

# Sub-monomeric strategy with minimal protection for the synthesis of C(2)-modified Peptide Nucleic Acids

Stefano Volpi,<sup>a</sup> Andrea Rozzi,<sup>a</sup> Nicola Rivi,<sup>a</sup> Martina Neri,<sup>a</sup> Wolfgang Knoll,<sup>a,b</sup> Roberto Corradini<sup>a\*</sup>.

<sup>a</sup>Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parco Area delle Scienze 17/A, Parma, 43123, Italy. <sup>b</sup> Biosensor Technologies' AIT-Austrian Institute of Technology GmbH, Konrad-Lorenz-Straße 24, 3430, Tulln an der Donau, Austria.

## Supporting Information

### Table of contents

<b>1</b>	<b>General</b> .....	<b>3</b>
<b>2</b>	<b>Synthesis of the minimally protected submonomers</b> .....	<b>4</b>
2.1	C(2)-Lys-based submonomers .....	4
2.2	C(2)-Arg-based submonomers.....	5
<b>3</b>	<b>Fmoc deprotection of the minimally-protected submonomers</b> .....	<b>6</b>
<b>4</b>	<b>Synthesis of cytosine(Boc)-1-acetic acid</b> .....	<b>8</b>
<b>5</b>	<b>PNA synthesis</b> .....	<b>9</b>
5.1	Unmodified PNAs.....	9
5.2	'Chiral box' PNAs.....	10
<b>6</b>	<b>Derivatization procedures for enantiomeric analysis</b> .....	<b>12</b>
<b>7</b>	<b>CG-MS analysis</b> .....	<b>13</b>
7.1	Achiral analyses.....	13
7.2	Chiral analyses.....	13
<b>8</b>	<b>Measurements of T<sub>m</sub> values</b> .....	<b>14</b>
<b>9</b>	<b>NMR spectra</b> .....	<b>15</b>
<b>10</b>	<b>ESI-MS spectra</b> .....	<b>24</b>
10.1	Deprotection of compounds 6b-D and 6b-L.....	24
10.2	UPLC-MS analysis for the coupling of submonomers 2-D and 3-D.....	24
10.3	UPLC-MS analysis for the synthesis of the 'Chiral box' part of PNA 2 .....	25
10.4	UPLC-MS chromatograms of pure PNAs .....	27
<b>11</b>	<b>GC-MS spectra</b> .....	<b>29</b>
11.1	Achiral analyses .....	29

11.2	Chiral analyses.....	31
<b>12</b>	<b>CD melting curves .....</b>	<b>33</b>
<b>13</b>	<b>References .....</b>	<b>35</b>

## 1 General

The DNA oligomers for evaluation of melting temperatures and the Fmoc-PNA(Bhoc)-OH monomers for PNA synthesis were purchased from Integrated DNA Technologies, Inc. (IDT) and LGC link, Inc., respectively, and used as such. Thymine-1-acetic acid (**CMT**), D-Arg(Pbf)-OH, L-Arg(Tos)-OH, D-Lys(Boc)-OH, L-Lys(Boc)-OH, and all the other commercially available reagents and solvents were bought from Sigma-Aldrich, Fluka, Merck, Carlo Erba, TCI Europe and used without further purification. Fmoc-aminoacetaldehyde (**1**) and Adenine(Boc)-1-acetic acid (**CMA**) were synthesized according to a procedure previously reported,<sup>1,2</sup> while Cytosine(Boc)-1-acetic acid (**CMC**) by adapting a protocol found in literature.<sup>2</sup> DMF was dried over 4 Å molecular sieves and purged with nitrogen to avoid the presence of dimethylamine. THF was dried by distillation over sodium and stored over 3 Å molecular sieves. TLCs were run on Merck 5554 silica 60 aluminum sheets. Flash column chromatography was performed on a CombiFlash NextGen 300 system by TeleDyne ISCO, using Silica or C18-silica PUREZZA-Daily Standard Cartridges (pore size 60Å, 50 µm particle size) for normal- or reversed phase separations, respectively.

NMR spectra were registered on Bruker Avance 400 and 300 MHz or on a JEOL ECZ600R 600MHz instruments.  $\delta$  values are expressed in ppm relatively to CDCl<sub>3</sub> (7.29 ppm for <sup>1</sup>H and 76.9 ppm for <sup>13</sup>C), DMSO-*d*<sup>6</sup> (2.50 ppm for <sup>1</sup>H and 39.5 ppm for <sup>13</sup>C) or CD<sub>3</sub>OD (3.34 ppm for <sup>1</sup>H and 49.86 ppm for <sup>13</sup>C). The following abbreviations are used to explain the multiplicities: s=singlet, d=doublet, t=triplet, q=quartet, m=multiplet, and br s.=broad signal.

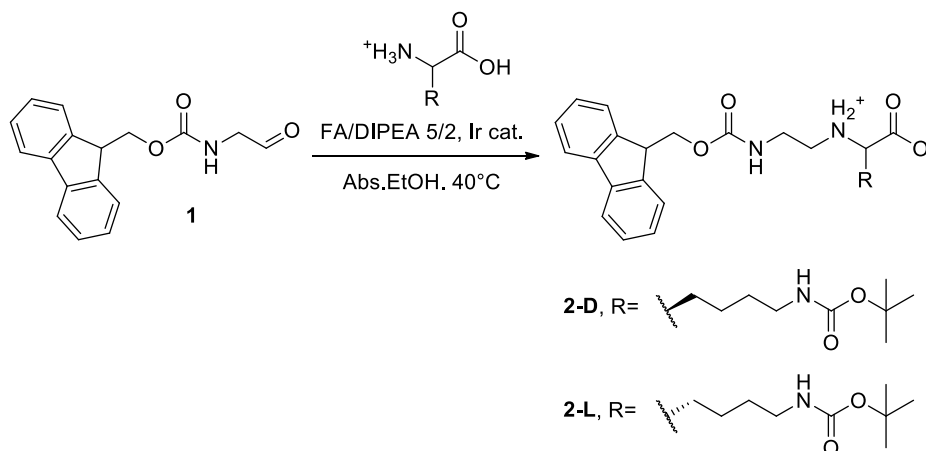
Reversed-phase HPLC purifications were performed on an Agilent 1200 Series LC system with a semi-preparative Jupiter column (C18, 5 mm, 300 Å, 10 x 250 mm) by Phenomenex. GC-MS analyses were performed using a 7820A GC System (Agilent Technologies, Palo Alto, CA) equipped with a MSD 5977B mass spectrometer (Agilent Technologies) and a 25 m × 0.25 mm, df 0.12 µm CP-ChirasilVal capillary column (Agilent Technologies) or a 30 m × 0.25 mm, df 0.25 µm DB-5MS capillary column (Agilent Technologies) for chiral and achiral separations, respectively.

UPLC-ESI-MS was carried out by using a Waters Acquity Ultra Performance LC with Waters Acquity SQ Detector and with ESI interface, and equipped with a Waters Acquity UPLC BEH 300 column (50x2.1 mm, 1.7 µm, C18) (Gradient for PNAs: 0.90 minutes in H<sub>2</sub>O + 0.2% FA, then linear gradient to 50% MeCN + 0.2% FA in 5.70 minutes at a flow rate of 0.25 mL/min. Gradient for small molecules: 0.90 minutes in H<sub>2</sub>O + 0.2% FA, then linear gradient to 100% MeCN + 0.2% FA in 5.70 minutes at a flow rate of 0.25 mL/min). HR-MS was performed on a Thermo LTQ ORBITRAP XL.

UV-ViS measurements were performed by a Thermo Scientific Evolution 260 BIO spectrophotometer, while circular dichroism analyses were carried out on a Jasco J715 spectropolarimeter.

## 2 Synthesis of the minimally protected submonomers

### 2.1 C(2)-Lys-based submonomers



**Scheme S 1:** Synthesis of the minimally-protected, C(2)-lysine-PNA-submonomers.

In a round-bottom flask **1** (285 mg, 1.01 mmol) and D-Lys(Boc)-OH or L- Lys(Boc)-OH (250 mg, 1.02 mmol) were suspended in absolute EtOH. The eutectic mixture of Formic Acid/*N,N*-diisopropylethylamine (DIPEA) 5/2 (535  $\mu$ L) and the iridium catalyst (31 mg, 0.05 mmol) were added, and a calcium chloride valve was used for protecting the reaction mixture from moisture. The temperature was raised to 40°C with an oil bath in order to solubilize the reagents, and the reaction was stirred overnight. The next day a white solid was precipitated, the reaction was cooled in an ice bath and the precipitate was filtered on Buchner and washed with cold ethanol to give pure **2-D** or **2-L**.

**N<sup>o</sup>-(Fmoc-2-aminoethyl)-N<sup>o</sup>-Boc-D-lysine (2-D):** white solid (313 mg, 0.61 mmol, 61% yield, *e.e.*: 93.4%).

<sup>1</sup>H NMR (600 MHz, DMSO-*d*<sup>6</sup>)  $\delta$  7.85 (d, *J* = 7.5 Hz, 2H), 7.64 (d, *J* = 7.4 Hz, 2H), 7.37 (t, *J* = 7.4 Hz, 2H), 7.30 (dd, *J* = 15.3, 7.7 Hz, 2H), 6.71 (s, 1H), 4.28 (d, *J* = 6.9 Hz, 2H), 4.18 (t, *J* = 6.7 Hz, 1H), 3.14 (d, *J* = 6.2 Hz, 2H), 3.06 (t, *J* = 5.9 Hz, 1H), 2.87 – 2.81 (m, 2H), 2.77 – 2.71 (m, 1H), 2.66 (dt, *J* = 12.4, 6.3 Hz, 1H), 1.54 (d, *J* = 5.7 Hz, 2H), 1.32 (s, 9H), 1.30 (d, *J* = 7.4 Hz, 2H), 1.27 – 1.20 (m, 2H). (**Figure S1**).

<sup>13</sup>C NMR (151 MHz, DMSO-*d*<sup>6</sup>)  $\delta$  157.0, 156.7, 156.1, 144.4, 141.3, 128.1, 127.6, 125.6, 120.6, 77.9, 65.9, 62.0, 47.3, 47.0, 40.3, 39.4, 31.5, 29.9, 28.7, 23.1. (**Figure S2**).

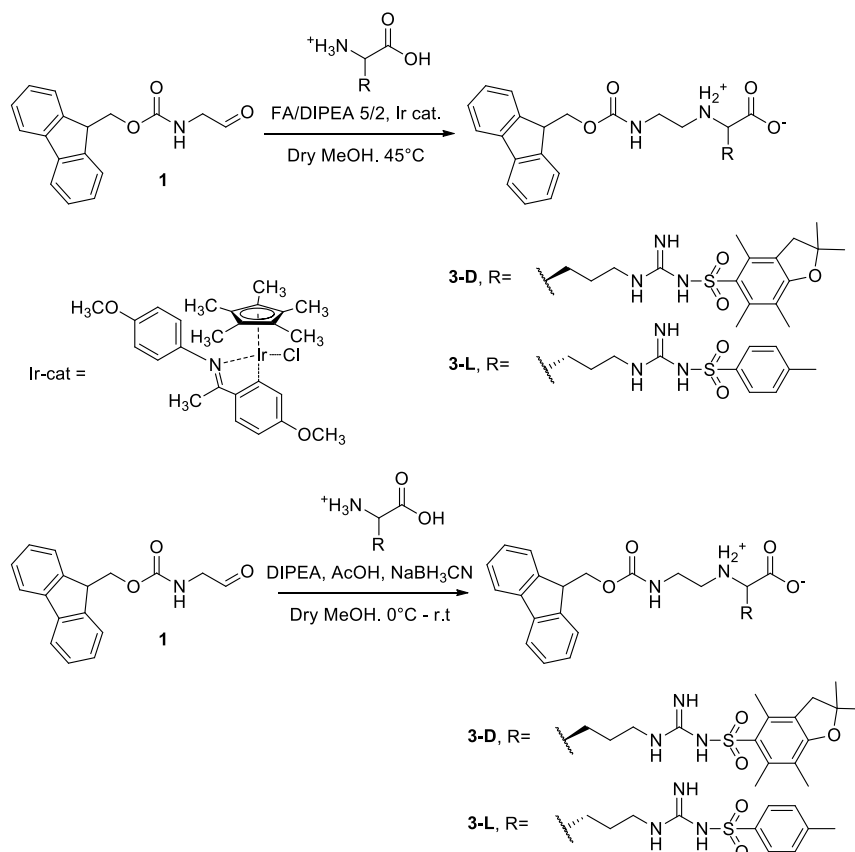
MS (ESI) *m/z*: 512.4 [M + H]<sup>+</sup>, 534.4 [M + Na]<sup>+</sup>, 1023.7 [2M + H]<sup>+</sup>, 1045.7 [2M+Na]<sup>+</sup>.

HRMS (ESI-LTQ-Orbitrap) *m/z*: [M – H]<sup>–</sup> Calcd for C<sub>28</sub>H<sub>36</sub>N<sub>3</sub>O<sub>6</sub> 510.2610; Found 510.2600.

Melting poin: 167 – 169° C

**N<sup>o</sup>-(Fmoc-2-aminoethyl)-N<sup>o</sup>-Boc-L-lysine (2-L):** White solid (296 mg, 0.58 mmol, 58% yield, *e.e.*: 94.4%). The target molecule shows the same spectroscopic properties of compound **2-D**.

## 2.2 C(2)-Arg-based submonomers



**Scheme S 2:** Synthesis of the minimally-protected, C(2)-arginine-PNA submonomers according to Method A (top) and Method B (bottom).

**Method A:** in a round-bottom flask, **1** (300 mg, 1.07 mmol) and D-Arg(Pbf)-OH (547 mg, 1.28 mmol) or L-Arg(Tos)-OH ·  $\text{CF}_3\text{COOH}$  (570 mg, 1.29 mmol) were dissolved in 15 mL of dry MeOH, and the iridium catalyst (33mg, 0.05 mmol) was added. The eutectic mixture of formic acid/DIPEA 5/2 (1.1 mL) was added and the reaction was let stirring at 45° C for 3 hours (oil bath). Then, the solvent was evaporated and the residue was dissolved in MeOH/EtOAc 1/9 (30 mL) and washed with water (4 x 20 mL). The organic phase was treated with anhydrous sodium sulfate and dried under reduced pressure. The crude (around 590 mg for both enantiomers) was divided in four aliquots and purified by reversed-phase flash chromatography ( $\text{H}_2\text{O}/\text{ACN}$  9/1-  $\text{H}_2\text{O}/\text{ACN}$  4/6) on a column containing 12 g of C-18 silica, to give **3-D** or **3-L**. Attempts to purify larger amounts of crude in a single step gave a lower yield due to overloading of the column, despite for this size of cartridges the manufacturer protocols suggested 460 mg of material as maximum loading limit.

**Method B:** in a round-bottom flask, **1** (300 mg, 1.07 mmol), D-Arg(Pbf)-OH (547 mg, 1.28 mmol) or L-Arg(Tos)-OH ·  $\text{CF}_3\text{COOH}$  (570 mg, 1.29 mmol) and DIPEA (0.22 mL, 1.28 mmol) were dissolved in 15 mL of dry MeOH and the resulting mixture was stirred for 30 min. Then, glacial acetic acid (66  $\mu\text{L}$ , 1.07 mmol) and  $\text{NaBH}_3\text{CN}$  (62 mg, 1.07 mmol) were added at 0°C (ice bath) and the reaction was let proceeding while warming up to room temperature. After 2 hours the reaction was quenched with water (15 mL) and the mixture was extracted with EtOAc (4 x 20 mL). The combined organic layers were dried with anhydrous sodium sulfate and the solvent was removed over reduced pressure. The crude (around 600 mg for both enantiomers) was divided in four aliquots and purified by reversed-phase flash chromatography ( $\text{H}_2\text{O}/\text{ACN}$  9/1-  $\text{H}_2\text{O}/\text{ACN}$  4/6)

on a column containing 12 g of C-18 silica, to give **3-D** or **3-L**. Attempts to purify larger amounts of crude in a single step gave a lower yield due to overloading of the column, despite for this size of cartridges the manufacturer protocols suggested 460 mg of material as maximum loading limit.

**N<sup>α</sup>-(Fmoc-2-aminoethyl)-N<sup>γ</sup>-Pbf-D-arginine (3-D)**: White solid (Method A: 360 mg, 0.52 mmol, 49% yield; Method B: 305 mg, 0.44 mmol, 41% yield).

<sup>1</sup>H NMR (400 MHz, DMSO-*d*<sup>6</sup>) δ 77.89 (d, J = 7.3 Hz, 2H), 7.68 (d, J = 7.1 Hz, 2H), 7.45 - 7.37 (m, 3H), 7.33 (t, J = 7.1 Hz, 2H), 7.22 (s, 1H), 6.65 (s, 1H), 4.33 (d, J = 6.5 Hz, 2H), 4.23 (t, J = 6.3 Hz, 1H), 3.20 (s, 3H), 3.04 (s, 2H), 2.95 (s, 2H), 2.85 - 2.63 (m, 2H), 2.50 (s, 3H) 2.43 (s, 3H), 2.00 (s, 3H), 1.59 (s, 2H), 1.48 (s, 2H), 1.40 (s, 6H). (**Figure S3**).

<sup>13</sup>C NMR (101 MHz, DMSO-*d*<sup>6</sup>) δ 171.9, 157.9, 156.7, 156.6, 144.3, 141.2, 137.7, 134.8, 131.9, 128.1, 127.5, 125.6, 124.8, 120.6, 116.7, 86.7, 65.9, 61.6, 56.3, 47.2, 46.5, 42.9, 38.6, 31.2, 28.8, 27.8, 19.4, 18.1, 12.8. (**Figure S4**).

MS (ESI) *m/z*: 692.42 [M + H]<sup>+</sup>, 714.40 [M + Na]<sup>+</sup>.

HRMS (ESI-LTQ-Orbitrap) *m/z*: [M + H]<sup>+</sup> Calcd for C<sub>36</sub>H<sub>46</sub>N<sub>5</sub>O<sub>7</sub>S 692.3113; Found 692.3113.

Melting point: 161 – 163° C

**N<sup>α</sup>-(Fmoc-2-aminoethyl)-N<sup>γ</sup>-Tos-L-arginine (3-L)**: White solid (Method A: 300 mg, 0.51 mmol, 47% yield; Method B: 344 mg, 0.58 mmol, 54% yield).

<sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD) δ 7.80 (d, J = 7.5 Hz, 2H), 7.75 (d, J = 8.0 Hz, 2H), 7.65 (d, J = 7.8 Hz, 2H), 7.39 (t, J = 7.4 Hz, 2H) 7.41 - 7.27 (m, 4H), 4.61 (br. s, 1H), 4.41 (d, J = 6.7 Hz, 2H), 4.21 (t, J = 6.7 Hz, 1H), 3.59 – 3.56 (m, 1H), 3.44 (t, J = 6.5 Hz, 2H), 3.26 – 3.20 (m, 2H), 3.19 (t, J = 6.5 Hz, 2H), 2.37 (s, 3H), 1.85 (br s., 2H), 1.64 (br s., 2H). (**Figure S5**).

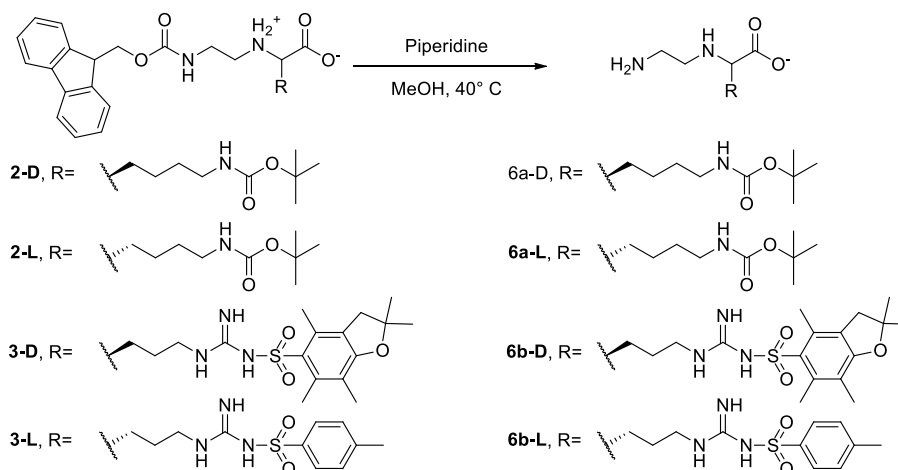
<sup>13</sup>C NMR (151 MHz, CD<sub>3</sub>OD) δ 171.9, 157.4, 144.0, 143.9, 142.3, 141.3, 140.8, 129.0, 127.5, 127.4, 126.9, 126.8, 125.8, 124.8, 119.7, 119.6, 66.8, 66.5, 60.6, 48.6, 42.9, 37.2, 27.1, 20.1. (**Figure S6**).

MS (ESI) *m/z*: 594.42 [M + H]<sup>+</sup>; 616.39 [M + Na]<sup>+</sup>, 1187.65 [2M + H]<sup>+</sup>, 1209.70 [2M + Na]<sup>+</sup>.

HRMS (ESI-LTQ-Orbitrap) *m/z*: [M + H]<sup>+</sup> Calcd for C<sub>30</sub>H<sub>36</sub>N<sub>5</sub>O<sub>6</sub>S 594.2381; Found 594.2380.

Melting point: 143 – 146° C

### 3 Fmoc deprotection of the minimally-protected submonomers



**Scheme S 3:** Removal of the Fmoc protecting group from the minimally-protected submonomers.

**General procedure for Fmoc-deprotection of the minimally protected submonomers:** Removal of the N-terminal Fmoc group was optimized on compound **2-D** as follows: the C(5)-modified backbone (511 mg, 1 mmol) was dissolved in methanol (50 mL) in the presence of piperidine (1.5 mL, 15 mmol). The reaction was stirred for 1 hour at 40° C in an oil bath, and the solvent was removed under reduced pressure. The residue was sonicated in EtOAc (25 mL, 2 x 15 min) and in a EtOAc/EtOH 9/1 mixture (25 mL, 1 x 15 min) to give pure **6a-D** (checked by TLC: *t*-BuOH/EtOH/H<sub>2</sub>O/AcOH 6/2/1/1, ninhydrin stain). The other Lysine- (**2-L**) or Arginine-modified (**3-D**, **3-L**) submonomers were deprotected according to the same procedure, but in a smaller scale (0.06 mmol), since the reaction was performed as a preliminary step for derivatization of the resulting products (see section 6).

