Supporting Information

DNA-Compatible Nitro Reduction and Synthesis of Benzimidazoles

*Huang-Chi Du and Hongbing Huang**

Center for Drug Discovery, Department of Pathology and Immunology

Baylor College of Medicine, Houston, TX 77030

*to whom correspondence should be addressed: hongbing.huang@bcm.edu

Table of Contents
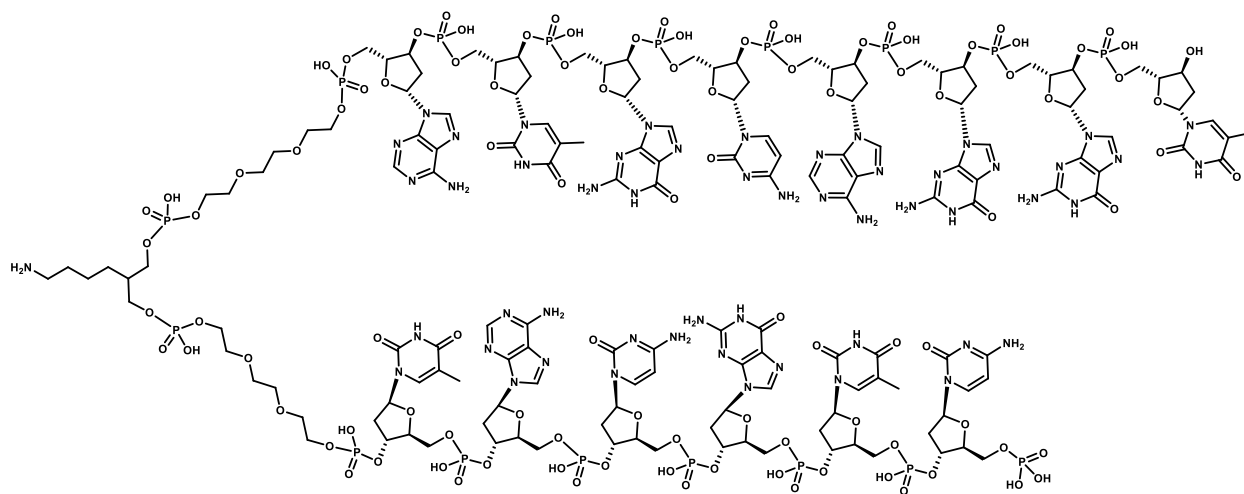
1. Structure of DNA Headpiece

Figure 1S. Structure of DNA headpiece, MW = 4992.3

2. Oligonucleotide Sequences

● The DNA sequence of headpiece (HP) : 5' d Phos-CTG CAT-Spacer 9-Amino C7- Spacer 9-ATG CAG GT 3'

● The DNA sequence for scheme 4 : 5' d Phos-TAT GAT ACT AAA GTA AGT CAC ACA CAA TTG GAG CAG TCC TGA GTG AAT ACC TGC AT -Spacer 9-Amino C7- Spacer 9-ATG CAG GTA TTC ACT CAG GAC TGC TCC AAT TGT GTG TGA CTT ACT TTA GTA TCA TAT C3')

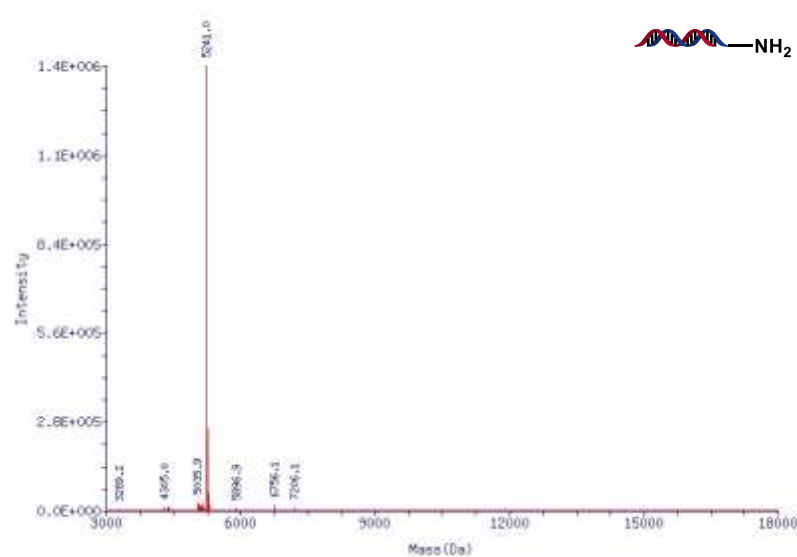# 3. Representative Mass Spectrum of DNA-Conjugates Characterization

