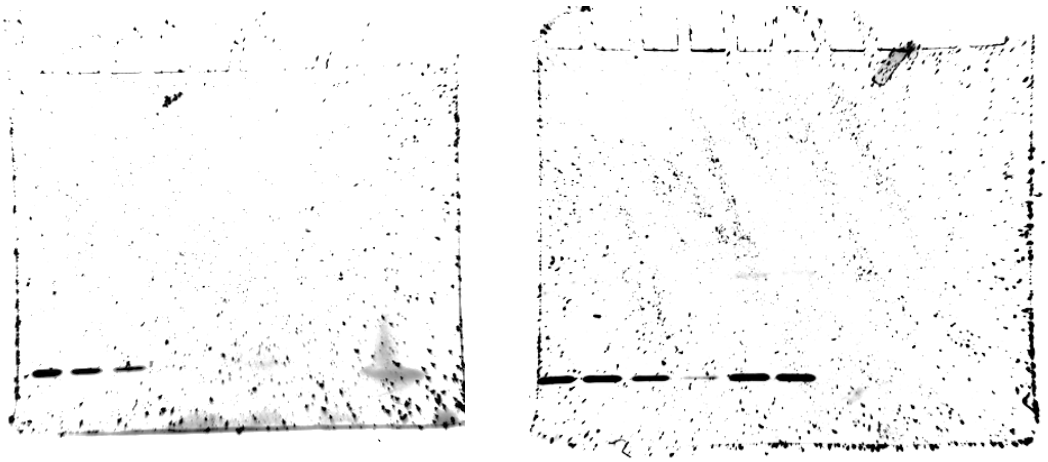


Figure 38 (left). EMSA titration of the dipodal peptidosteroid **8** to the CRE dsDNA. Lane 1-10 contain respectively 0, 0.167, 0.501, 0.668, 0.751, 0.835, 1.002, 1.169, 1.336, 1.67 µM of the mimic, stained with SybrGold (right). EMSA titration of the dipodal peptidosteroid **10** to the CRE dsDNA. Lane 1-10 contain respectively 0, 0.167, 0.501, 0.668, 0.751, 0.835, 1.002, 1.169, 1.336, 1.67 µM of the mimic, stained with SybrGold (left).

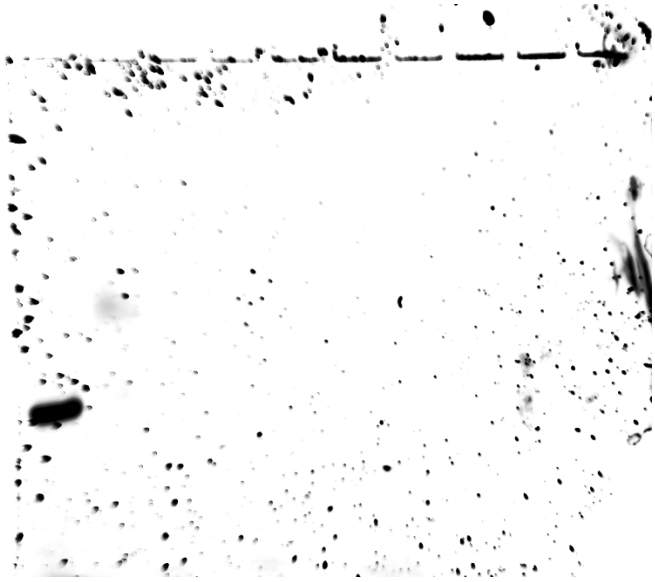


Figure 39. EMSA titration of the dimeric GCN4 **12** to the CRE binding site. Lane 1-10 contain respectively 0, 0.25, 0.5, 0.68, 0.75, 1.0, 1.25, 1.5, 2.0, 2.5 µM of the mimic, stained with SybrGold.

ELECTROPHORETIC MOBILITY SHIFT ASSAY (EMSA) FOR QUANTIFICATION OF THE DISSOCIATION CONSTANT.