**2-aminoethyl-N<sup>o</sup>-Boc-D-lysine (6a-D):** White solid (230 mg, 0.80 mmol, 80 % yield)

<sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD), δ: 3.16 (t, J = 6.0 Hz, 1H), 3.05 (t, J = 6.4 Hz, 2H), 3.01 (t, J = 5.6 Hz, 2H), 2.92 (t, J = 6.4 Hz, 2H), 1.70 (m, 2H), 1.45 (m, 13H). (**Figure S7**).

<sup>13</sup>C NMR (151 MHz, CD<sub>3</sub>OD), δ: 178.9, 157.2, 78.4, 63.5, 45.3, 39.8, 38.5, 38.5, 32.5, 29.5, 27.4, 22.8. (**Figure S8**).

MS (ESI) m/z: 290.38 [M + H]<sup>+</sup>, 312.36 [M + Na]<sup>+</sup>, 579.57 [2M + H]<sup>+</sup>, 601.48 [2M + Na]<sup>+</sup>.

HRMS (ESI-LTQ-Orbitrap) m/z: [M + H]<sup>+</sup> Calcd for C<sub>13</sub>H<sub>28</sub>N<sub>3</sub>O<sub>4</sub> 290.2074; Found 290.2074.

Melting point: decomposes at 178° C

**2-aminoethyl-N<sup>o</sup>-Boc-L-lysine (6a-L):** White solid, (15 mg, 0.05 mmol, 80 % yield). This product shows the same spectroscopic properties of compound **6a-D**.

**2-aminoethyl-N<sup>γ</sup>-Pbf-D-Arginine (6b-D):** White solid (20 mg, 0.04 mmol 70% yield).

<sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD) δ 3.18 (br s., 2H), 3.08 (t, J = 5.4 Hz, 1H), 3.02 (s, 2H), 2.96 (t, J = 5.8 Hz, 2H) 2.84 (t, J = 5.8 Hz, 2H), 2.59 (s, 3H), 2.52 (s, 3H), 2.06 (s, 3H), 1.57 (br s., 4H), 1.47 (s, 6H). (**Figure S9**).

<sup>13</sup>C NMR (151 MHz, CD<sub>3</sub>OD) δ: 179.9, 158.4, 156.9, 138.0, 132.9, 132.1, 124.6, 117.0, 86.2, 63.2, 45.2, 42.5, 38.8, 30.3, 27.3, 18.1, 16.9, 11.1. (**Figure S10**).

MS (ESI) m/z: 470.35 [M+H]<sup>+</sup>, 492.46 [M + Na]<sup>+</sup>, 939.70 [2M + H]<sup>+</sup>, 961.68 [2M + Na]<sup>+</sup>.

HRMS (ESI-LTQ-Orbitrap) m/z: [M + H]<sup>+</sup> Calcd for C<sub>21</sub>H<sub>36</sub>N<sub>5</sub>O<sub>5</sub>S 470.2432; Found 470.2435.

Melting point: 175 – 177° C.

**2-aminoethyl-N<sup>γ</sup>-Tos-L-Arginine (6b-L):** Light brown solid (16 mg, 0.04 mmol, 74% yield)

<sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD) δ 7.76 (d, J = 7.8 Hz, 2H), 7.32 (d, J = 7.8 Hz, 2H), 3.19 (br s., 2H), 3.08 (t, J = 5.7 Hz, 1H), 2.99 (t, J = 5.6 Hz, 2H), 2.87 (t, J = 5.6 Hz, 2H), 2.40 (s, 3H), 1.65 (br s., 4H). (**Figure S11**).

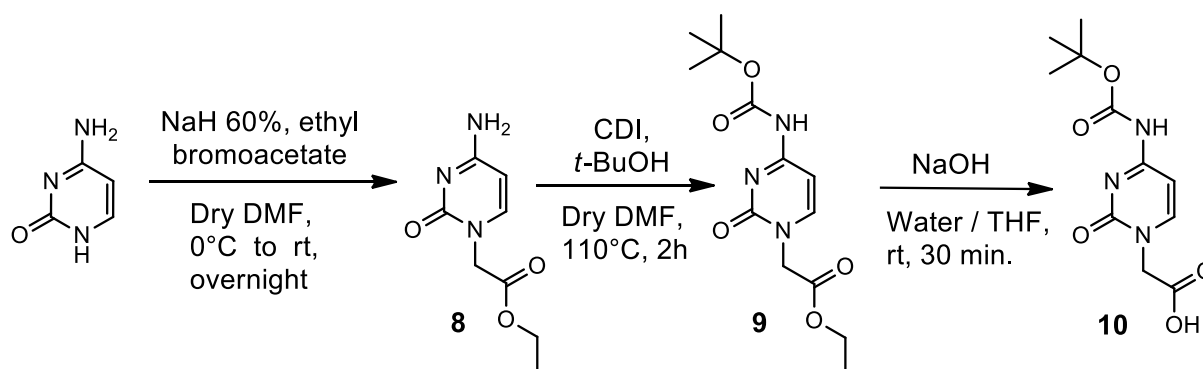
<sup>13</sup>C NMR (151 MHz, CD<sub>3</sub>OD) δ: 179.9, 157.3, 142.2, 140.9, 129.0, 125.8, 63.3, 45.3, 40.4, 38.9, 30.2, 25.8, 20.1. (**Figure S12**).

MS (ESI) m/z: 371.82 [M + H]<sup>+</sup>, 743.50 [2M+H]<sup>+</sup>.

HRMS (ESI-LTQ-Orbitrap) m/z: [M + H]<sup>+</sup> Calcd for C<sub>15</sub>H<sub>26</sub>N<sub>5</sub>O<sub>4</sub>S 372.1700; Found 372.1701.

Melting point 128 – 129° C

## 4 Synthesis of cytosine(Boc)-1-acetic acid



Scheme S 4: Synthesis of cytosine(Boc)-1-acetic acid

**Ethyl cytosine-1-acetate (8):** in a round-bottom flask, cytosine (1.96 g, 17.72 mmol) was suspended in 30 mL of dry DMF, under nitrogen atmosphere and under stirring. NaH 60% was added in three portions (3 x 0.287 g, 3 x 7.09 mmol) and the solution became pale white. The mixture was let stirring for 2.5 h at room temperature, and then ethyl 2-bromoacetate was added dropwise (2.3 mL, 19.85 mmol) at 0°C (ice bath). The mixture turned orange and the reaction was let stirring overnight. The reaction was quenched with AcOH (12.2 mL, 0.21 mol, 12 eq) under vigorous magnetic stirring. The solvent was evaporated under high vacuum and the crude mixture was purified by two consecutive chromatographic flash column (AcOEt/MeOH 7:3). The resulting solid was the pure product (1.22 g, 6.19 mmol, 35% yield).

<sup>1</sup>H NMR (400 MHz, DMSO-*d*<sup>6</sup>) δ: 7.55 (d, J = 7.2 Hz, 1H), 7.12 (d, J = 42.0 Hz, 2H), 5.69 (d, J = 7.2 Hz, 1H), 4.43 (s, 2H), 4.12 (q, J = 7.1 Hz, 2H), 1.20 (t, J = 7.1 Hz, 3H). (**Figure S13**).

<sup>13</sup>C NMR (101 MHz, DMSO-*d*<sup>6</sup>) δ: 169.2, 166.8, 156.2, 146.8, 94.0, 61.2, 50.4, 14.5. (**Figure S14**).

MS (ESI) *m/z*: 198.10 [M + H]<sup>+</sup>, 395.19 [2M + H]<sup>+</sup>.

HRMS (ESI-LTQ-Orbitrap) *m/z*: [M + H]<sup>+</sup> Calcd for C<sub>8</sub>H<sub>12</sub>N<sub>3</sub>O<sub>3</sub> 198.0873; Found 198.0868;

[M + Na]<sup>+</sup>. Calcd for C<sub>8</sub>H<sub>11</sub>O<sub>3</sub>N<sub>3</sub>Na 220.0693; Found 220.0687.

**Melting point:** 254 – 255° C.

**Ethyl (4-N-(t-Butoxycarbonyl)cytosine)-1-acetate (9):** compound **8** (1.22 g, 6.19 mmol) was dissolved in 26 mL of dry DMF and carbonyldiimidazole (CDI, 1.07 g, 6.59 mmol) was added. The reaction mixture, which became brownish, was let stirring for 2.5 hours at room temperature. Then, anhydrous *t*-BuOH (2 mL, 20.59 mmol, dried on 3 Å molecular sieves) was added and the temperature was raised at 80°C for 1.5 h with an oil bath. After that, the reaction was left stirring overnight at room temperature and was then quenched by addition of MeOH (334 μL, 8.23 mmol) under vigorous magnetic stirring. The solvent was evaporated under high vacuum and the resulting brownish oil was dissolved in 20 mL of EtOAc and washed with a saturated KHSO<sub>4</sub> aqueous solution (4 x 20 mL). The organic phase was dried over anhydrous Na<sub>2</sub>SO<sub>4</sub> and evaporated under reduced pressure. The crude mixture was purified by flash chromatography (EtOAc/ MeOH 8:2) to give the pure product (1.10 g, 3.71 mmol, 60% yield).



<sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD) δ 7.93 (d, J = 7.3 Hz, 1H), 7.30 (d, J = 7.3 Hz, 1H), 4.66 (s, 2H), 4.25 (q, J = 7.1 Hz, 2H), 1.55 (s, 9H), 1.30 (t, J = 7.1 Hz, 3H). (**Figure S15**).

<sup>13</sup>C NMR (101 MHz, CD<sub>3</sub>OD) δ 167.9, 164.4, 157.2, 152.0, 149.5, 95.4, 81.8, 61.5, 50.8, 26.9, 13.0. (**Figure S16**).

UPLC-MS (ESI): r.t. = 4.19 min, m/z: 298.2 [M + H]<sup>+</sup>, 242.17 [M - *t*-Bu]<sup>+</sup>, 595.5 [2M + H]<sup>+</sup>.

HRMS (ESI-LTQ-Orbitrap) m/z: [M + Na]<sup>+</sup> Calcd for C<sub>13</sub>H<sub>19</sub>N<sub>3</sub>O<sub>5</sub>Na 320.1217; Found 320.1212.

Melting point: > 300° C.

**4-N-(*t*-Butoxycarbonyl)cytosine)-1-acetic acid (10):** compound **9** (0.42 g, 1.41 mmol) was put in a round-bottom flask and dissolved in a mixture of H<sub>2</sub>O (4 mL), MeOH (12 mL) and ACN (12 mL). LiOH·H<sub>2</sub>O (0.453 g, 10.80 mmol) was dissolved by sonication in 4.3 mL of water into a vial and was added to the reaction mixture. A precipitate was immediately formed and the solution became pale yellow. The reaction was quenched after 10 min by dropwise addition of aqueous 1M HCl (about 10 mL) until the pH dropped down to 3-3.5. After that, the reaction flask was stored in fridge at 5°C for 2.5 hours and the precipitate was filtered and washed with cold water to give the title compound as a white solid (0.31 g, 1.15 mmol, 82% yield).

<sup>1</sup>H NMR (300 MHz, DMSO-*d*<sup>6</sup>) δ 10.33 (s, 1H), 7.99 (d, J = 7.3 Hz, 1H), 6.99 (d, J = 7.3 Hz, 1H), 4.51 (s, 2H), 3.34 (s, 1H), 1.46 (s, 9H). (**Figure S17**).

<sup>13</sup>C NMR (101 MHz, DMSO-*d*<sup>6</sup>) δ 169.9, 163.9, 155.6, 152.6, 150.5, 94.5, 81.5, 50.9, 28.3. (**Figure S18**).

UPLC-MS (ESI): r.t. = 3.53 min, m/z: 270.17 [M + H]<sup>+</sup>, 214.11 [M - *t*-Bu]<sup>+</sup>, 539.32 [2M + H]<sup>+</sup>.

HRMS (ESI-LTQ-Orbitrap) m/z: [M - H]<sup>-</sup> Calcd for C<sub>11</sub>H<sub>14</sub>N<sub>3</sub>O<sub>5</sub> 268.0939; Found 268.0938.

Melting point: > 300° C

## 5 PNA synthesis

All the PNAs were synthesized on Rink Amide-ChemMatrix resin loaded with Fmoc-Gly-OH (0.17 mmol/g) in 5 μmol scale, by using the following solutions:

- Deprotection: 20% piperidine in DMF
- Capping: acetic anhydride/DIPEA/dry DMF 5/6/89
- DIPEA wash: 5% DIPEA in DMF
- Cleavage: TFA/*m*-cresol 9/1

### 5.1 Unmodified PNAs

PNA without modifications (**PNA 1** and **4**) were obtained by automatic synthesis on a Biotage Syro I system according to the standard manufacturer protocols, consisting in the following steps: a) Deprotection 2 x 8 min; b) DCM wash; c) DMF Wash; d) Coupling 2 x 1 h (3 eq of monomer, 3eq of HBTU, 6 eq of DIPEA in dry DMF); e) DMF Wash; f) Capping 2 x 2 min; g) DMF Wash. A periodic control of the synthesis progress was performed by UPLC-MS analysis.

For both PNAs, a final treatment of the dried resins with the cleavage solution (2 x 1 h) resulted in the release of the oligomers and in the removal of the protecting groups from the nucleobases and the side-chains on C(2)-position. The crude samples were precipitated with 50 volumes of Et<sub>2</sub>O at -20°C and purified by reversed-phase HPLC (Gradient: 5 minutes in H<sub>2</sub>O + 0.1% TFA, then linear gradient from H<sub>2</sub>O + 0.1% TFA to 50% MeCN + 0.1 % TFA in 30 minutes at a flow rate of 4.0

ml/min). The purity of the resulting PNAs was checked by UPLC-MS analysis, while their concentration was calculated from the UV absorbance of diluted solutions at  $\lambda = 260$  nm, assuming an additive contribution of all nucleobases to the total  $\epsilon$  (260 nm) of the oligomers. Accordingly, the molar extinction coefficients were calculated by considering the following values for each nucleobase: T = 8600 M<sup>-1</sup>cm<sup>-1</sup>, C = 6600 M<sup>-1</sup>cm<sup>-1</sup>, A = 13700 M<sup>-1</sup>cm<sup>-1</sup>, G = 11700 M<sup>-1</sup>cm<sup>-1</sup>.

## 5.2 'Chiral box' PNAs

Chiral-box PNAs (**PNA 2**, **3**, **5** and **6**) were synthesized manually by using standard Fmoc/Bhoc synthesis for the attachment of the first unmodified monomers.<sup>3</sup> A sub-monomeric approach was instead used for the construction of the chiral monomers: a) Fmoc deprotection 2 x 8 min; b) DCM wash; c) DMF wash; d) Kaiser test (1 min, positive);<sup>4</sup> e) coupling of the minimally protected submonomer (**2-D** or **3-D**) 2 x 30min (activation for 2 min; Activation solution: 5 eq of submonomer, 4.9 eq of PyBOP or HBTU as activator, 10 eq of DIPEA in dry DMF); f) DMF wash; g) coupling of the carboxymethylnucleobase (**CMA**, **CMT** or **CMC**) 2 x 2 h (activation for 15min; Activation solution: 7 eq of carboxymethylnucleobase, 7 eq of DIC, 7 eq of DhBtOH and 1 eq of DIPEA in dry DMF); h) DMF wash; i) DCM wash. These steps were repeated for all the three C2-modified monomers and then the PNA sequences were completed by using standard Fmoc/Bhoc synthetic strategies.

The occurrence of significant self-coupling processes affecting the attachment of submonomers **2-D** and **3-D** was assessed by UPLC-MS analysis during the first cycle of chain elongation (step e) for **PNA 2** and **3**, respectively. The crudes resulting from the insertion of the C(2)-modified backbones contained transient tris(pyrrolidine)phosphoroamidate adducts, which derive from reaction with the PyBop activating agent and are normally removed by mild basic treatments (i. e. the subsequent Fmoc deprotection steps, data not shown). Hence, the samples used for monitoring the attachment of the compounds **2-D** and **3-D** (0.15  $\mu$ mol) were reacted 1 hour with MeNH<sub>2</sub> (1.5  $\mu$ mol) prior to UPLC-MS analysis for an easier interpretation of the chromatograms. The traces were in fact composed by well-separated peaks corresponding to deprotected PNAs and small molecules generated by the cleavage of their terminal Fmoc groups or the aforementioned tris(pyrrolidine)phosphoroamidate adducts (**Figures S21** and **S22**). Significant over-coupling reactions involving the unprotected N(3)-position of compounds **2-D** and **3-D** were excluded by extraction of the chromatograms referred to the corresponding byproducts, thus validating the use of these minimally-protected submonomers.

A general evaluation of the submonomeric procedure was instead carried out for the Lys-based **PNA2**, after the construction of its complete 'Chiral box' stretch (step g, third synthetic cycle). A crude portion of the oligomer (0.15  $\mu$ mol) was submitted to UPLC-MS and the extracted ion chromatograms corresponding to uncompleted PNAs were generated, evidencing minor amounts of compounds truncated on their N-term or lacking a nucleobase (**Figures S23** and **S24**, respectively). At the end of their synthesis, each 'Chiral box' PNA was removed from the resin, purified, characterized and quantified according to the general procedures reported above for the unmodified derivatives.

**PNA1**: H-OO-CTA CGC CAT CAG CT-Gly-NH<sub>2</sub>,  $\epsilon$  (260 nm): 129900 M<sup>-1</sup>cm<sup>-1</sup>, 11.1 % yield. UPLC-MS (ESI): r.t. = 2.69 min, m/z: [M] Calcd 4079.1; Found 1020.7 [M + 4H]<sup>4+</sup>, 816.7 [M + 5H]<sup>5+</sup>, 680.8 [M + 6H]<sup>6+</sup>, 583.7 [M + 7H]<sup>7+</sup>, 510.8 [M + 8H]<sup>8+</sup>, 454.1 [M + 9H]<sup>9+</sup> (**Figure S25**).

**PNA2:** H-CTA CGC CA<sub>D-Lys</sub>T<sub>D-Lys</sub> C<sub>D-Lys</sub>AG CT-Gly-NH<sub>2</sub>,  $\epsilon$  (260 nm): 129900 cm<sup>-1</sup>M<sup>-1</sup>, 5.0 % yield.

UPLC-MS (ESI): r.t. = 2.45 min, m/z: [M] Calcd 4002.0; Found 1335.1 [M + 3H]<sup>3+</sup>, 1001.4 [M + 4H]<sup>4+</sup>, 801.3 [M + 5H]<sup>5+</sup>, 668.1 [M + 6H]<sup>6+</sup>, 572.6 [M + 7H]<sup>7+</sup>, 501.2 [M + 8H]<sup>8+</sup>. (**Figure S26**).

**PNA3:** H-OO-CTA CGC CCA<sub>D-Arg</sub>T<sub>D-Arg</sub> C<sub>D-Arg</sub>AG CT-Gly-NH<sub>2</sub>,  $\epsilon$  (260 nm): 129900 cm<sup>-1</sup> M<sup>-1</sup>, 4.0 % yield.

UPLC-MS (ESI): r.t. = 2.57 min, m/z: [M] Calcd 4376.0 [M]; Found 876.3 [M + 5H]<sup>5+</sup>, 730.3 [M + 6H]<sup>6+</sup>, 626.1 [M + 7H]<sup>7+</sup>, 548.0 [M + 8H]<sup>8+</sup>. (**Figure S27**).

**PNA4:** H-O-CGC CAT CAG CT-Gly-NH<sub>2</sub>,  $\epsilon$  (260 nm): 101000 cm<sup>-1</sup> M<sup>-1</sup>, 2.5 % yield.

UPLC-MS (ESI): r.t. = 2.67 min, m/z: [M] Calcd 3141.1; Found 1048.0 [M + 3H]<sup>3+</sup>, 786.3 [M + 4H]<sup>4+</sup>, 629.2 [M + 5H]<sup>5+</sup>, 524.5 [M + 6H]<sup>6+</sup>, 449.6 [M + 7H]<sup>7+</sup>. (**Figure S28**).

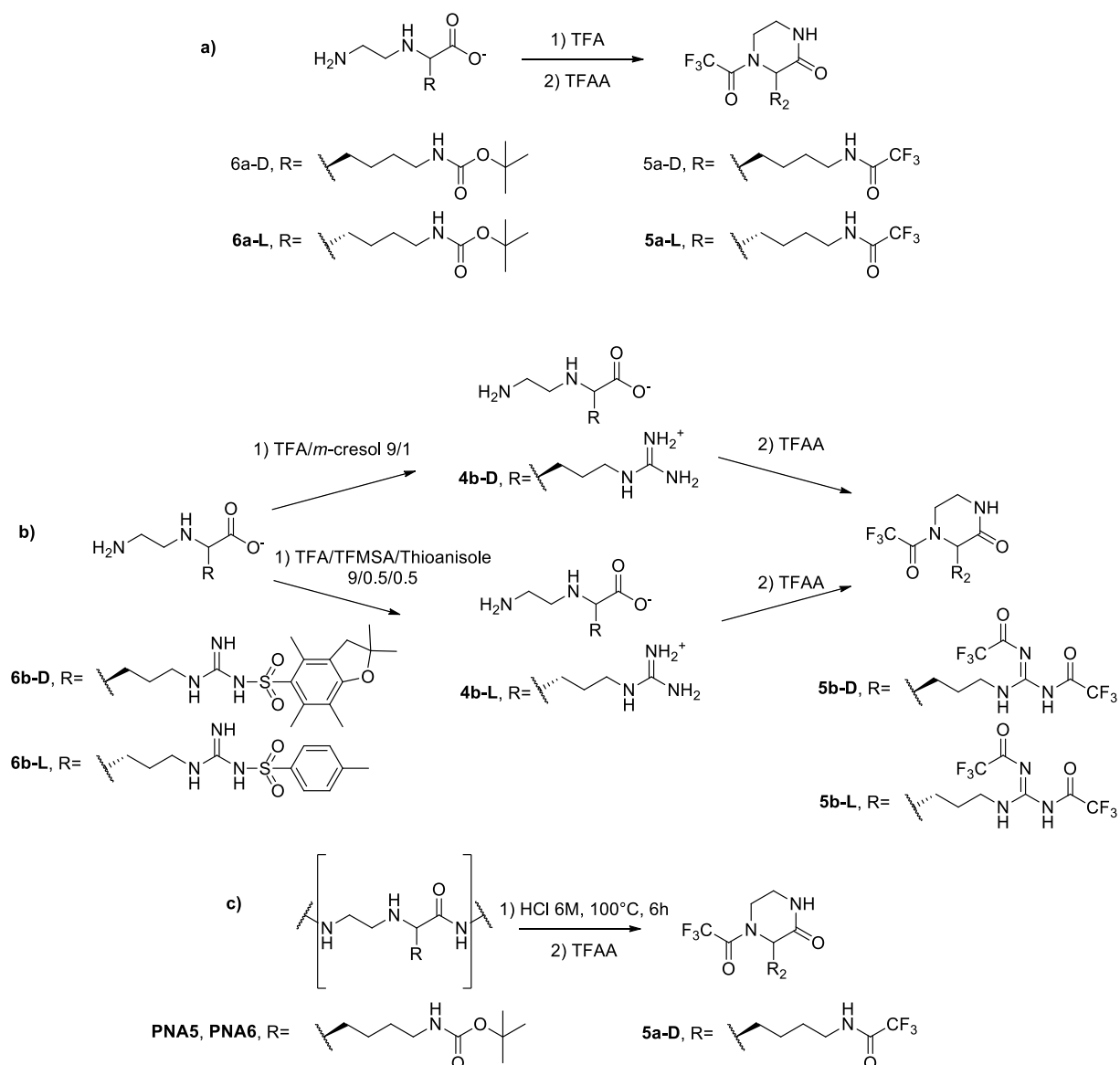
**PNA5:** H-O-CGC CA<sub>D-Lys</sub>T<sub>D-Lys</sub> C<sub>D-Lys</sub> AG CT-Gly-NH<sub>2</sub>,  $\epsilon$  (260 nm): 101000 cm<sup>-1</sup> M<sup>-1</sup>, 2.5 % yield, *e.e.*: 94.6%.