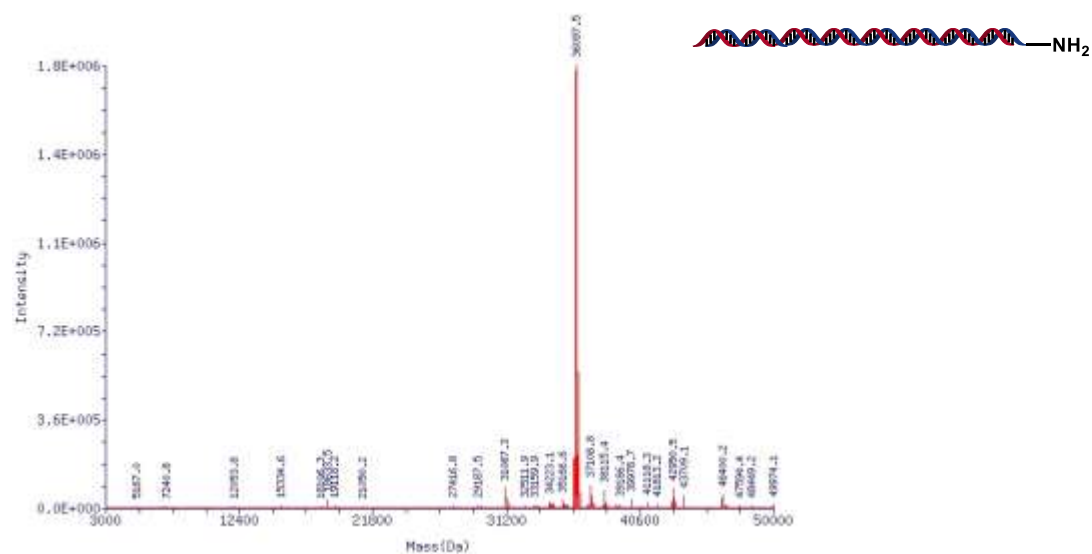**Figure S1.** Deconvoluted mass spectrum of DNA headpiece, expected: 5239.6; observed 5241.0.

Figure S2. Deconvoluted mass spectrum of the 58 b.p DNA (for scheme 4), expected: 36089.6; observed 36097.5.
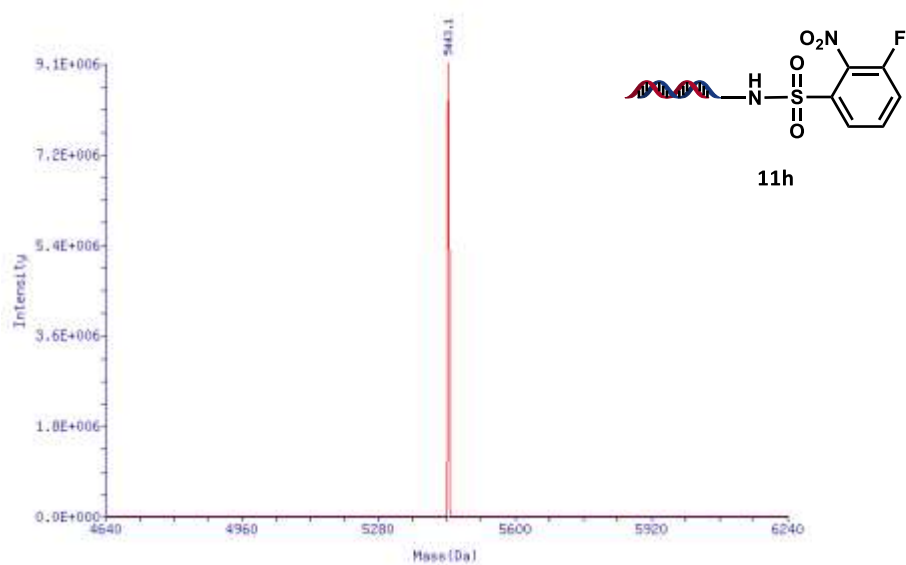
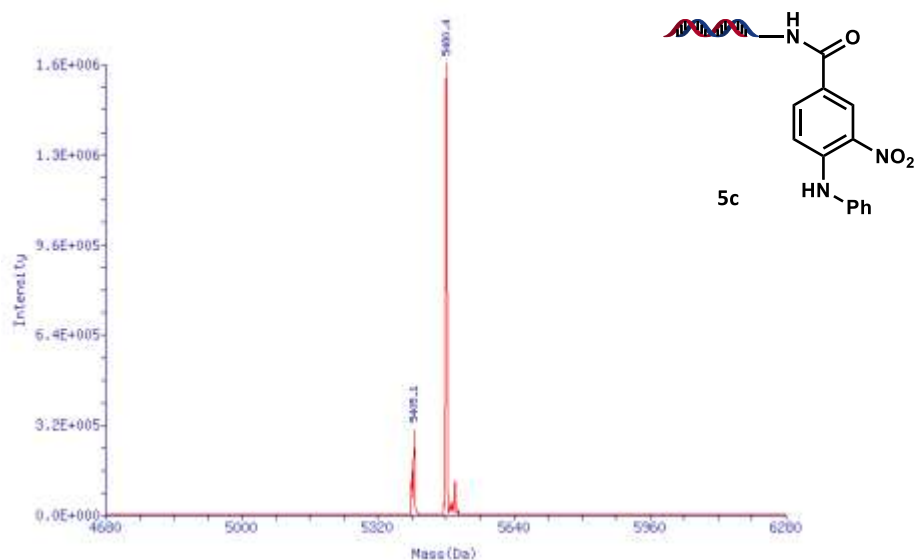**Figure S3.** Deconvoluted mass spectrum of compound **11h**, expected: 5442.7; observed 5443.1.