Preparation of ^{32}P-labeled dsDNA target

Oligonucleotide CRE (5' – CGG ATG ACG TCA TTT TTT TTC – 3') was 5'-labeled using [γ-^{32}P] ATP and T4 polynucleotide kinase (Fermentas) according to the manufacturer's protocol and then purified using QIAquick Nucleotide Removal Kit (Qiagen). The 5'-end labeled pyrimidine oligonucleotide was annealed with the unlabeled complementary strand. An amount of 5 nM dsDNA was prepared diluting 20 µL 0.5 M Tris, pH = 8, 40 µL 2.5 M NaCl, 40 µL 0.025 M EDTA and then adding milliQ water such that the total volume is 1 mL. The DNA was annealed in a heat block by heating from 95°C during 5 minutes followed by slow cooling to room temperature. As competitor DNA the following sequence (5'-AGCAGAGGGCGTGGGGGAAAAGAAAAAAGATCCACCGGTCGCCAC-3') and its complementary strand was used.

Loading buffer: 20 µL Tris 1 M, pH = 7.6, 20 µL KCl 0.2 M, 20 µL $MgCl_2$ 0.1 M, 40 µL EDTA 0.025 M.

Sucrose: 30% sucrose in mQ (300 mg/mL)

Peptides: 10 µL stock solutions (10x) were prepared in MiliQ water (0, 0.6, 1.25, 2.5, 5, 10, 20 µM)

Loading mixture: The loading mixture comprised of: 10 µL mQ, 4 µL sucrose, 2 µL loading buffer, 2 µL DNA, 2 µL peptide. The loading mixture was prepared only 1 h prior to running of gels and kept on ice as soon as ready.

Preparation of Gels (for 1 Gel):

In a clean glass beaker the following reagents were added (in given order): 21.57 mL mQ, 0.6 mL TBE 10X, 7.5 mL of 40% acrylamide solution (29:1), 0.3 mL APS (10% w/w in mQ) and 30 µL of TEMED was then added to the mixture and mixed properly before pouring it gently along parallel glass plates. The glass plates were tapped gently to ensure removal of all air bubbles and the markers were squeezed between the plates to ensure uniform width of each well. Sufficient time was given for polymerization (40 minutes).

Electrophoresis:

A pre-run of the gels was performed prior to loading them. Care was taken to see that the gels were properly immersed in 0.2x TBE buffer (non-denaturing gel, without urea) and the loading wells were free from any air bubbles. The wells were washed after the pre-run. Instrument settings: 150 V, 100 mA, 19 W for 30 minutes with circulation water-cooling. 5 µL of the loading mixture was then loaded onto the wells. Instrument settings: 150 V, 100 mA, 19 W for 45 minutes with circulation water-cooling.

The gels were frozen and analyzed by phosphor imaging using Molecular Imager FX and the data were processed using Quantity One software (BioRad).

From the electrophoresis experiments with competitor DNA, we could calculate K_D values of **7** and **9** of the binding to the CRE sequence**.**

Calculation of the dissociation constant (K_D)

Molecular Imager FX and the Quantity One software (BioRad) were used for scanning and analysis the gel, respectively. The band intensities were fitted in the equation: K_D = [Peptide][D]/[SB] where [D] and [SB] are the concentrations of target DNA and shifted band, respectively. By assuming a 1:1 stoichiometry of Peptide and duplex DNA and an unknown K_D. We assume that the concentration of free Peptide is essentially constant and equal to its initial concentration C0, because the Peptide is in excess over the target DNA. The equation used for the curve fit was $\%D=100/(1+C0/K_D)$[5,6].