UPLC-MS (ESI): r.t. = 2.55 min, m/z: [M] Calcd 3355.0; Found 1118.9 [M + 3H]<sup>3+</sup>, 839.7 [M + 4H]<sup>4+</sup>, 671.9 [M + 5H]<sup>5+</sup>, 560.1 [M + 6H]<sup>6+</sup>, 480.2 [M + 7H]<sup>7+</sup>. (**Figure S29**).

**PNA6:** Biotin-OO-CGC CA<sub>D-Lys</sub>T<sub>D-Lys</sub> C<sub>D-Lys</sub>AG CT-Gly-NH<sub>2</sub>,  $\epsilon$  (260 nm): 101000 cm<sup>-1</sup> M<sup>-1</sup>, 5.8 % yield. *e.e.*: 90.0%.

UPLC-MS (ESI): r.t. = 2.86 min, m/z: [M] Calcd 3725.5; Found 1242.6 [M + 3H]<sup>3+</sup>, 932.2 [M + 4H]<sup>4+</sup>, 746.1 [M + 5H]<sup>5+</sup>, 621.9 [M + 6H]<sup>6+</sup>, 533.1 [M + 7H]<sup>7+</sup>. (**Figure S30**).

## 6 Derivatization procedures for enantiomeric analysis.



**Scheme S 5:** Derivatization of **a)** Lys- or **b)** Arg-based minimally-protected submonomers and **c)** Chiral-box PNAs for GC analysis.

**C(2)-Lysine-submonomers:** Derivatization of Lys-based submonomers was performed according to a known procedure.<sup>5</sup> Briefly: in a screw-cap tube, compounds **6a-D** or **6a-L** (10  $\mu$ mol) were dissolved in TFA (1 mL) and let stirring for 30 min, then trifluoroacetic anhydride (TFAA) was added (2 mL) and the temperature was raised to 40°C with an oil bath. The progression of the two steps (deprotection and derivatization) was checked by TLC (*t*-BuOH/EtOH/H<sub>2</sub>O/AcOH 6/2/1/1, ninhydrin stain) and the reaction was quenched after 3 hours by removal of the reactants at 40°C (oil bath) under nitrogen flux. The residues were taken in 0.5 mL of EtOAc and the optical purity of the corresponding piperazine-2-one **5a-D** and **5a-L** was checked by chiral GC analysis (see section 7.2).

**Arg-submonomers:** Derivatization of Arg-based submonomers was performed by adapting a known procedure.<sup>5</sup> In a screw-cap tube, compounds **6b-D** or **6b-L** (10  $\mu$ mol) were dissolved in

TFA/m-cresol 9/1 (1 mL) or TFA/TFMSA/thioanisole 9/0.5/0.5 (1 mL), respectively. The cleavage of the side-chain protecting group was completed after 1h for Pbf and 3h for Tos to give intermediates **4-D** and **4-L**, whose formation was monitored by TLC (*t*-BuOH/EtOH/H<sub>2</sub>O/AcOH 6/2/1/1, ninhydrin stain) and ESI-MS analysis (*m/z*: 218.18 for **4b-D** and 218.25 for **4b-L**, [M+H]<sup>+</sup>, **Figure S19** and **S20**, respectively). Then, the reaction mixtures were dried under nitrogen flux at 50° C (oil bath) and the residues were triturated by sonication in AcOEt (3 mL, 2 x 15 min) and in Et<sub>2</sub>O (3 mL, 1 x 15 min). The resulting residues were taken in trifluoroacetic anhydride (TFAA, 2 mL) and reacted at 40°C for 3 hours in an oil bath. Then, unreacted TFAA was removed under nitrogen flux and the resulting piperazine-2-one **5b-D** and **5b-L** were dissolved in 0.5 mL of EtOAc for GC analysis (see section 7.1).

**Chiral-box PNAs:** Derivatization of C(2)-modified PNAs was performed according to a known procedure.<sup>5</sup> Briefly: In a screw-cap tube, an aqueous solution of PNA **3**, **5** or **6** (30 nmol) was mixed with 37% HCl (0.3 mL, 6M final concentration) and digested overnight at 100°C with an oil bath. The solvent was removed at 50° C (oil bath) under nitrogen flux and then the residues were taken in trifluoroacetic anhydride (TFAA, 2 mL) and reacted for 3h at 40°C (oil bath). After that, unreacted TFAA was removed under nitrogen flux and the corresponding mixtures of modified- and unmodified piperazine-2-ones were dissolved in 0.3 mL of EtOAc for chiral GC analysis (see section 7.2).

## 7 CG-MS analysis

### 7.1 Achiral analyses

Solutions of the derivatized Arg-based submonomers **5b-D** and **5b-L** and “Chiral-box” PNA **3** (2 µL) were submitted to achiral GC analysis under the Total Ion Current (TIC) mode for a preliminary determination of their fragmentation patterns. Then, the experiments were repeated by Selected Ion Monitoring (SIM) for simulating their enantiomeric analysis with maximum sensitivity. The tested samples were: a) 1/1 mixture of **5b-D** and **5b-L**; b) **5b-D**; c) **5b-L**; d) PNA **3** hydrolyzate. The experimental conditions were selected by adapting a known procedure:<sup>5</sup>

- Selected ions: *m/z* 149, 167, 279
- Modality: Splitless
- Injector temperature: 230°C
- Solvent delay: 4 min
- Column temperature: 190°C for 75 min

**5b-D/5b-L** 1/1, r.t.: 65.52 min. (**Figure S31**)

**5b-D**, r.t.: 65.49 min. (**Figure S32**)

**5b-L**, r.t.: 65.66 min. (**Figure S33**)

PNA **3**, r.t.: 65.51 min. (**Figure S34**)

### 7.2 Chiral analyses

Lys-submonomers **5a-D** and **5a-L** and Chiral-box PNA **5** and **6** were directly submitted to chiral GC investigations according to what reported in literature,<sup>5</sup> and Arg-derivatives were analyzed in

the same way. The experiments were carried out under SIM mode on 2  $\mu$ L aliquots of the following solutions: a) 1/1 mixture of **5a-D** and **5a-L**; b) **5a-D**; c) **5a-L**; d) **PNA 5** hydrolysate; e) **PNA 6** hydrolysate; f) **5b-D**; g) **5b-L**. The detailed experimental conditions are listed below:

- Selected ions: m/z: 140, 167, 196 for **5a-D**, **5a-L**, **PNA 5** and **PNA 6**,  
m/z: 149, 167, 279 for **5b-D** and **5b-L**
- Modality: splitless
- Injector temperature: 230°C
- Solvent delay: 4 min
- Column temperature: 190°C for 75 min

**5a-D/5a-L** 1/1, r.t.: 39.09 min, area %:  $52.0 \pm 0.8\%$ ; 42.64 min, area %:  $48.0 \pm 0.6\%$  (**Figure S35**)

**5a-D**, r.t.: 39.54 min, area %:  $96.7 \pm 0.5\%$ ; 42.58 min, area %:  $3.3 \pm 0.5\%$ ; *e.e.*: 93.4% (**Figure S36**)

**5a-L**: r.t. 39.74 min, area %:  $2.8 \pm 1.5\%$ ; 42.61 min, area %:  $97.2 \pm 1.5\%$ ; *e.e.*: 94.4% (**Figure S37**)

**PNA 5**: 40.09 min, area %:  $97.3 \pm 0.3\%$ ; 43.34 min, area %:  $2.7 \pm 0.3\%$ ; *e.e.*: 94.6% (**Figure S38**)

**PNA 6**: 39.60 min, area %:  $95.0 \pm 2.3\%$ ; 42.68 min, area %:  $5.0 \pm 2.3\%$ ; *e.e.*: 90.0% (**Figure S39**)

The Arg-derivatives **5b-D** and **5b-L** were not eluted under the tested experimental conditions for chiral GC analysis, probably due to the high molecular weight of the corresponding piperazine-2-ones and to the high polarity of the stationary phase. Additional experiments, performed by increasing the temperature (200°C, the maximum value according to the column manufacturer) or the time of analysis (300 min), failed in the elution of the given compounds (**Figure S40** and **S41**). Accordingly, **PNA 3** was not submitted to the same analysis.

## 8 Measurements of $T_m$ values

Thermal denaturation profiles were measured at circular dichroism in 1mL quartz cuvette, by monitoring both the CD and UV (HT channel) signals in the 300 - 200 nm wavelength range. 5  $\mu$ M Solutions of ssDNA (KRAS wild-type and G12D mutated), ssPNA (**PNA1**, **2**, **3**, **4** and **5**) and corresponding PNA-DNA duplexes in PBS buffer (100 mM NaCl, 10 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 7) were monitored from 18 °C to 90 °C with a heating rate of 1°C/min. The highest differences between the profiles were found at 266 nm. Melting temperatures were calculated from the first derivative of the 10<sup>th</sup> order polynomial that best fitted the experimental data by using a dedicated Matlab script (Figures **S42** –**S46**).

## 9 NMR spectra

### Compound 2-D

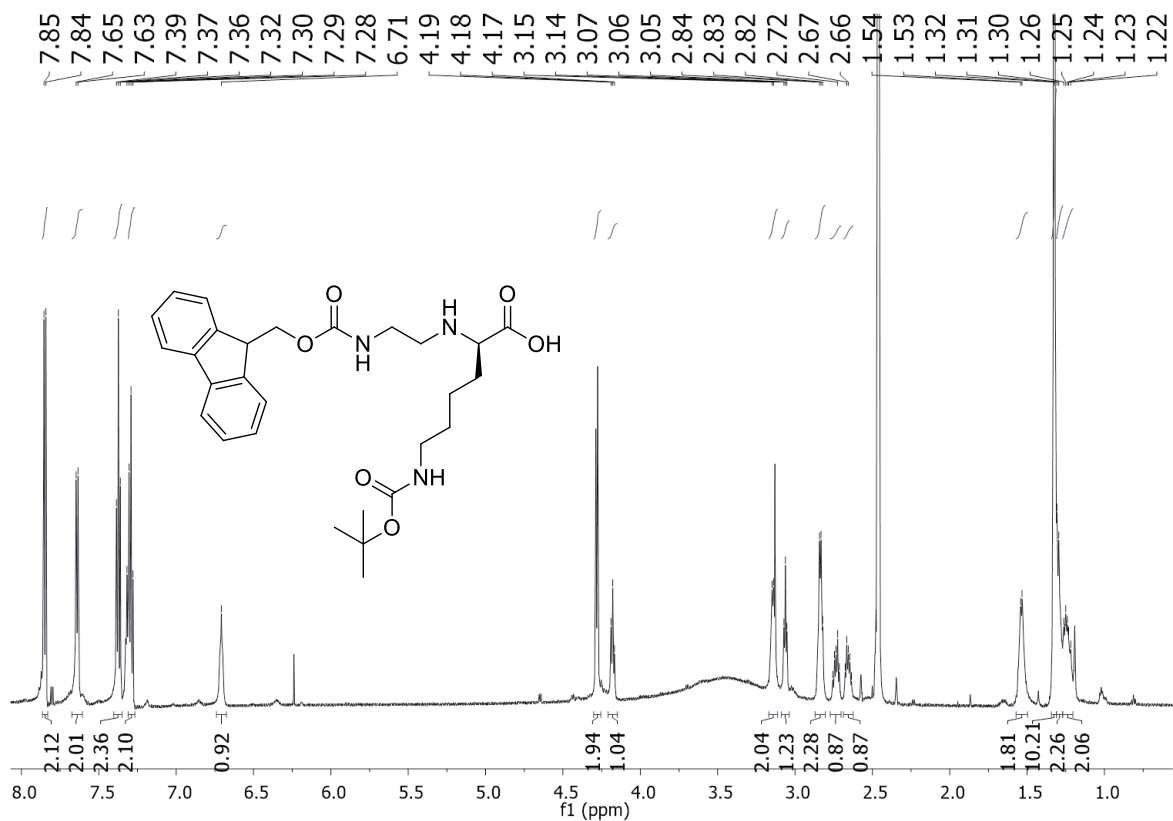


Figure S 1:  $^1\text{H-NMR}$  (600 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound 2-D.

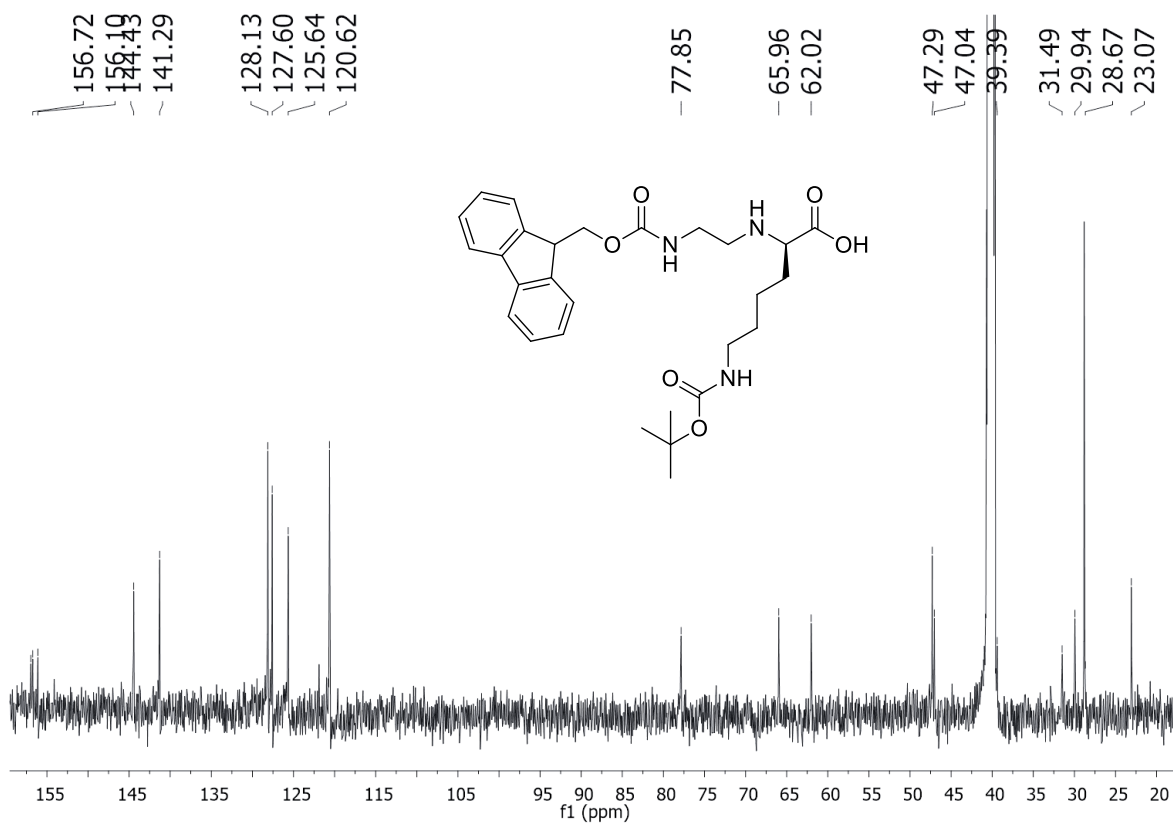
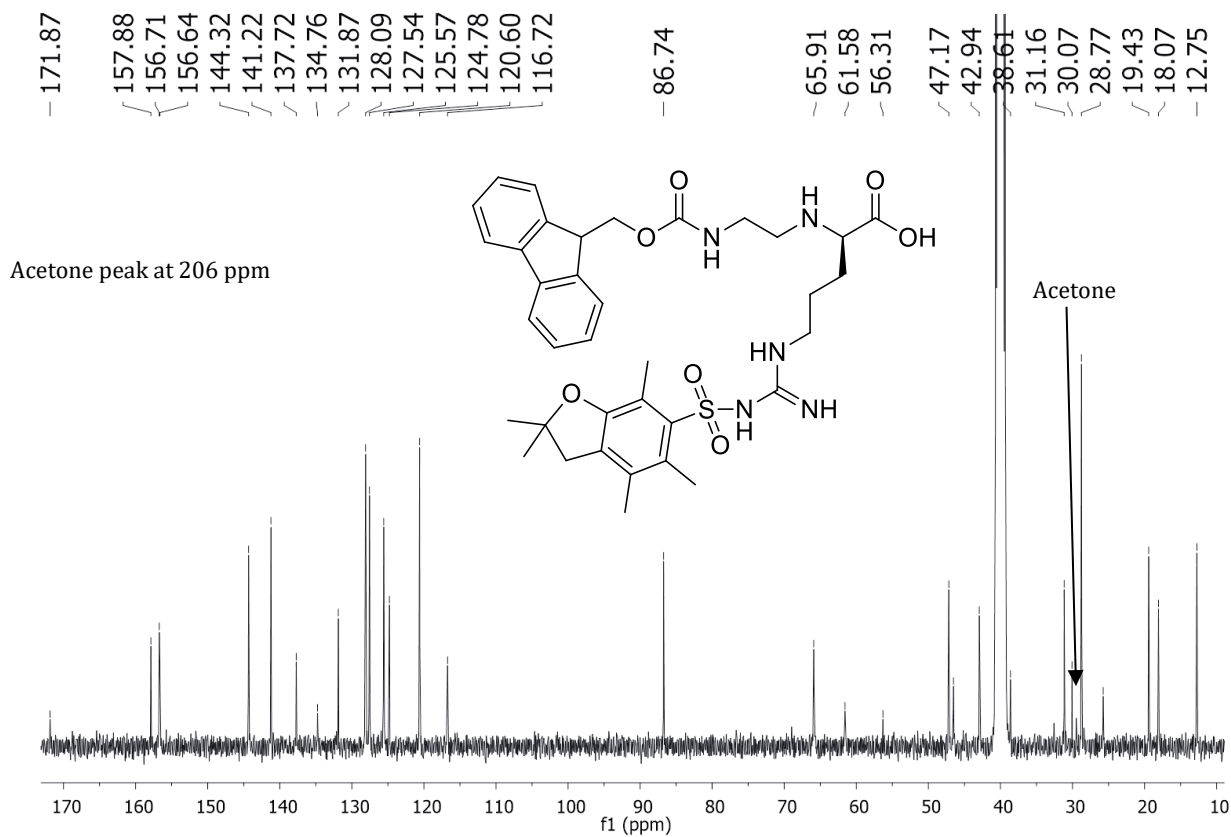
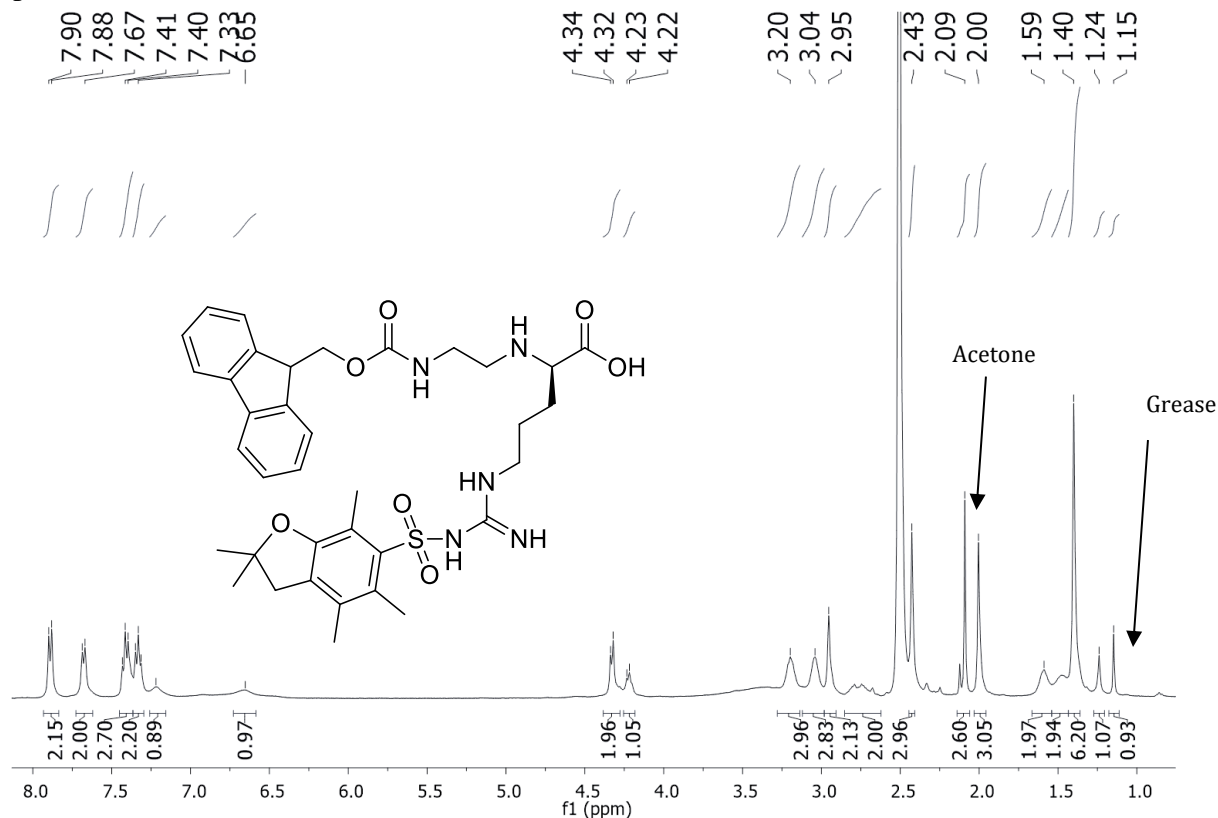


Figure S 2:  $^{13}\text{C-NMR}$  (151 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound 2-D.

Compound 3-D





Compound **3-L**

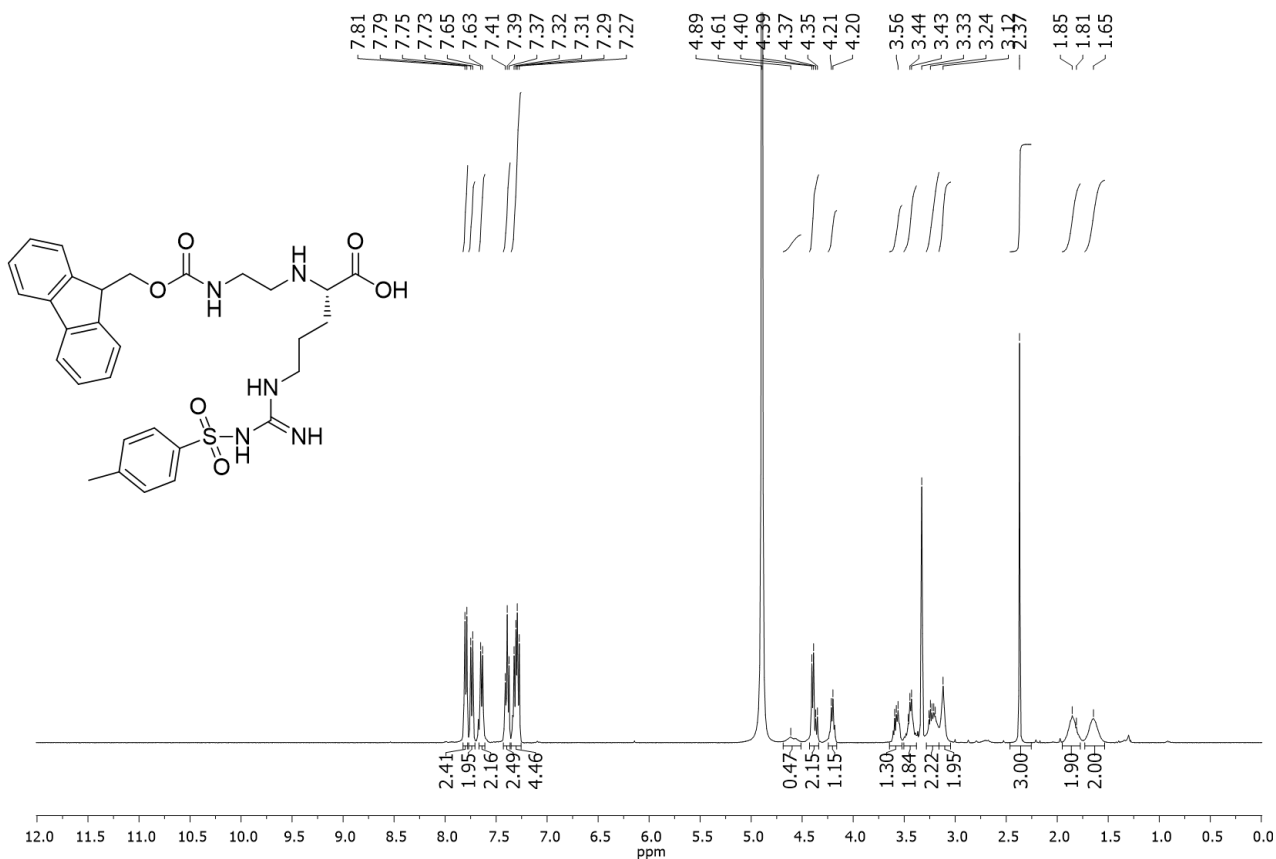


Figure S 5: <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD) of compound **3-L**.

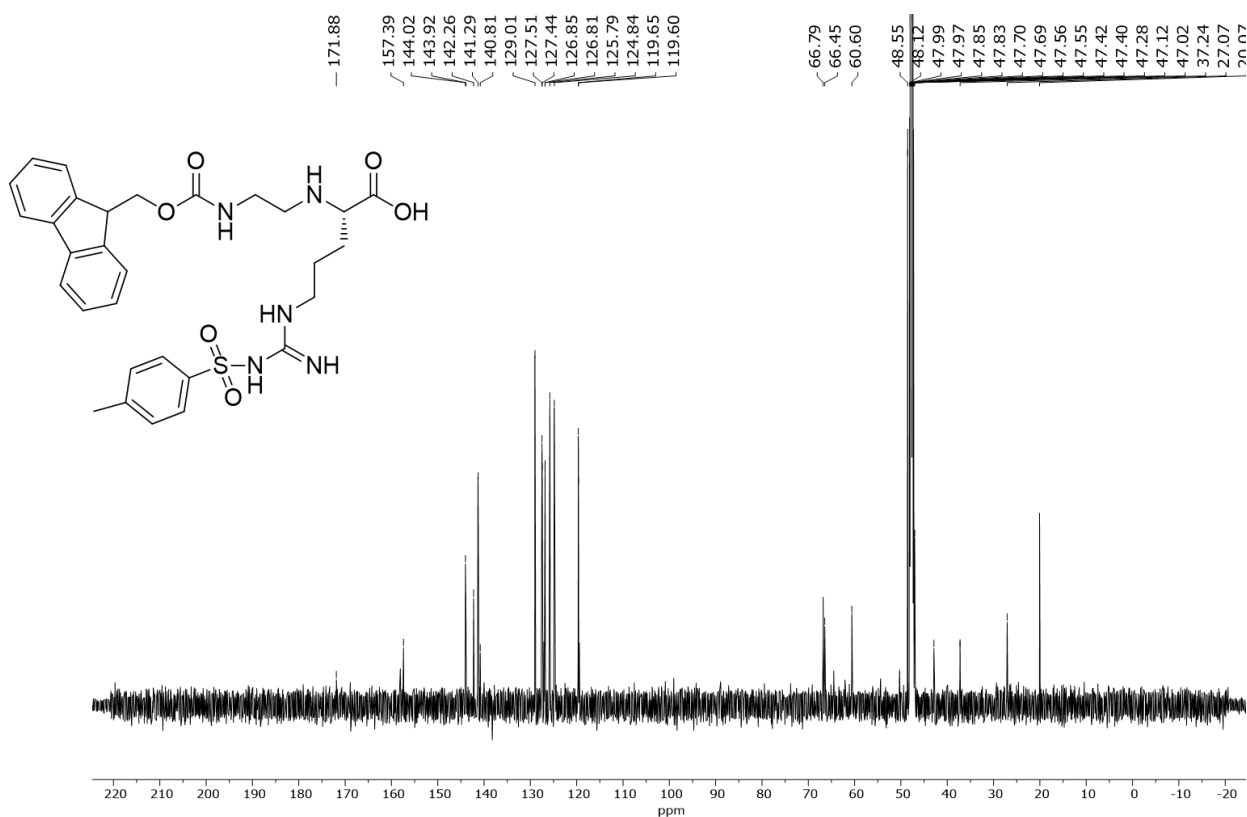


Figure S 6: <sup>13</sup>C NMR (151 MHz, CD<sub>3</sub>OD) of compound **3-L**.

Compound **6a-D**

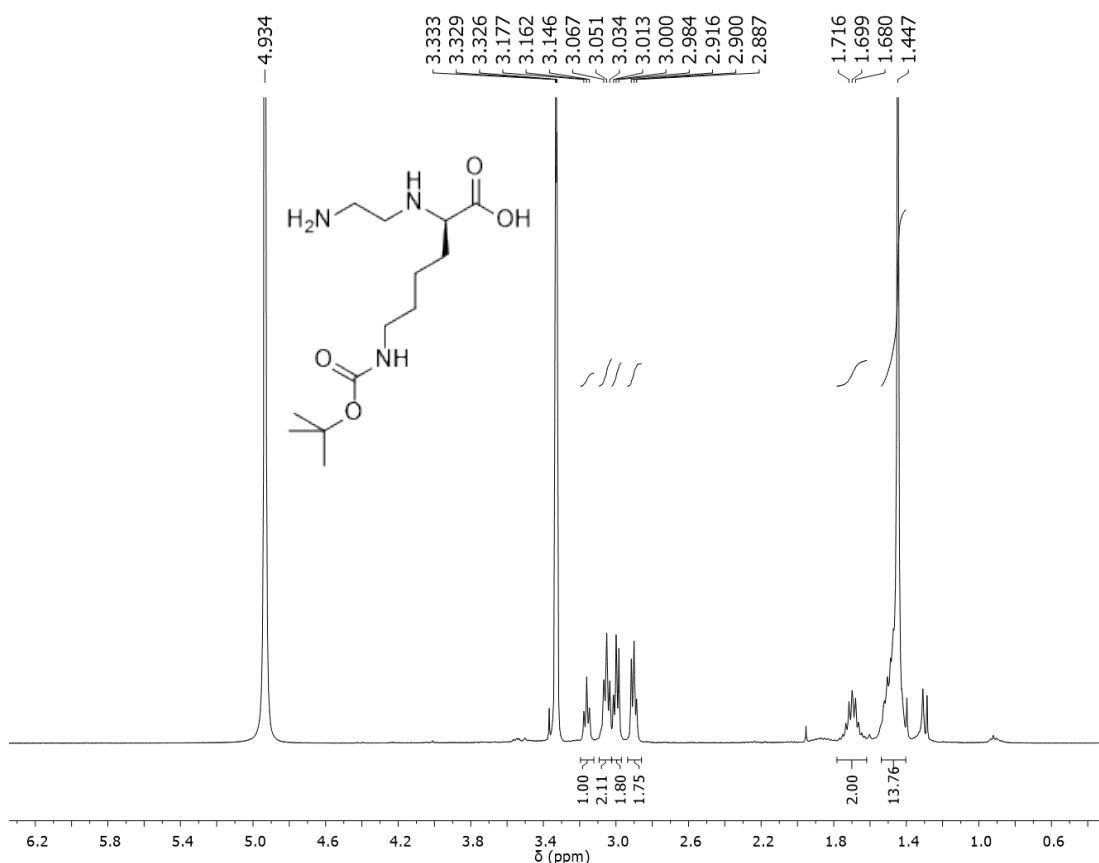


Figure S 7:  $^1\text{H NMR}$  (400 MHz,  $\text{CD}_3\text{OD}$ ) of compound **6a-D**.

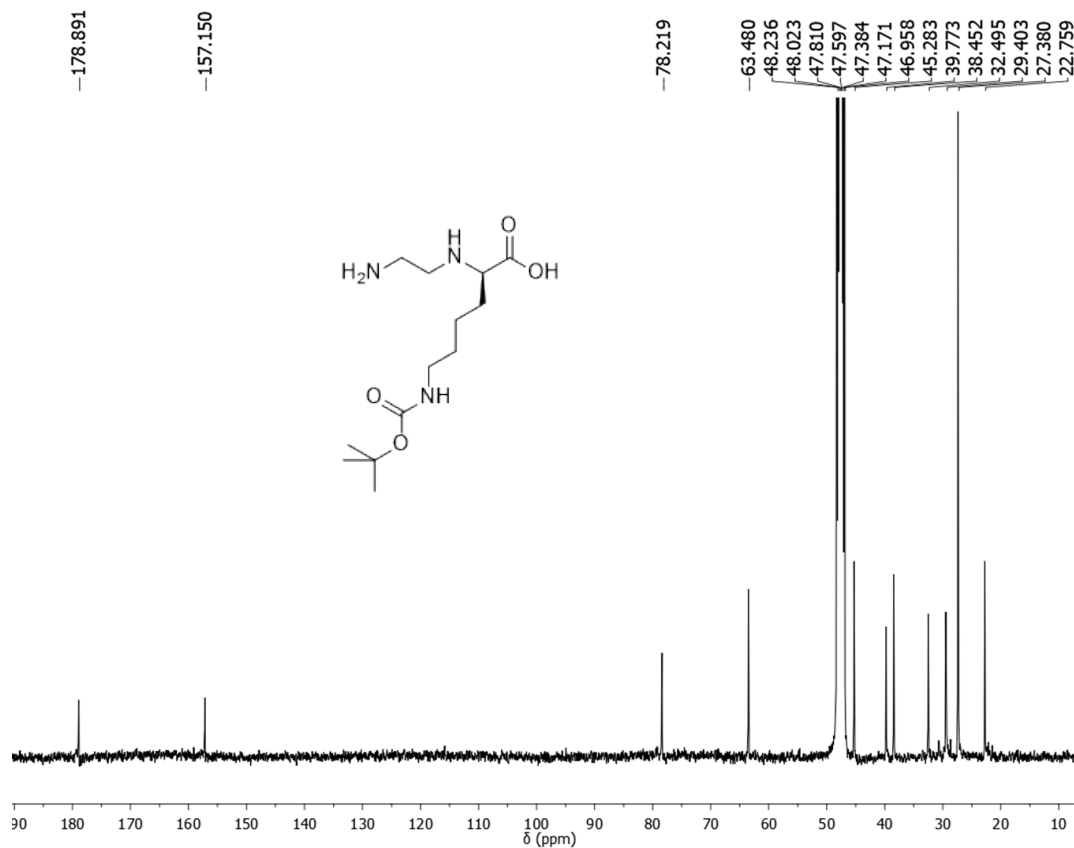
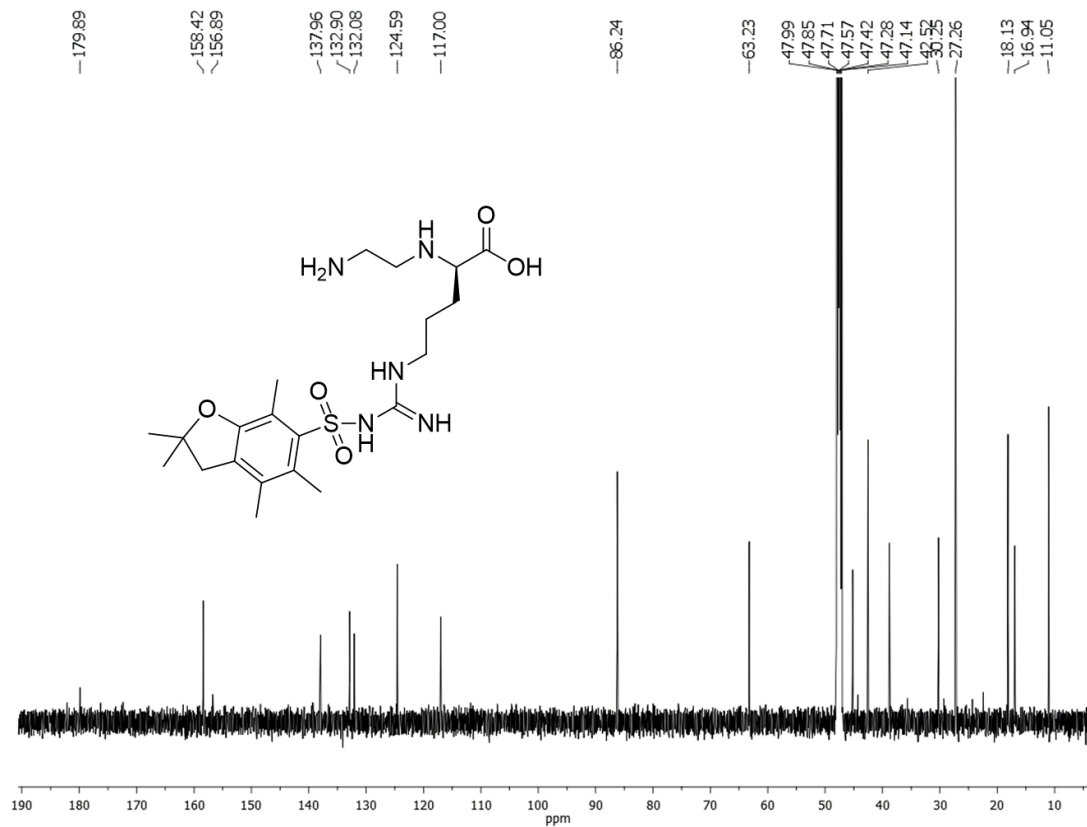
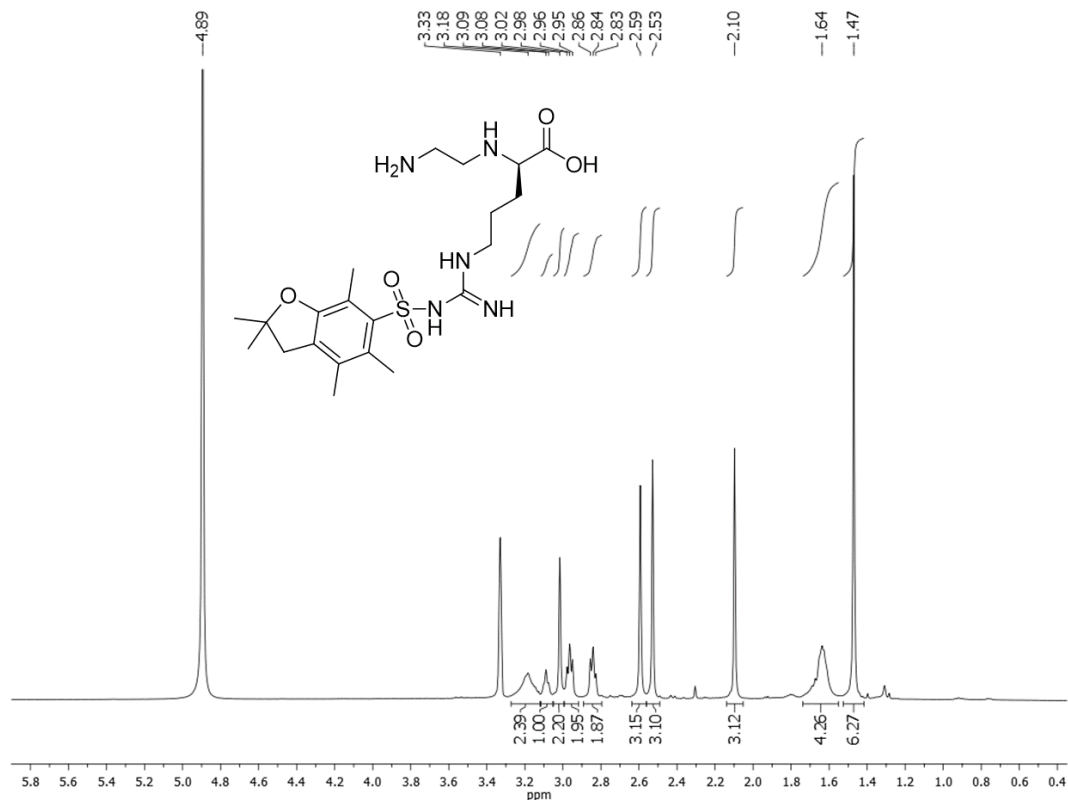


Figure S 8:  $^{13}\text{C NMR}$  (151 MHz,  $\text{CD}_3\text{OD}$ ) of compound **6a-D**.

Compound **6b-D**



Compound **6b-L**

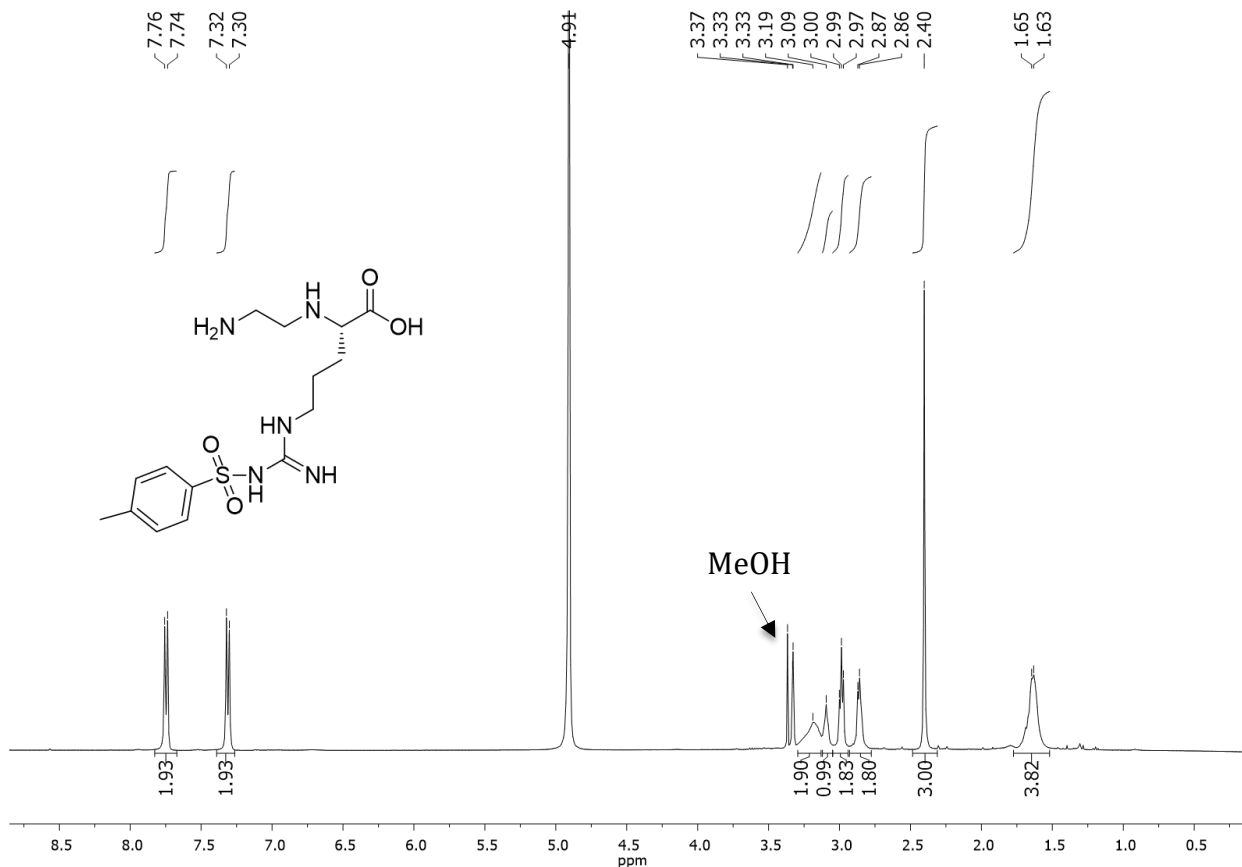


Figure S 11:  $^1\text{H NMR}$  (400 MHz,  $\text{CD}_3\text{OD}$ ) of compound **6b-L**.