Table 1

Experimental set-up	**K_D (nM) of 7**	**K_D (nM) of 9**
Peptide/duplex ratio 1:100 (figure 41)	76.7 ± 6.2	136.7 ± 21.9
Peptide/duplex ratio 1:2000 exp. 1 (figure 42)	44.5 ± 22.3	269.6 ± 198.9
Peptide/duplex ratio 1:2000 exp. 2 (figure 43)	49.3 ± 32.5	273.1 ± 204.8

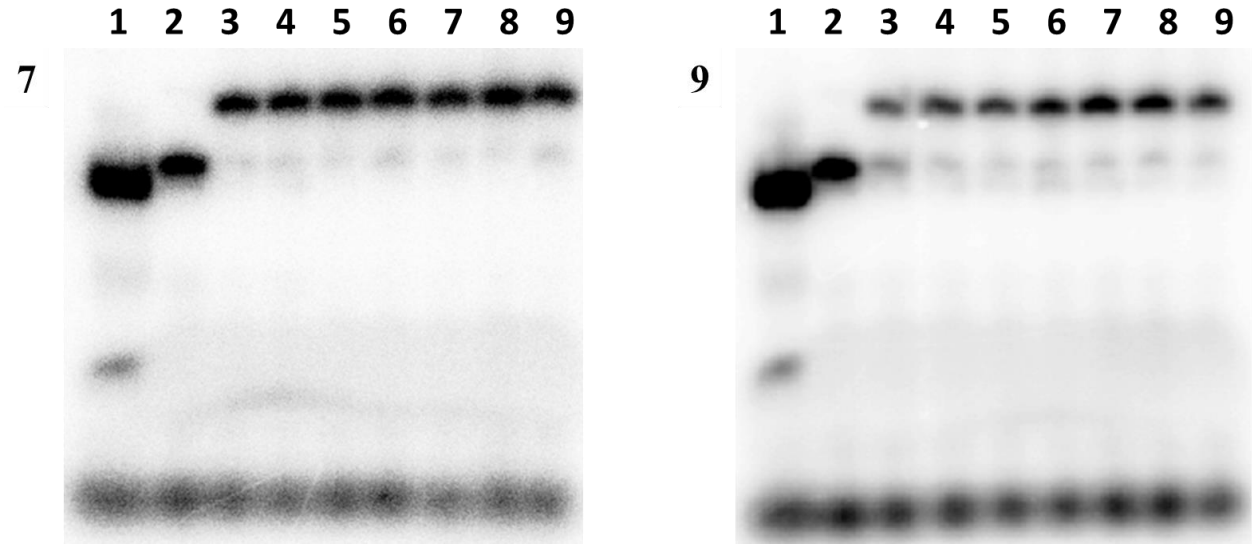


Figure 40. EMSA titration of the dipodal peptidosteroid conjugates **7** and **9** to the 5'-labeled ^{32}P-CRE sequence (5' – CGG ATG ACG TCA TTT TTT TTC – 3') at 5 nM: First lane in all the gels: pyrimidine strand. Lanes 2-9 contain peptide concentrations of 0, 0.05, 0.0625, 0.075, 0.0875, 0.1, 0.1125 and 0.125 µM for **7** and **9**.