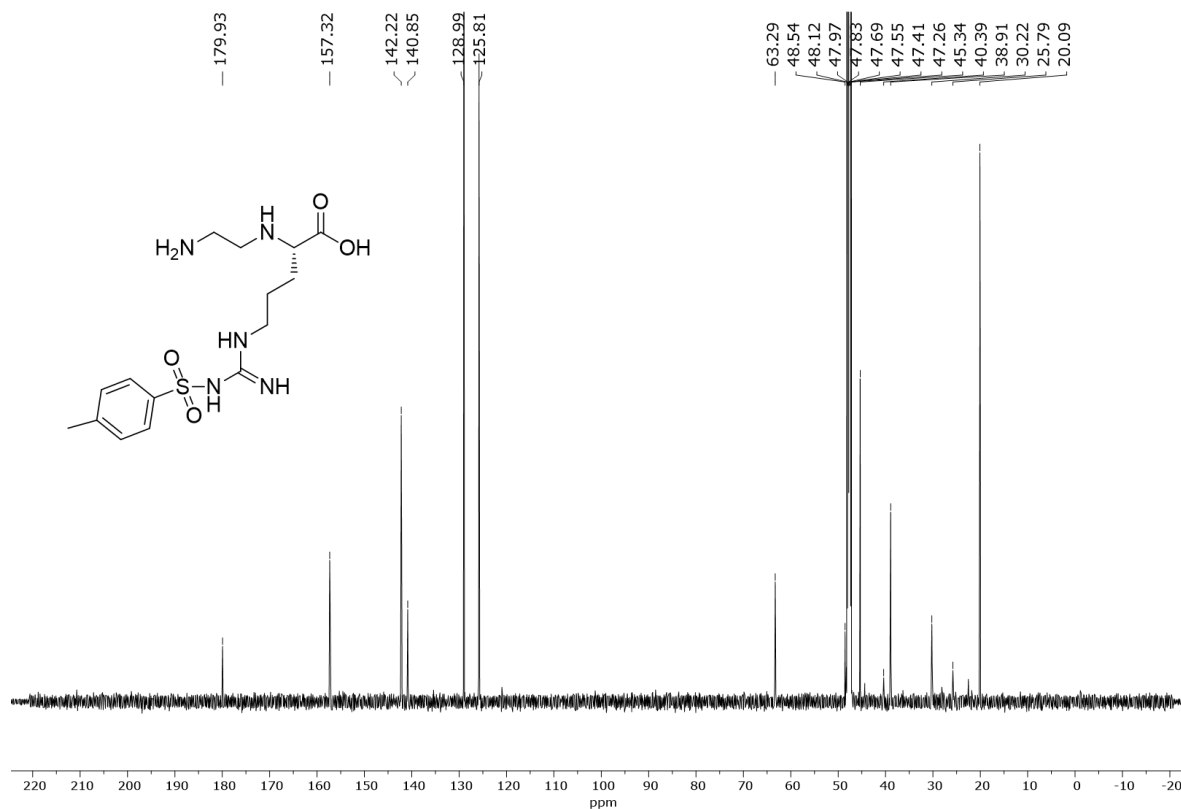


Figure S 12:  $^{13}\text{C NMR}$  (151 MHz,  $\text{CD}_3\text{OD}$ ) of compound **6b-L**.

# Compound 8

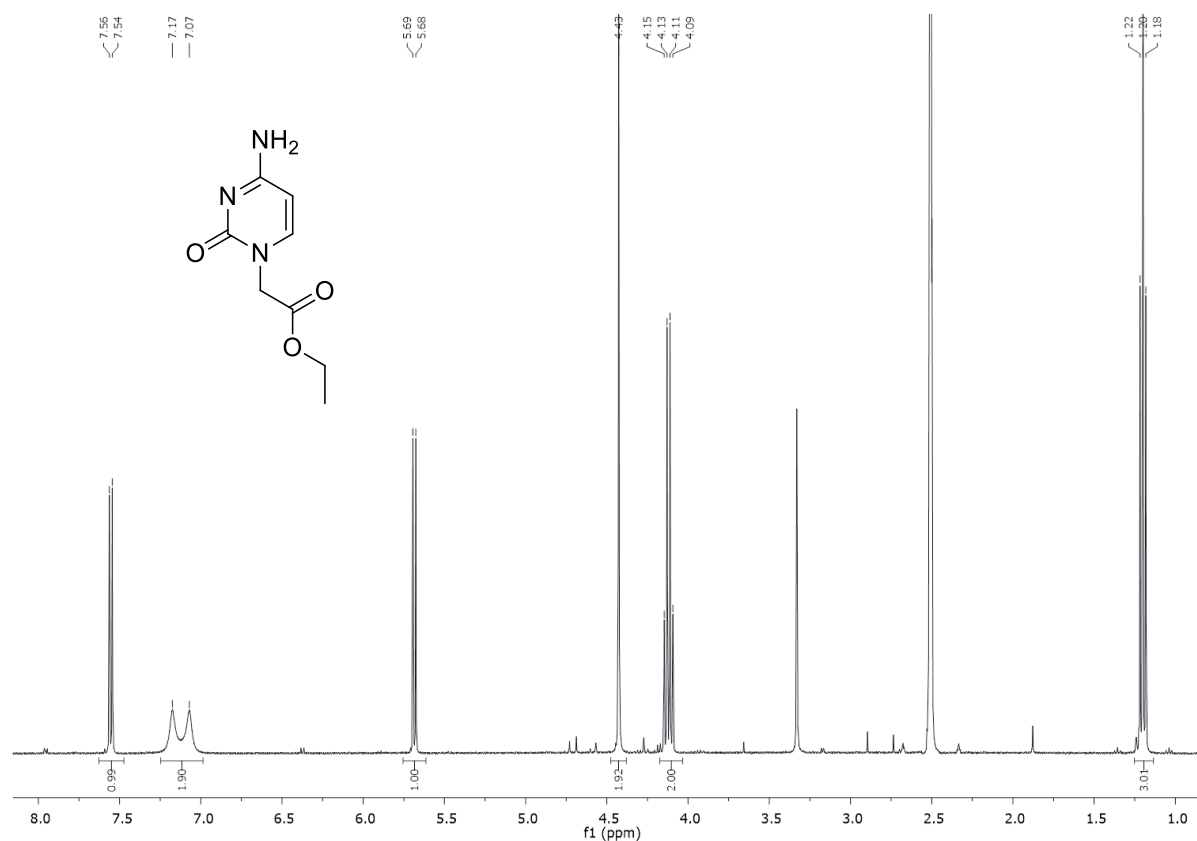


Figure S 13:  $^1\text{H NMR}$  (400 MHz,  $\text{DMSO-}d_6$ ) of compound 8.

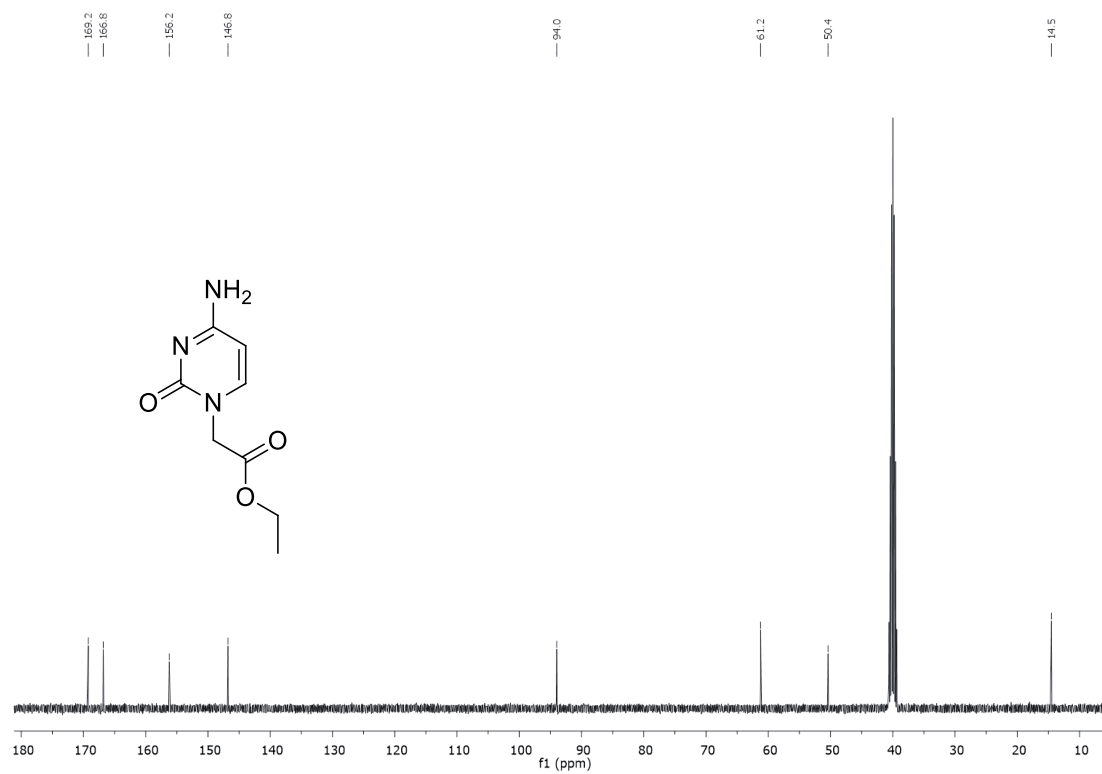


Figure S 14:  $^{13}\text{C NMR}$  (101 MHz,  $\text{DMSO-}d_6$ ) of compound 8.

# Compound 9

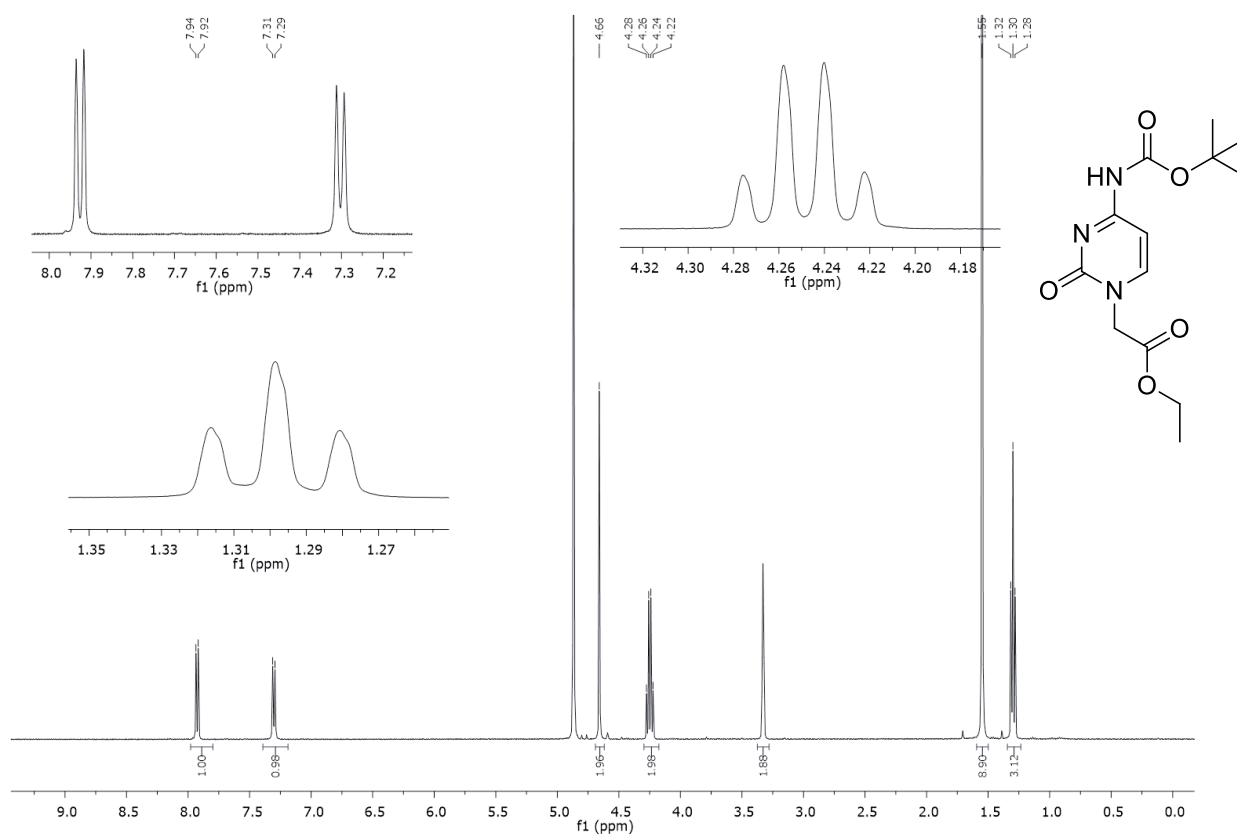


Figure S 15: <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD) of compound 9.

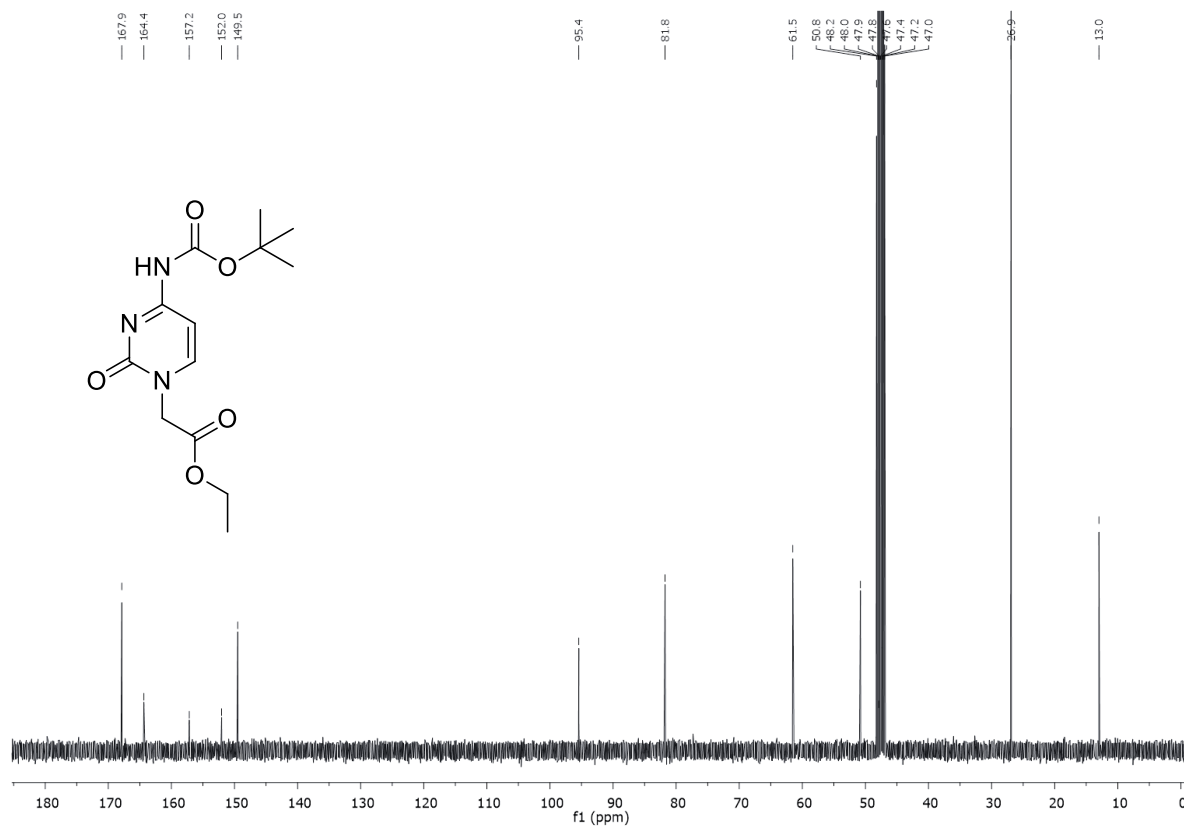
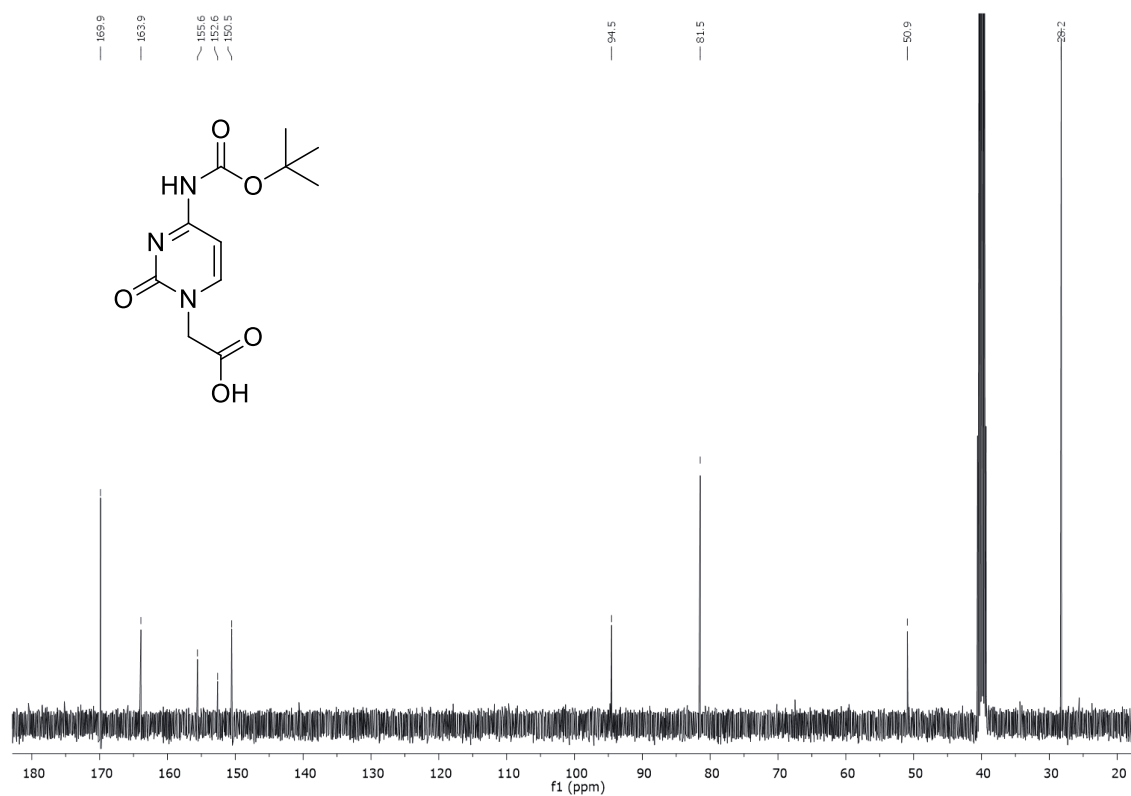
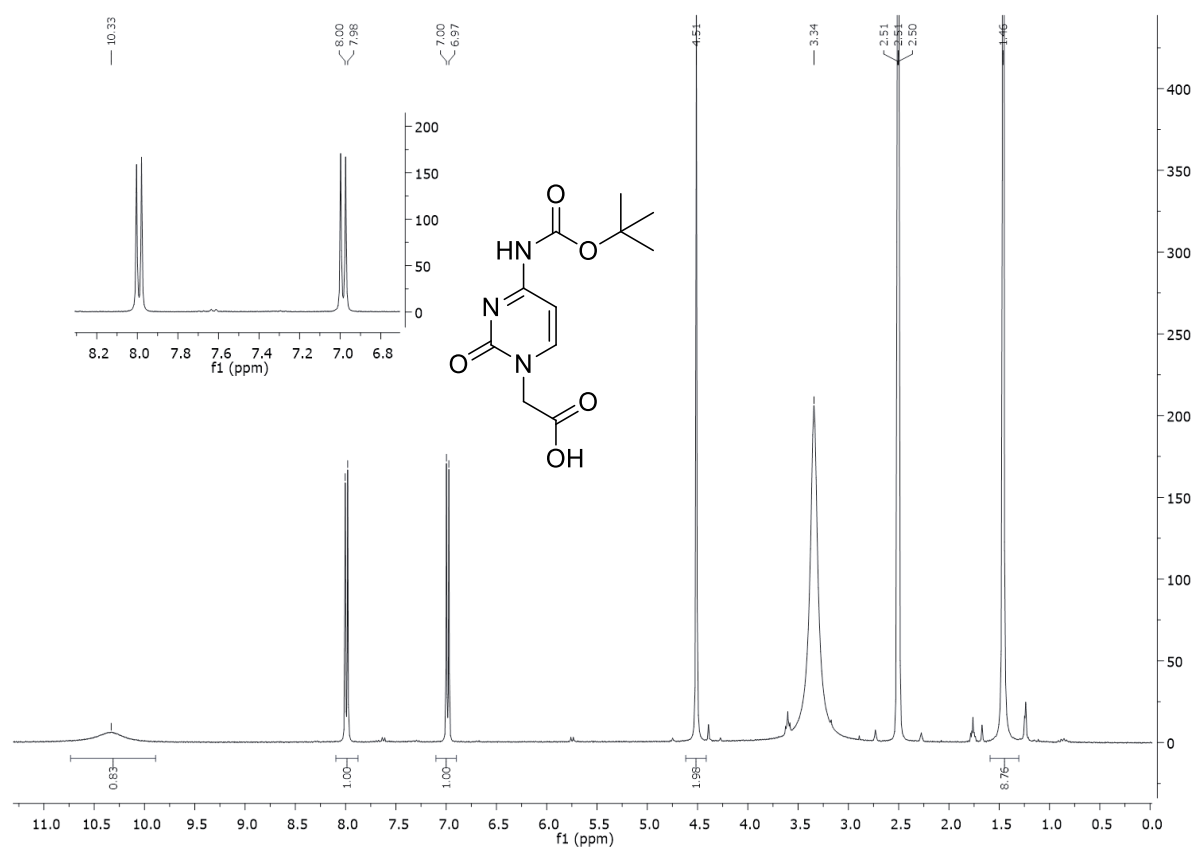


Figure S 16: <sup>13</sup>C NMR (101 MHz, CD<sub>3</sub>OD) of compound 9.