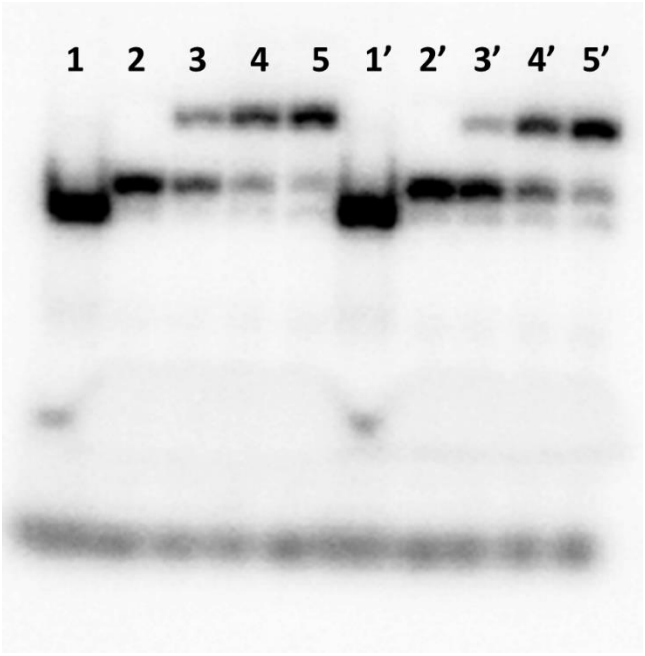


Figure 41. EMSA titration of the dipodal peptidosteroid conjugates **7** (lanes 1-5) and **9** (lanes 1'-5') to the 5'-labeled ^{32}P-CRE sequence (5' – CGG ATG ACG TCA TTT TTT TTC – 3') at 5 nM in the presence of competitor DNA sequence (5'-AGCAGAGGGCGTGGGGGAAAAGAAAAAAGATCCACCGGTCGCCAC-3') at 500 nM: First lane in all the gels: pyrimidine strand. Lanes 2 and 2' to 5 and 5' contain peptide concentrations of 0, 0.05, 0.125 and 0.312 µM for **7** and **9** respectively.

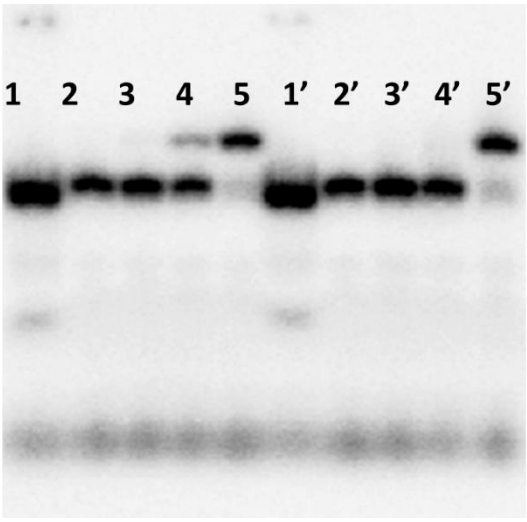


Figure 42. (Experiment 2). EMSA titration of the dipodal peptidosteroid conjugates **7** and **9** to the 5'-labeled ^{32}P-CRE sequence (5' – CGG ATG ACG TCA TTT TTT TTC – 3') at 1 nM in the presence of competitor DNA sequence (5'-AGCAGAGGGCGTGGGGGAAAAGAAAAAAGATCCACCGGTCGCCAC-3') at 2 µM: First lane in all the gels: pyrimidine strand. Lanes 2 and 2' to 5 and 5' contain peptide concentrations of 0, 0.01, 0.025 and 0.0625 µM for **7** and **9** respectively.

Figure 43. (Experiment 3). EMSA titration of the dipodal peptidosteroid conjugates **7** and **9** to the 5'-labeled ^{32}P-CRE sequence (5' – CGG ATG ACG TCA TTT TTT TTC – 3') at 1 nM in the presence of competitor DNA sequence (5'-AGCAGAGGGCGTGGGGGAAAAGAAAAAAGATCCACCGGTCGCCAC-3') at 2 μM: Lanes 1 and 1' to 4 and 4' contain peptide concentrations of 0, 0.01, 0.025 and 0.0625 μM for **7** and **9** respectively.

PREVIOUS MODELS

We hereby include a summary to understand where the constructs synthesized stand in terms of relative binding affinity to the CRE compared to previously published models by various groups[7]. Due to differences in the length of the peptide chain and the conditions used for gel electrophoresis including different DNA concentrations, buffers, acrylamide percentages in gels, etc. a direct comparison is difficult. Therefore, we have only chosen those models which have the same peptide length as in the current models.

The construct synthesized by Morii[8] has a K_D < 100 nm. The construct is based on the non-covalent interaction between cyclodextrin and adamantane. Although a K_D value has not been published, it can be derived approximately from the EMSA data shown in the article. This extrapolated value is comparable to our models, although based on our calculations the steroid-peptide conjugate **7** is somewhat better than the construct by Morii in addition to being synthesized in a convergent manner.

The mimic developed by the lab of Mascareñas[9] based on the photo switchable diazobenzene linker in the cis form has a K_D < 5 nm as stated in the article. The dissociation constant of this construct is lower than our best construct (K_D ≈ 44 nm).

Yet another TF design from the lab of Mascareñas[10] has a K_D of 299 ± 26 nM for CRE. The added advantage of this TF mimic is the ability to bind selectively either CRE or its inverse sequence by changing the reaction conditions. This ability of modulating the DNA recognition makes the design unique. However, if only CRE DNA is to be targeted this mimic has a K_D that is nearly an order of magnitude higher than the steroid scaffold based **7**.

The fourth and final construct although published by the lab of Mascareñas and Vázquez is actually a shortened version of the dimer from the mimic of Kim[11]. The K_D, although not reported is <150 nm based on the EMSA data from the article. This again appear to be higher than for our models and hence displaying lower binding affinity to the CRE.

6. DLS Measurements

A 100 µL of 1 µM was prepared in a small cuvette in filtered Dulbecco's Modified Eagle culture medium (DMEM-CM). The Dynamic light scattering was measured with the Zetasizer Nanoseries Nano-S (Malvern). Data was processed by Graph Pad.

DLS was measured in a solution of 1 µM peptide concentration in filtered DMEM-CM. Compound **13** was measured in an overall concentration of 2 µM. The DMEM contains 10 % (v/v) of FBS. Figure 40 shows the size distribution of a mixture of the peptides and cell culture medium separated by a small shift. No complexation for the GCN4 peptide **11** occurred because the size distribution in function of volume fraction was similar to that of filtered DMEM-CM. The compounds with a deoxycholic acid scaffold (**13, 14**) complexated with serum proteins to some extent probably due to their higher hydrophobicity. However, by comparing the volume fraction of these peptides to the one of pure DMEM, it was clear that almost no complexation occurred due to the limited shift of the size distribution.

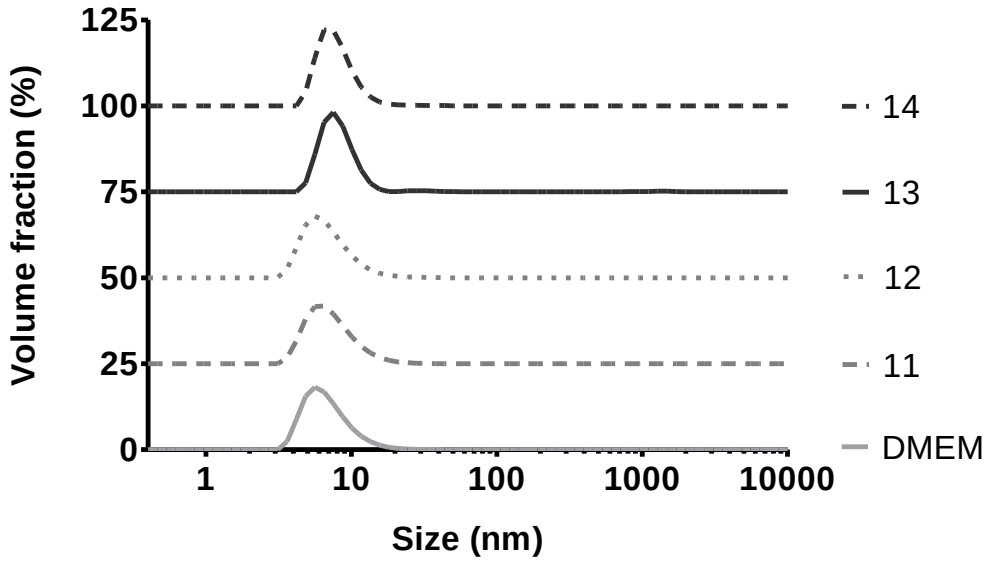


Figure 44: Distribution of size of particles in solution in function of volume fraction