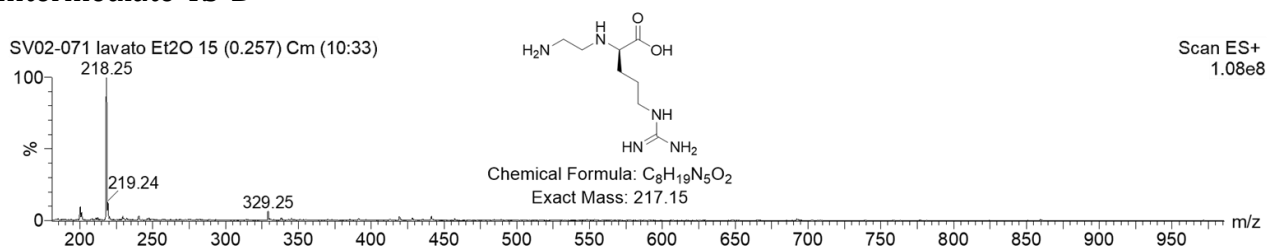
# Compound 10



## 10 ESI-MS spectra

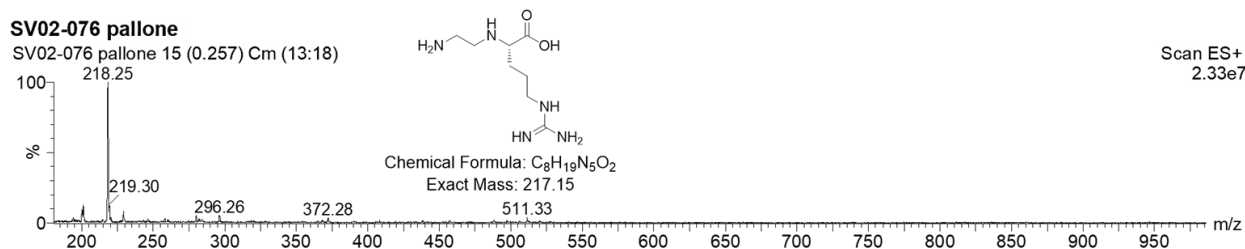
### 10.1 Deprotection of compounds 6b-D and 6b-L

#### Intermediate 4b-D



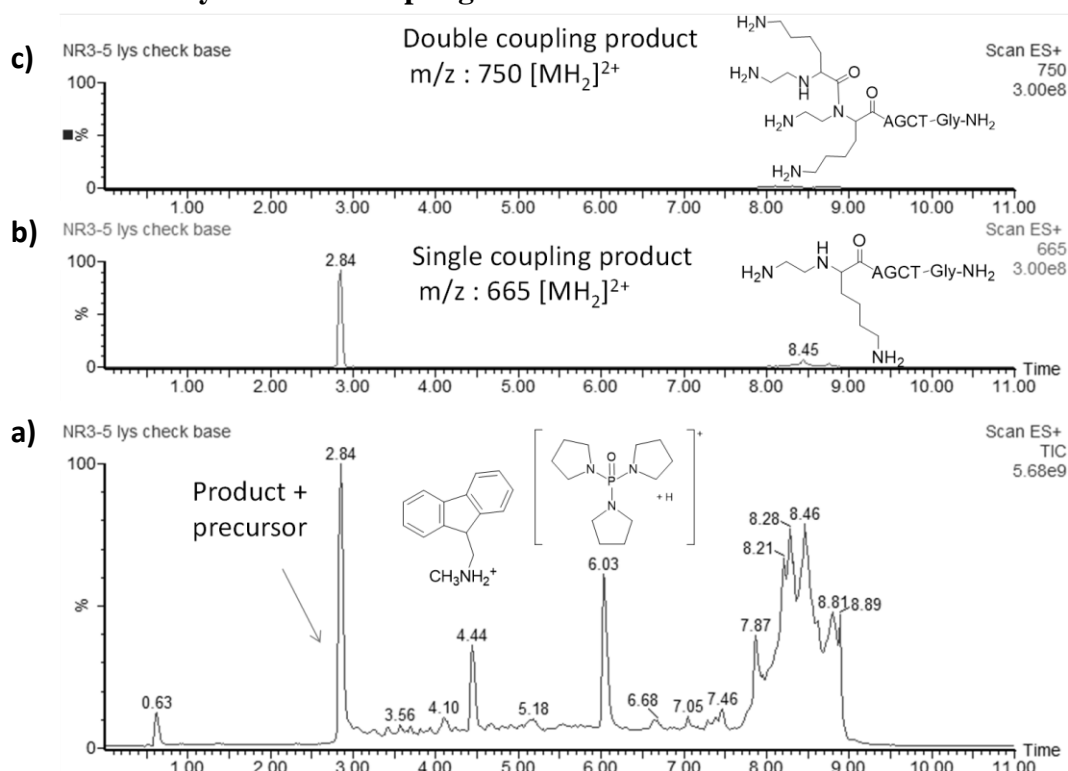
**Figure S 19:** ESI-MS (+) control for the formation of intermediate **4b-D** by removal of Pbf protecting group from **6b-D**.

#### Intermediate 4b-L



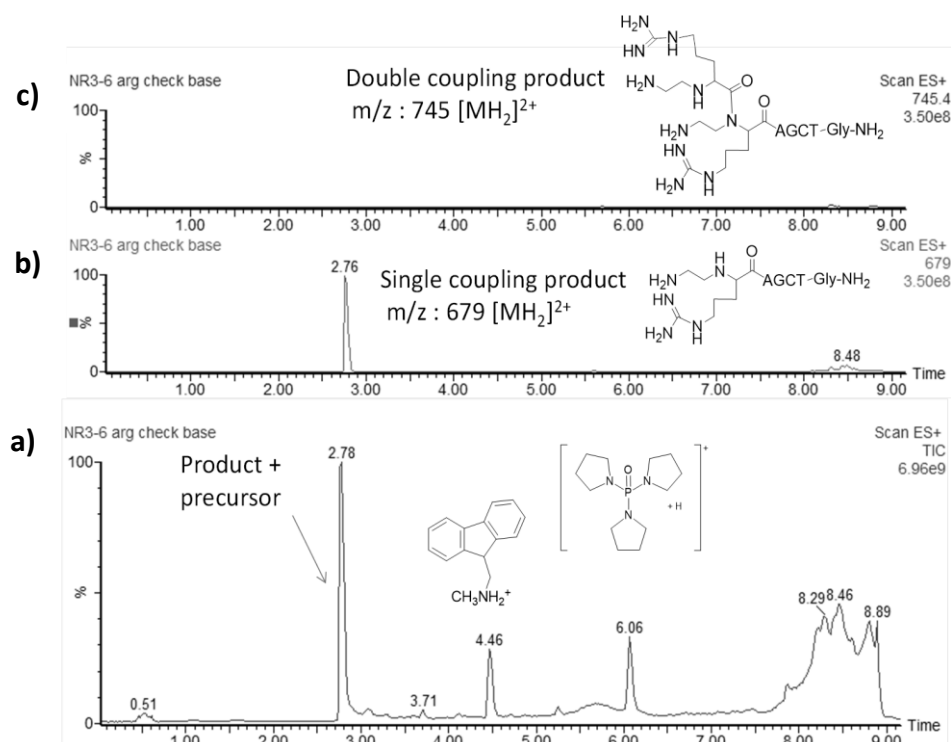
**Figure S 20:** ESI-MS (+) control for the formation of intermediate **4b-L** by removal of Tos protecting group from **6b-L**.

### 10.2 UPLC-MS analysis for the coupling of submonomers 2-D and 3-D



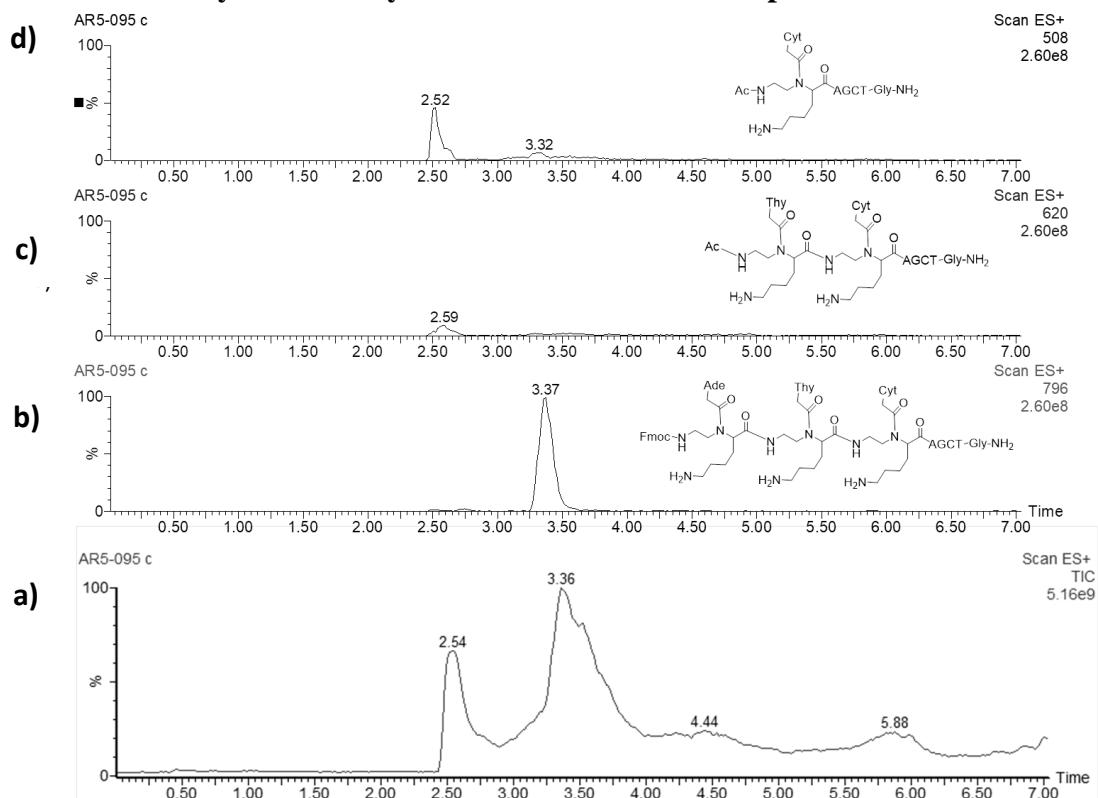
**Figure S 21:** UPLC-MS analysis for the coupling of the first Lys-based submonomer **2-D** on PNA **2**. **a)** Total ion current (TIC) chromatogram of a crude mixture after treatment with  $MeNH_2$  to remove the terminal Fmoc group and transient tris(pyrrrolidine)phosphoroamidate adducts. The traces generated by the  $[M+2H]^{2+}$  ions of derivatives containing **b)** one or **c)** two C(2)-modified backbones were extracted and normalized to the same intensity scale to exclude the occurrence of double-coupling reactions.



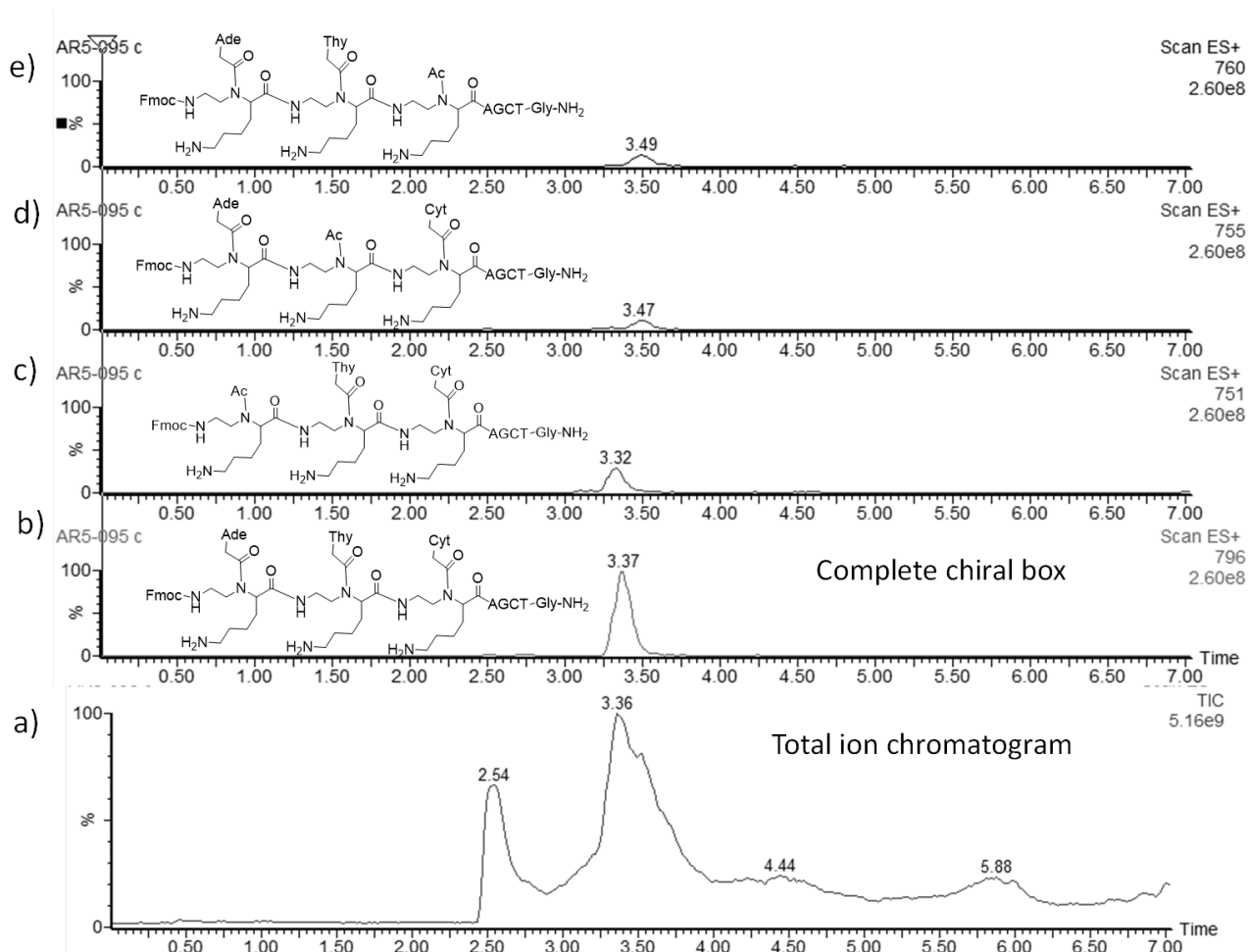


**Figure S 22:** UPLC-MS analysis for the coupling of the first Lys-based submonomer **3-D** on **PNA 3**. **a)** Total ion current (TIC) chromatogram of a crude mixture after treatment with  $\text{MeNH}_2$  to remove the terminal Fmoc group and transient tris(pyrrrolidine)phosphoroamidate adducts. The traces generated by the  $[\text{M}+2\text{H}]^{2+}$  ions of derivatives containing **b)** one or **c)** two C(2)-modified backbones were extracted and normalized to the same intensity scale to exclude the occurrence of double-coupling reactions.

### 10.3 UPLC-MS analysis for the synthesis of the ‘Chiral box’ part of PNA 2



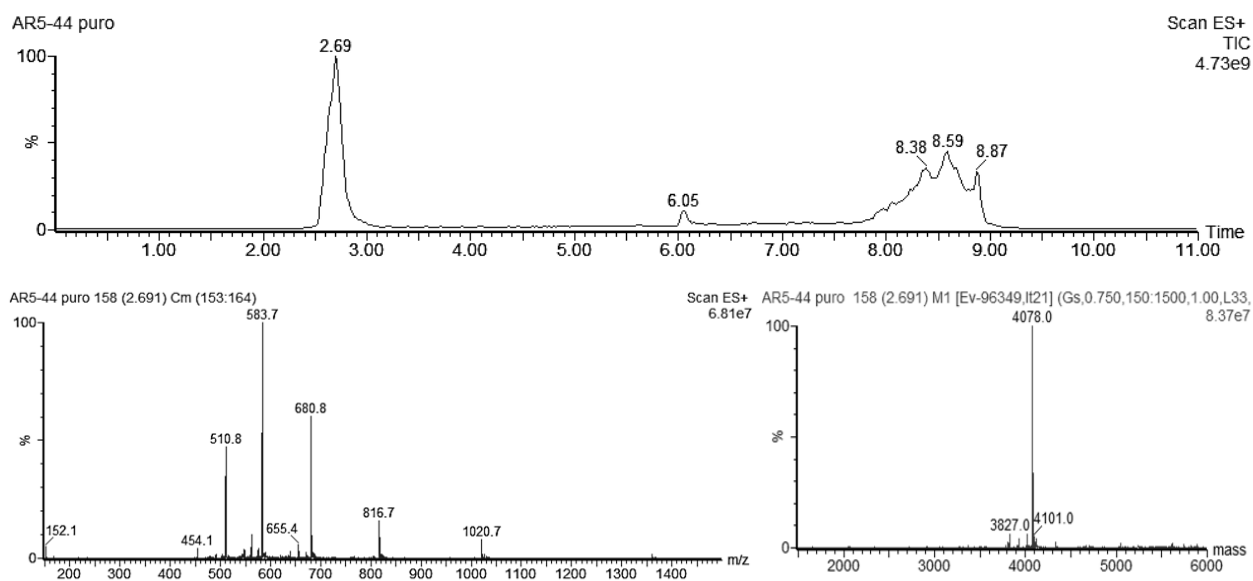
**Figure S 23:** UPLC-MS analysis of **PNA 2** after the completion of its ‘Chiral box’ stretch. **a)** Total ion current (TIC) chromatogram of a Fmoc-protected crude sample. The traces generated by the  $[\text{M}+3\text{H}]^{3+}$  ions of derivatives containing **b)** three, **c)** two or **d)** one C(2)-modified monomers were extracted and normalized to the same intensity scale for evaluating the efficiency of chain elongation (Ac = acetyl group, Ade = Adenine, Thy = Thymine, Cyt = cytosine).



**Figure S 24:** UPLC-MS analysis of PNA 2 after the completion of its 'Chiral box' stretch. **a)** Total ion current (TIC) chromatogram of a Fmoc-protected crude sample. The traces generated by the  $[M+3H]^{3+}$  ions of derivatives acetylated on **b)** none or **c) - e)** one of their N(3) positions were extracted and normalized to the same intensity scale for evaluating the efficiency of the nucleobase insertion (Ac = acetyl group, Ade = Adenine, Thy = Thymine, Cyt = cytosine).