In order to further estimate the influence of particles taken up by the cell, fluorescent microscopy was used. A drop of compound **14** solution on a cover glass is evaluated by fluorescent microscopy. The presence of a large amount of BSA peptide complex should show a dotted pattern in fluorescence microscopy. In this case, a homogeneous solution is present which means that complexation occurs to a very small extent (figure 44). Therefore, it is unlikely that complexation will positively influence the cellular uptake.

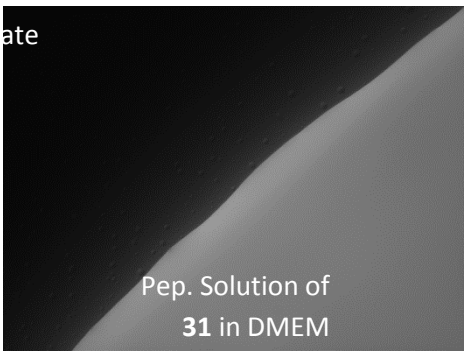


Figure 45: fluorescence microscopy image showing the absence of BSA-protein complexes

7. MTT Assay

RAW264.7 cells cultivated in DMEM (10^4/well, 200 µL) were plated in a 96-well plates and incubated overnight at 37 °C and 5 % CO_2. The peptides and control compounds were added in an overall concentration of 0.25 µM and 1 µM and incubated for 24h under the same conditions. The MTT solution was then added to the aspirated wells and incubated for 3 h. After removal of the cell medium, purple formazan crystals were dissolved in DMSO. Then, UV-measurement at 570 nm was performed with a plate reader to check cell viability quantitatively. An observation with the naked eye already gave a good idea about the toxicity of the compounds due to the disappearance of color in the well. Comparison of the absorbance of the formazan solution of the sample to the absorbance of a positive (incubation with ultrapure water) and a negative control (incubation with DMSO) gave quantitative results of cell viability as in the equation:

$$Cell\ viability\ (\%) = \frac{A - A_{pos}}{A_{neg} - A_{pos}} * 100$$

Cell viability was evaluated according to the ISO10993-5 norms which say that compounds are cytotoxic if the assay points to a cell viability lower than 70 %. Figure 42 shows the percentage of viable cells for all peptides. At 0.25 µM, the cell viability of **13** was close to 70 %. On the other hand, at 1 µM, the deoxycholic scaffold containing compounds (**13, 14**) and the two dimerized GCN4-strands (**12**) were cytotoxic. This was not unexpected because bile acids have detergent like properties that can damage the membrane resulting in cell death.

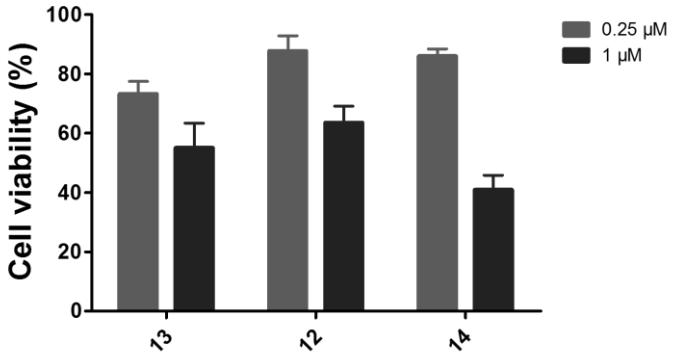


Figure 46: MTT assay for peptide measured at 0.25 µM and 1 µM

8. Confocal Microscopy

RAW264.7 cells cultivated in DMEM (10^5/well, 300 µL) were plated in a confocal plates and incubated overnight at 37 °C and 5 % CO_2. Peptides were added in an overall concentration of 0.25 µM and incubated for 2 h at the same conditions or on ice. Cells were washed and fixated with 2 % of Paraformaldehyde for 30 min at 37°C. Cells were stained by 30 min incubation with a 0.2 % solution of CTB-AF647 and 0.2 % of Hoechst in PBS with 1 % BSA. This was done to visualize the cell nucleus (blue) and cell membrane (red) respectively. Cells were resuspended in PBS and measured with confocal microscope (Leica SP5 equipped with a 63x (1.4 NA) oil immersion objective) at 3 different wavelengths (405 nm, 488 nm and 643 nm). The peptides used were labeled with fluorescein which emits green light. To understand the uptake mechanism better, three samples are made in duplicates and incubated at 4 °C to inhibit active transport. Due to the difference in emission wavelength, the three dyes can be detected separately. Processing the data with Image J gave overlay images whereby the nucleus, cell membrane and fluorescein are shown in blue, red and green respectively.

In order to evaluate the influence of the concentration of the steroid-peptide conjugate **11** on cellular uptake, confocal microscopy was also performed at a concentration of 1 µM (figure 47).

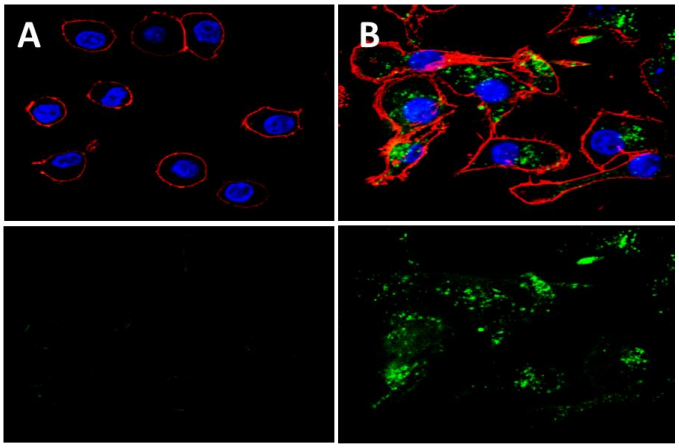


Figure 47: Confocal microscopy images of RAW264.7 cells incubated with compound **11 a**t (A) 4 °C and (B) 37 °C. Cell nuclei were labeled with Hoechst (blue) and cell membranes with AlexaFluor647 conjugated cholera toxin subunit B (red). The lower panels only show the green fluorescence channel.

9. Flow cytometry

RAW264.7 cells cultivated in DMEM (10^5/well, 1 ml) were plated in a 24-well plates and incubated overnight at 37 °C and 5 % CO_2. Peptides were added in an overall concentration of 0.25 µM and 1 µM and incubated for 2h at the same conditions or on ice. Cells were washed with PBS and detached with Na_4EDTA. Cells were re-suspended in PBS and added to the BD Accuri flow cytometer. Experiments were carried out in duplicate. The amount of peptide associated to the cell was measured by flow cytometry. Blank cells were measured and the signal received corresponded to auto-fluorescence. A threshold was chosen to separate cells emitting light due to auto-fluorescence from cells containing fluorescently labeled peptides. The area under the histograms beyond this threshold correspond to the number of cells which actually emit sufficient intensity of fluorescence to make sure fluorescently labeled peptides are present. This number is further referred to as the percent of peptide positive cells. The percent of peptide positive cells for two peptides may be 100 % yet one may better due to its higher mean fluorescence value. If a peptide is not taken up by all cells, the percentage of peptide positive cells is lower than 100 % proportional to the colored area under the curve. It is therefore important to both compare the mean fluorescence and the amount of peptide positive cells.

During flow cytometry measurements no comparison was possible between the results of single fluorescently labeled peptides and the peptides containing two fluoresceins. Therefore the single fluorescently labeled peptidse **11** and **13** are discussed separately from the ones containing two fluoresceins (**12** and **14**).