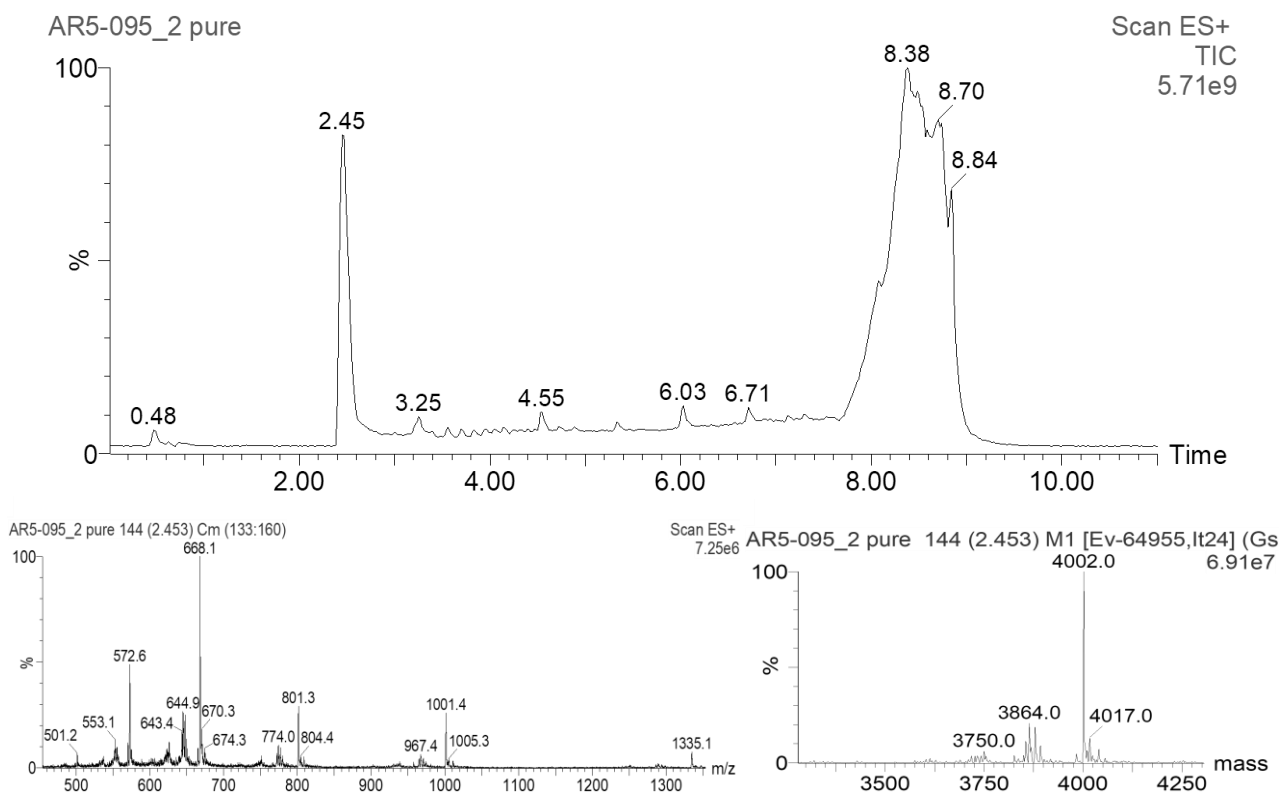
## 10.4 UPLC-MS chromatograms of pure PNAs

### PNA 1



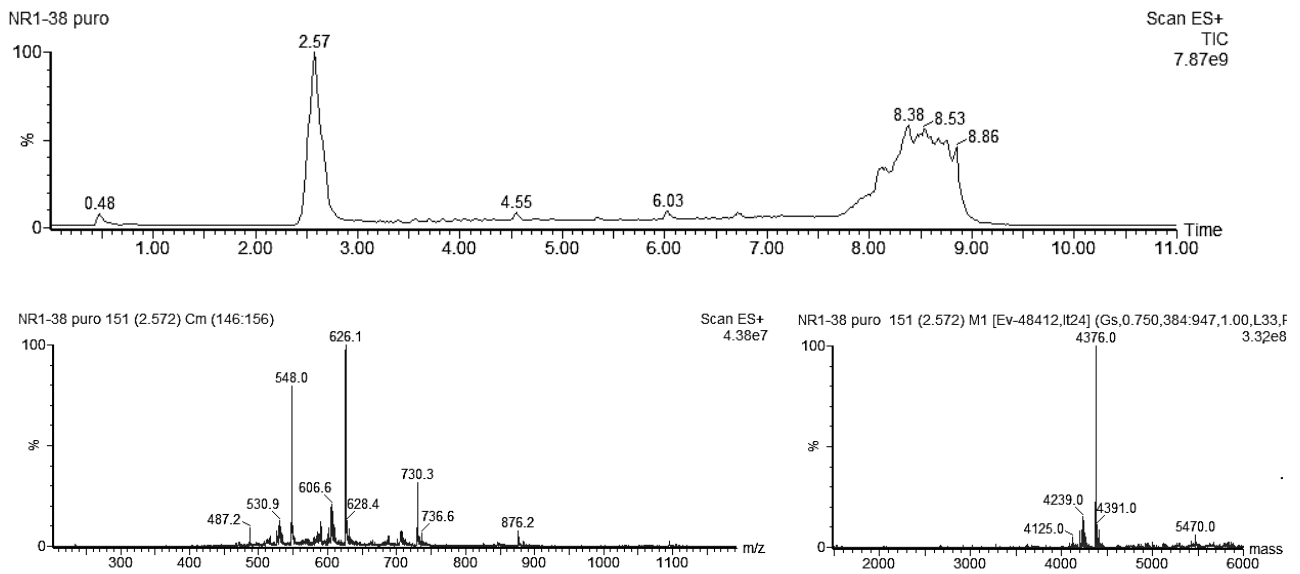
**Figure S 25:** UPLC-MS chromatogram of pure **PNA1** (top) and ESI-MS spectrum for the peak at 2.78 min with the corresponding mathematic deconvolution of the multicharged signals (bottom).

### PNA 2



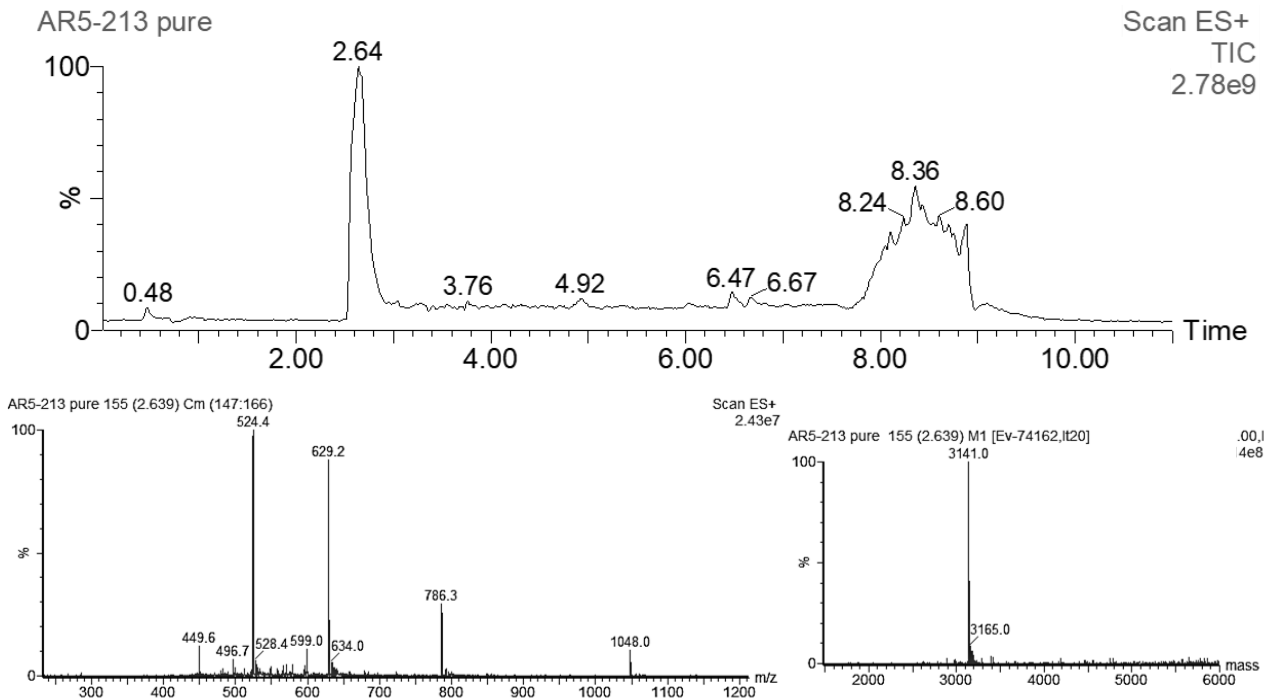
**Figure S 26:** UPLC-MS chromatogram of pure **PNA2** (top) and ESI-MS spectrum for the peak at 2.45 with min the corresponding mathematic deconvolution of the multicharged signals (bottom).

### PNA 3



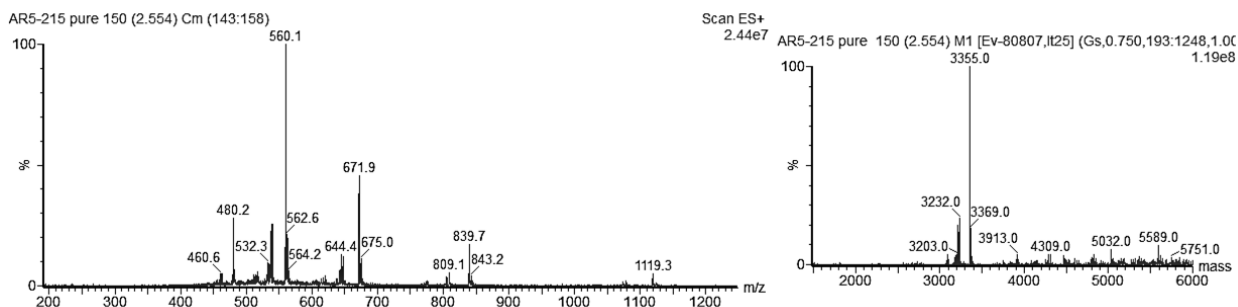
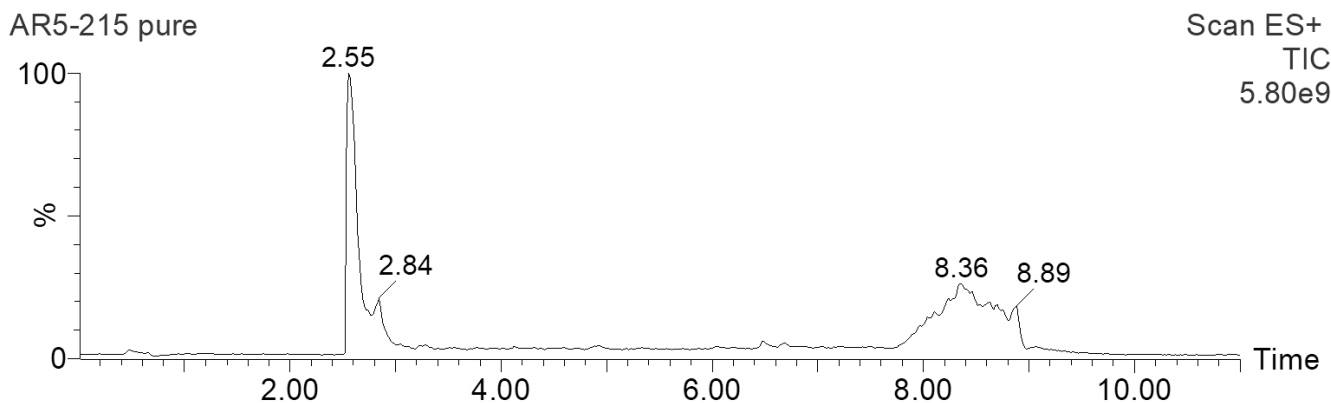
**Figure S 27:** UPLC-MS chromatogram of pure **PNA2** (top) and ESI-MS spectrum for the peak at 2.45 with min the corresponding mathematic deconvolution of the multicharged signals (bottom).

### PNA 4



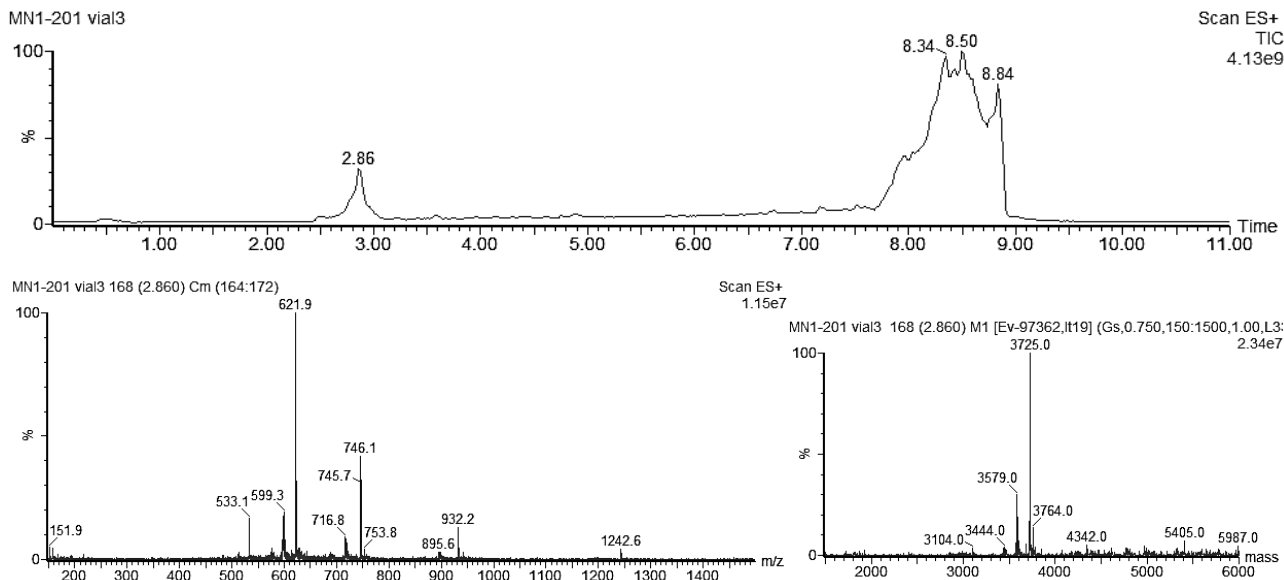
**Figure S 28:** UPLC-MS chromatogram of pure **PNA4** (top) and ESI-MS spectrum for the peak at 2.64 min with the corresponding mathematic deconvolution of the multicharged signals (bottom).

### PNA5



**Figure S 29:** UPLC-MS chromatogram of pure **PNA5** (top) and ESI-MS spectrum for the peak at 2.55 min with the corresponding mathematic deconvolution of the multicharged signals (bottom).

## PNA 6

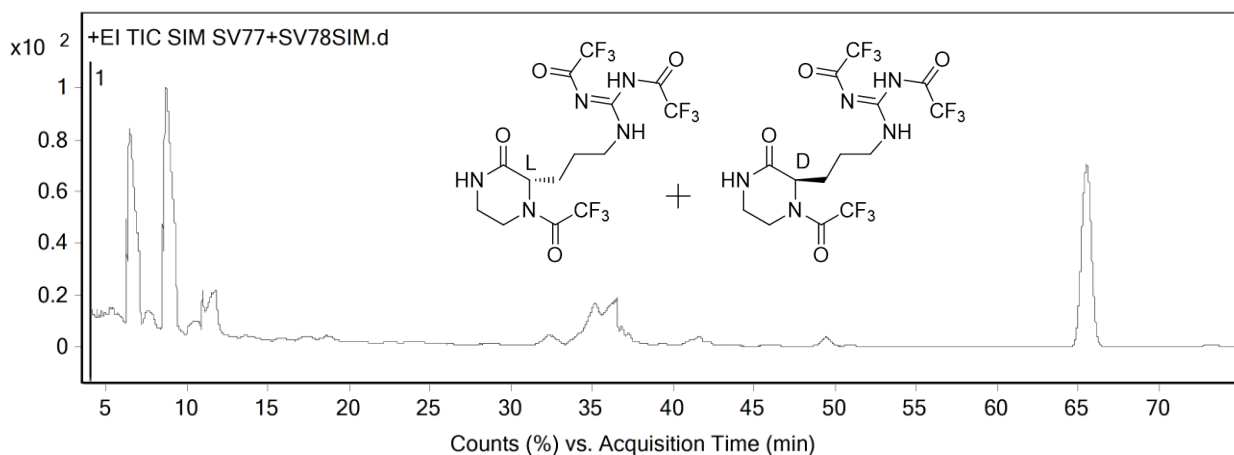


**Figure S 30:** UPLC-MS chromatogram of pure **PNA6** (top) and ESI-MS spectrum for the peak at 2.86 min with the corresponding mathematic deconvolution of the multicharged signals (bottom).

## 11 GC-MS spectra

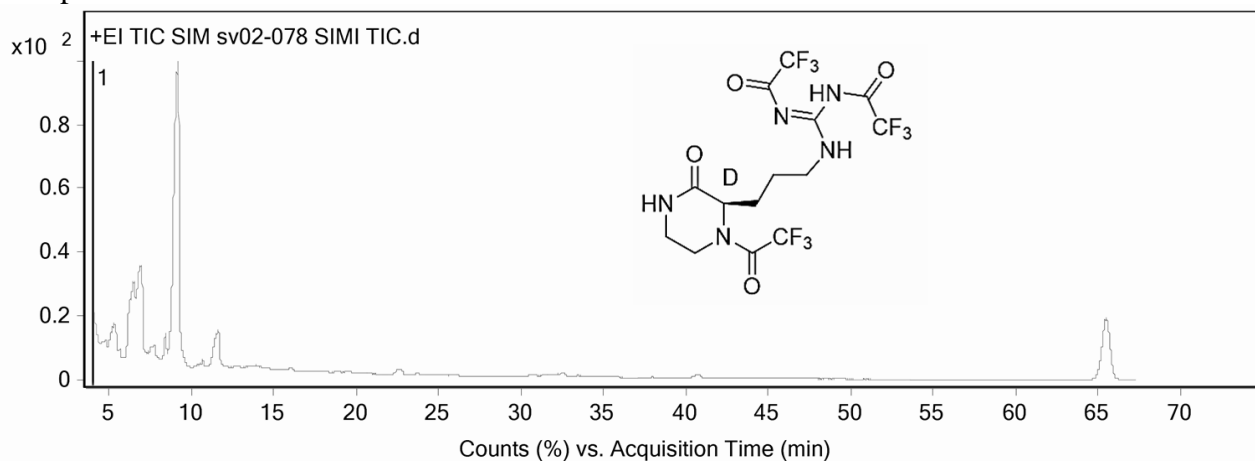
### 11.1 Achiral analyses

1/1 mixture of compounds **5b-D/5b-L**



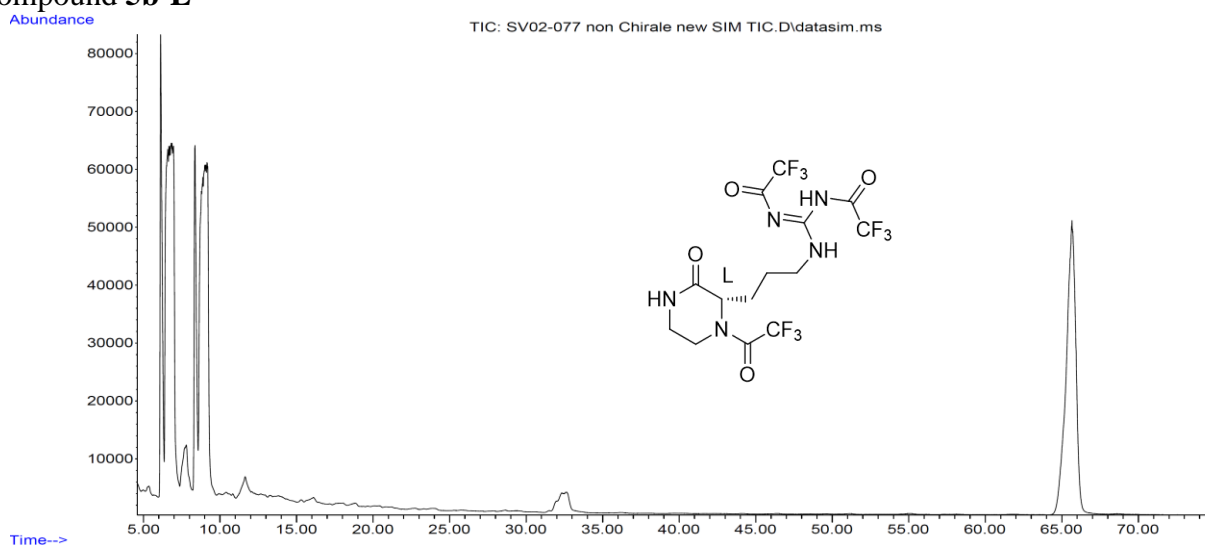
**Figure S 31:** GC-MS chromatogram (SIM mode:  $m/z$  149, 167, 279) of a 1/1 mixture of compounds **5b-D** and **5b-L**. The two piperazine-2-ones gave a common peak at 65.52 min.

### Compound **5b-D**



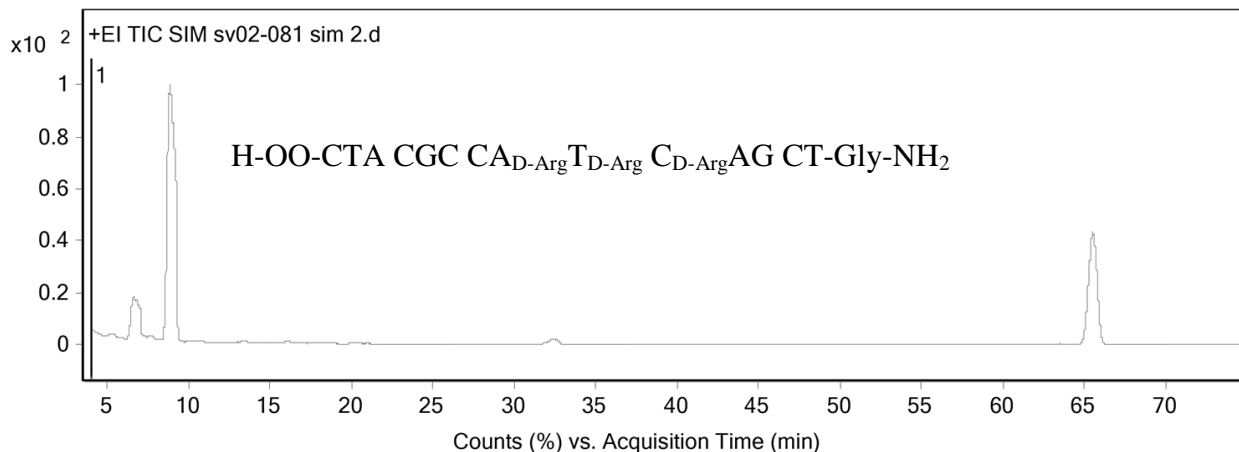
**Figure S 32:** GC-MS chromatogram (SIM mode:  $m/z$  149, 167, 279) of compound **5b-D**. The given piperazine-2-one eluted at 65.49 min.

### Compound **5b-L**



**Figure S 33:** GC-MS chromatogram (SIM mode:  $m/z$  149, 167, 279) of compound **5b-L**. The given piperazine-2-one eluted at 65.66 min.

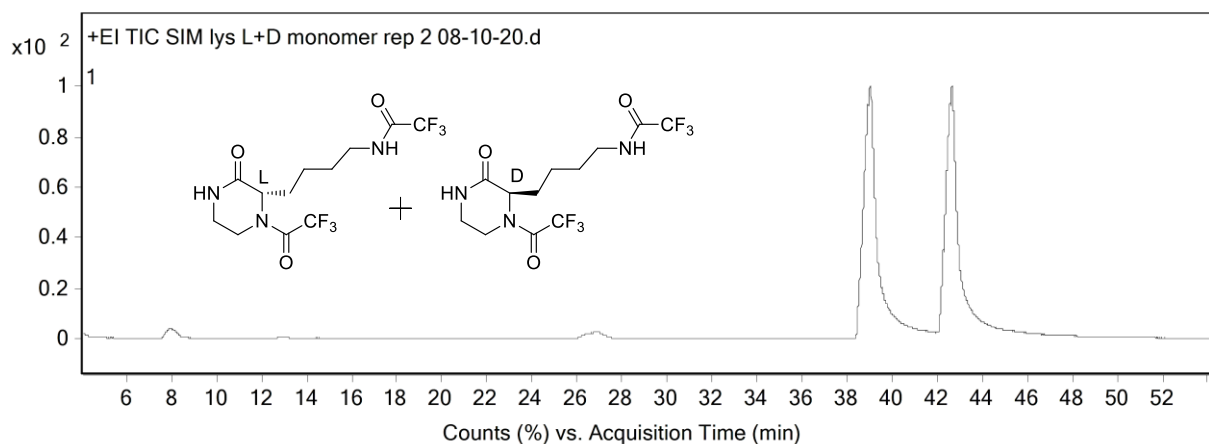
### PNA 3



**Figure S 34:** GC-MS chromatogram (SIM mode:  $m/z$  149, 167, 279) of **PNA 3**. The piperazine-2-ones deriving from Arg-modified submonomers eluted at 65.45 min.