A concentration dependent uptake was observed for all peptides whereby both mean fluorescence and percent of peptide positive cells increased upon increasing the concentration. Compound **13** which is a single GCN4-strand grafted on a deoxycholic scaffold had substantially higher uptake (figure 48). These findings emphasize the importance of conjugating peptides to a bile acid.

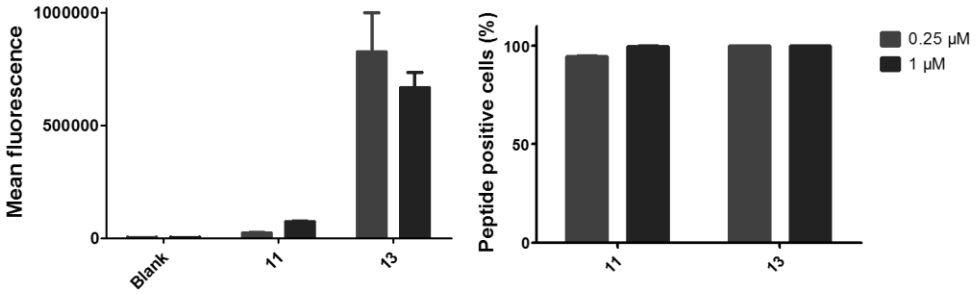


Figure 48: Mean fluorescence (left) and peptide positive cells (right) of fluorescently labeled peptides **11** and **13.**

The increased uptake of **13** at 0.25 µM and 1 µM compared to the other peptides was not unexpected. However, a decrease in mean fluorescence was observed when switching from 0.25 µM to 1 µM. The tendency of a bile acid to disrupt the cell membrane due to its detergent like properties can be responsible for a maximum uptake level. However, the reason for the decrease in the mean fluorescence when increasing the concentration from 0.25 µM to 1 µM is unknown.

10. References:

[1] S. Akoka, L. Barantin, M. Trierweiler, *Anal. Chem.* **1999**, *71*, 2554–7.

[2] D. Verzele, J. L. Goeman, A. Madder, *Ark. (Gainesville, FL, U. S.)* **2007**, 325–336.

[3] D. Verzele, A. Madder, *Eur. J. Org. Chem.* **2007**, 1793–1797.

[4] M. Wenska, M. Alvira, P. Steunenberg, A. Stenberg, M. Murtola, R. Stroemberg, *Nucleic Acids Res.* **2011**, *39*, 9047–9059.

[5] Y. Ruiz García, J. Zelenka, V. Pabon, A. Iyer, M. Buděšínský, T. Kraus, C. I. E. Smith, A. Madder, *Org. Biomol. Chem.* **2015**, DOI 10.1039/C5OB00609K.

[6] H. Bergquist, A. Nikravesh, R. D. Fernández, V. Larsson, C.-H. Nguyen, L. Good, R. Zain, *Chembiochem* **2009**, *10*, 2629–37.

[7] E. Pazos, J. Mosquera, M. E. Vázquez, J. L. Mascareñas, *Chembiochem* **2011**, *12*, 1958–73.

[8] M. Ueno, A. Murakami, K. Makino, T. Morii, *J. Am. Chem. Soc.* **1993**, *115*, 12575–12576.

[9] A. M. Caamano, M. E. Vazquez, J. Martinez-Costas, L. Castedo, J. L. Mascarenas, *Angew. Chem., Int. Ed.* **2000**, *39*, 3104–3107.

[10] J. Mosquera, A. Jiménez-Balsa, V. I. Dodero, M. E. Vazquez, J. L. Mascarenas, *Nat. Commun.* **2013**, *4*, 1874–1878.

[11] A. Jiménez-Balsa, E. Pazos, B. Martínez-Albardonedo, J. L. Mascareñas, M. E. Vázquez, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 8825–9.