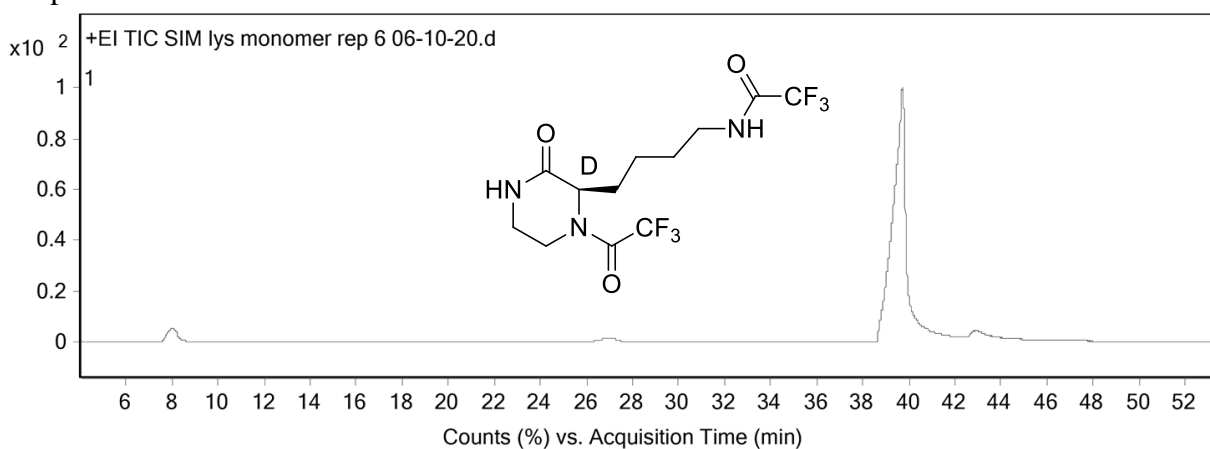
### 11.2 Chiral analyses

1/1 mixture of compounds **5a-D** and **5a-L**



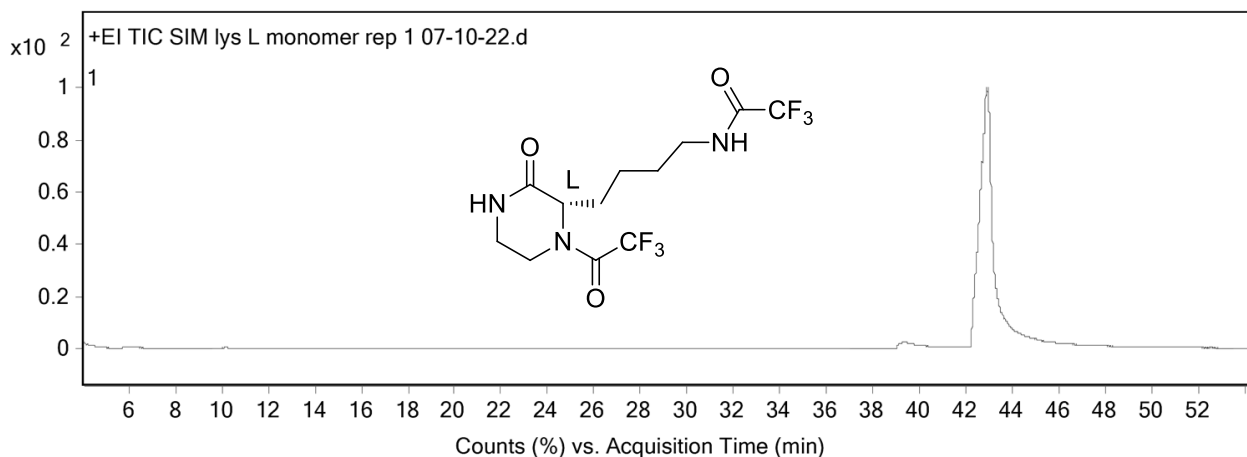
**Figure S 35:** GC-MS chromatogram (SIM mode:  $m/z$  140, 167, 196) of a 1/1 mixture of compounds **5a-D** and **5a-L**. The two piperazine-2-ones eluted at 39.09 min and 42.64 min, respectively.

### Compound **5a-D**



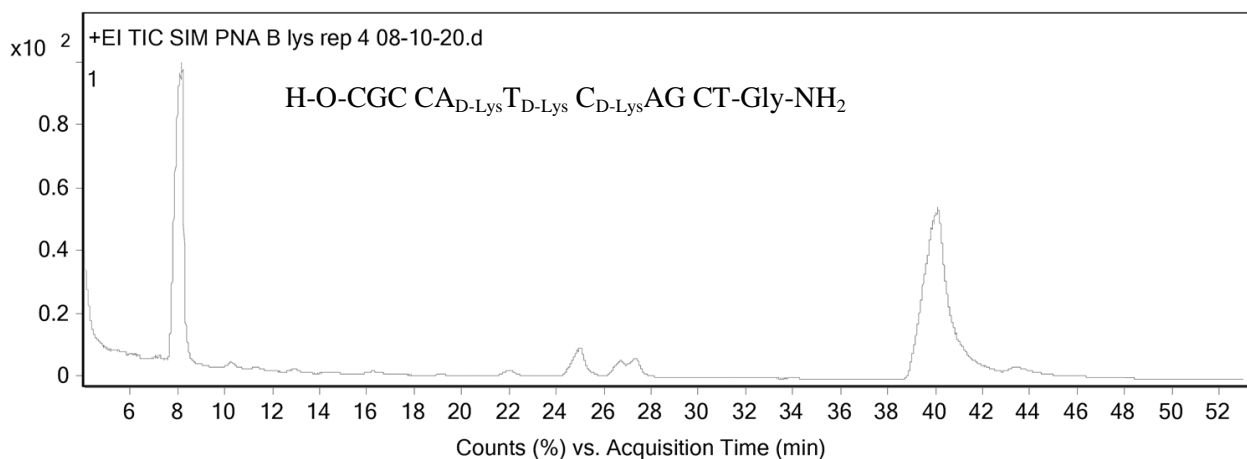
**Figure S 36:** GC-MS chromatogram (SIM mode:  $m/z$  140, 167, 196) of compound **5a-D**. The given piperazine-2-one and traces of its L-enantiomer eluted at 39.54 min and 42.58 min, respectively.

### Compound 5a-L



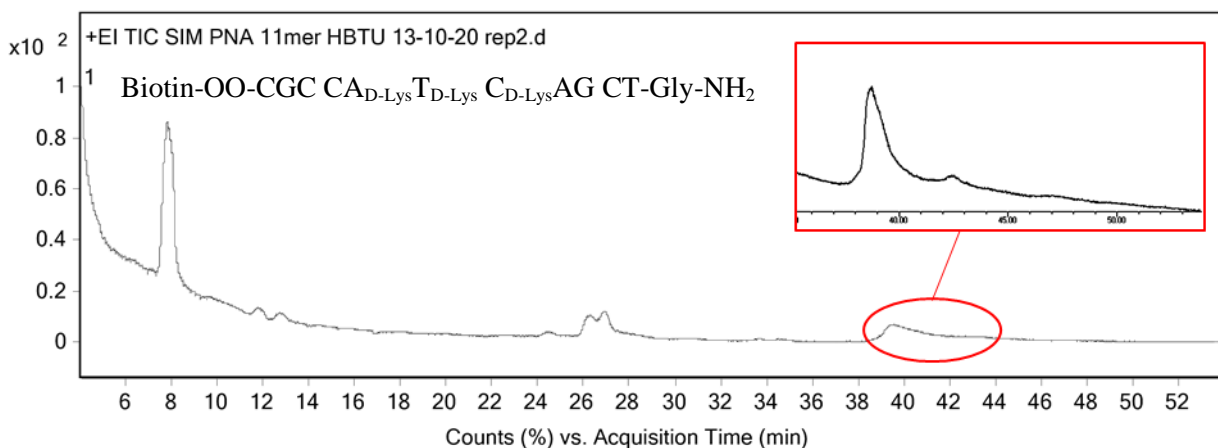
**Figure S 37:** GC-MS chromatogram (SIM mode:  $m/z$  140, 167, 196) of compound **5a-L**. The given piperazine-2-one and traces of its D-enantiomer eluted at 42.61 min and 39.74 min, respectively.

### PNA 5



**Figure S 38:** GC-MS chromatogram (SIM mode:  $m/z$  140, 167, 196) of **PNA5**. The piperazine-2-ones deriving from Lys-modified submonomers and traces of their L-enantiomer eluted at 40.09 min and 43.34 min, respectively.

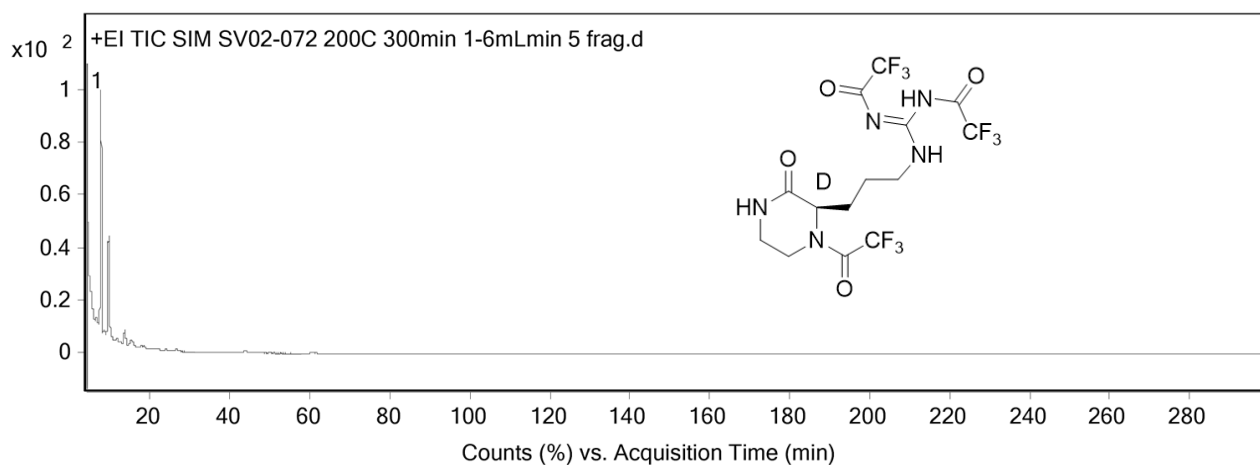
### PNA 6



**Figure S 39:** GC-MS chromatogram (SIM mode:  $m/z$  140, 167, 196) of **PNA2**. The piperazine-2-ones deriving from Lys-modified submonomers and traces of their L-enantiomer eluted at 39.60 min and 42.68 min, respectively.

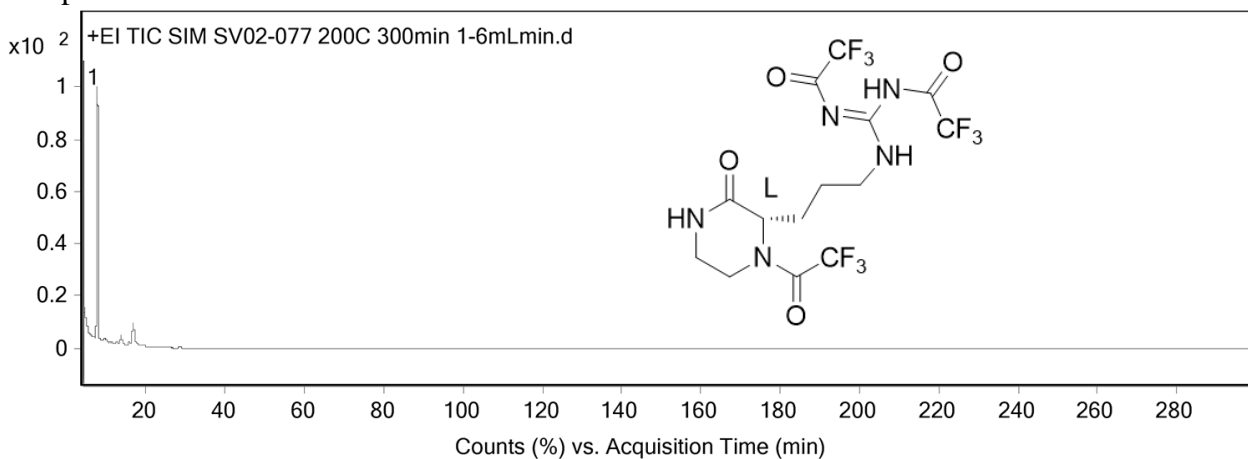


## Compound 5b-D



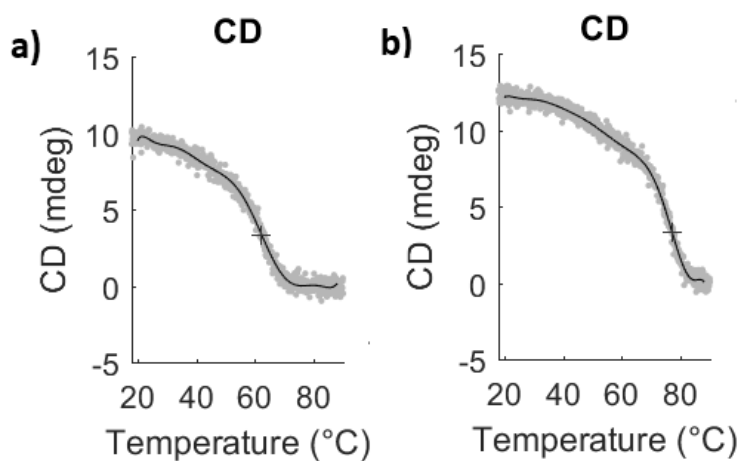
**Figure S 40:** Chiral GC-MS chromatogram (SIM mode: m/z 149, 167, 279) of compound **5b-D** performed at 200°C x 300 min. No peaks ascribable to the given piperazine-2-one were eluted under the tested experimental conditions.

## Compound 5b-L

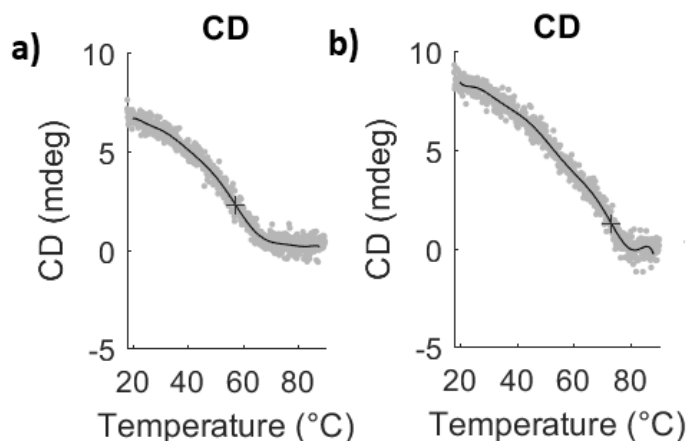


**Figure S 41:** GC-MS chromatogram (SIM mode: m/z 149, 167, 279) of compound **5b-L** performed at 200° C x 300 min. No peaks ascribable to the given piperazine-2-ones were eluted under the tested experimental conditions.

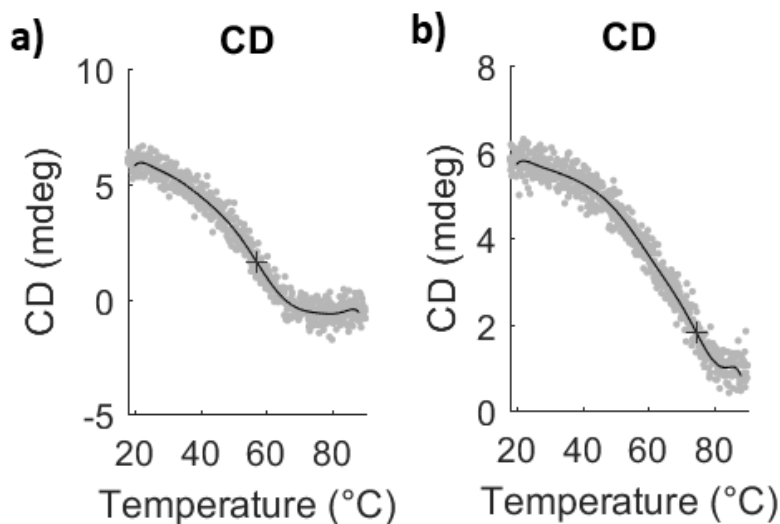
## 12 CD melting curves



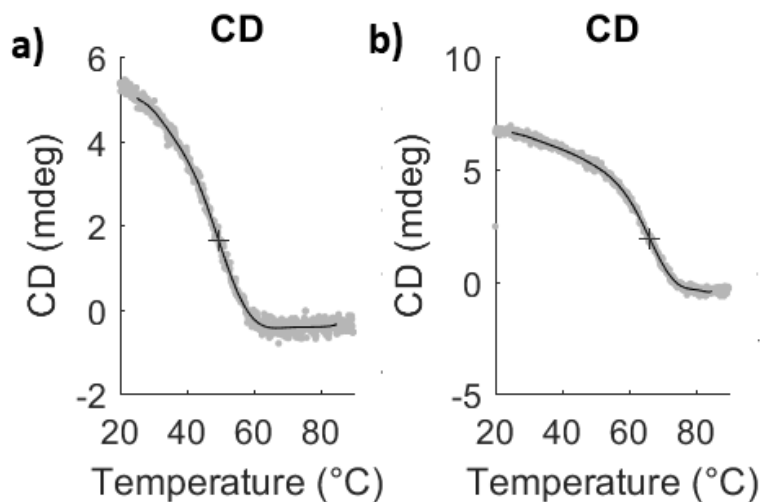
**Figure S 42:** Fitting of the experimental CD melting curves ( $\lambda = 266$  nm) for the complexes formed by **PNA1** with a) KRAS wild-type DNA (single-mismatch) and b) G12D mutated DNA (full-match).



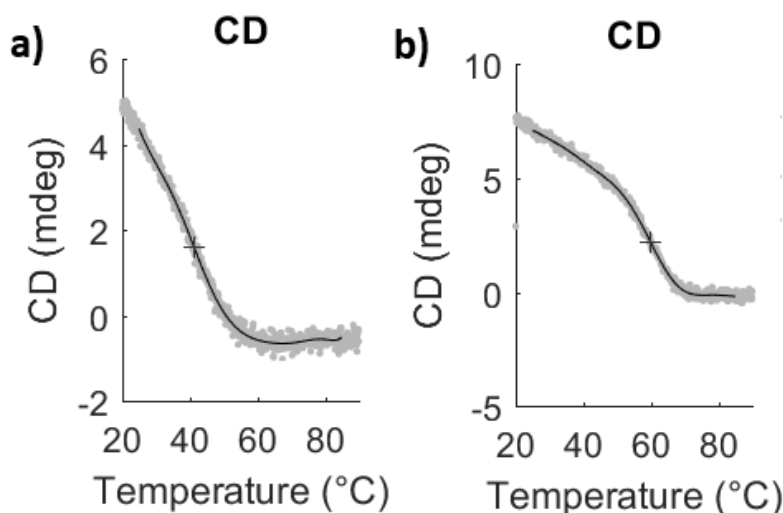
**Figure S 43:** Fitting of the experimental CD melting curves ( $\lambda = 266$  nm) for the complexes formed by **PNA2** with a) KRAS wild-type DNA (single-mismatch) and b) G12D mutated DNA (full-match).



**Figure S 44:** Fitting of the experimental CD melting curves ( $\lambda = 266$  nm) for the complexes formed by **PNA3** with a) KRAS wild-type DNA (single-mismatch) and b) G12D mutated DNA (full-match).



**Figure S 45:** Fitting of the experimental CD melting curves ( $\lambda = 266$  nm) for the complexes formed by **PNA4** with a) KRAS wild-type DNA (single-mismatch) and b) G12D mutated DNA (full-match).



**Figure S 46:** Fitting of the experimental CD melting curves ( $\lambda = 266$  nm) for the complexes formed by **PNA5** with **a)** KRAS wild-type DNA (single-mismatch) and **b)** G12D mutated DNA (full-match).

### 13 References

- (1) Tedeschi, T.; Sforza, S.; Maffei, F.; Corradini, R.; Marchelli, R. A Fmoc-Based Submonomeric Strategy for the Solid Phase Synthesis of Optically Pure Chiral PNAs. *Tetrahedron Lett.* **2008**, *49* (33), 4958–4961. <https://doi.org/10.1016/j.tetlet.2008.05.114>.
- (2) Pothukanuri, S.; Pianowski, Z.; Winssinger, N. Expanding the Scope and Orthogonality of PNA Synthesis. *European J. Org. Chem.* **2008**, No. 18, 3141–3148. <https://doi.org/10.1002/ejoc.200800141>.
- (3) Shaikh, A. Y.; Hansen, A. M.; Franzyk, H. Fmoc-Based Assembly of PNA Oligomers: Manual and Microwave-Assisted Automated Synthesis. In *Peptide Nucleic Acids: Methods and Protocols*; Nielsen, P. E., Ed.; Springer US: New York, NY, 2020; pp 1–16. [https://doi.org/10.1007/978-1-0716-0243-0\\_1](https://doi.org/10.1007/978-1-0716-0243-0_1).
- (4) Kaiser, E.; Colecott, R. L.; Bossinger, C. D.; Cook, P. I. Color Test for Detection of Free Terminal Amino Groups in the Solid-Phase Synthesis of Peptides. *Anal. Biochem.* **1970**, *34* (2), 595–598. [https://doi.org/10.1016/0003-2697\(70\)90146-6](https://doi.org/10.1016/0003-2697(70)90146-6).
- (5) Corradini, R.; Di Silvestro, G.; Sforza, S.; Palla, G.; Dossena, A.; Nielsen, P. E.; Marchelli, R. Direct Enantiomeric Separation of N-Aminoethylamino Acids: Determination of the Enantiomeric Excess of Chiral Peptide Nucleic Acids (PNAs) by GC. *Tetrahedron Asymmetry* **1999**, *10* (11), 2063–2066. [https://doi.org/10.1016/S0957-4166\(99\)00210-4](https://doi.org/10.1016/S0957-4166(99)00210-4).