Figure S4. Deconvoluted mass spectrum of compound **5c**, expected: 5479.8; observed 5480.4.
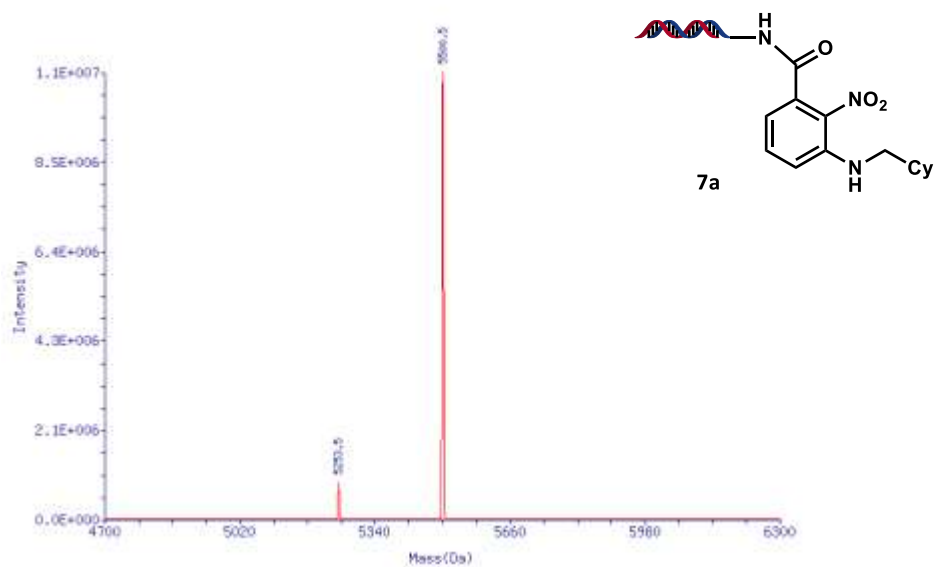
Figure S5. Deconvoluted mass spectrum of compound **7a**, expected: 5499.9; observed 5500.5.
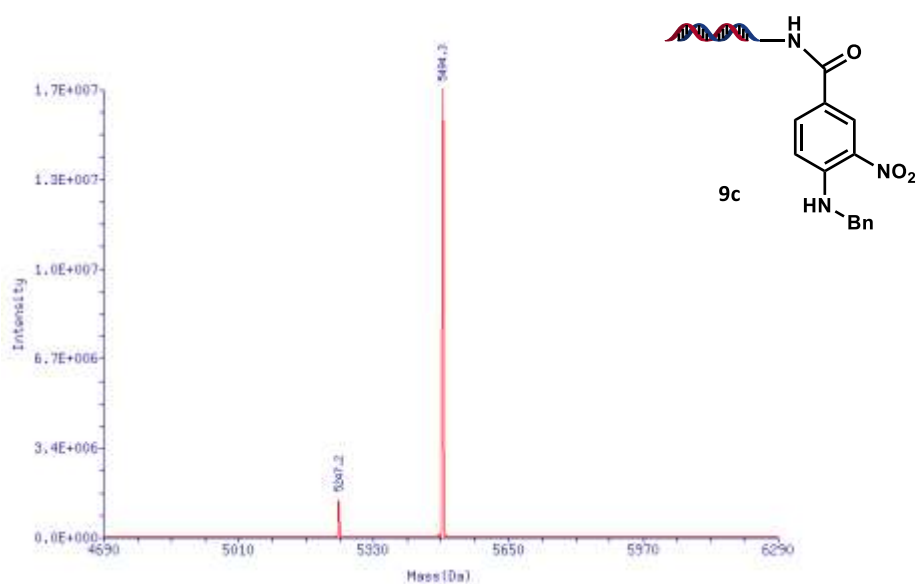

Figure S6. Deconvoluted mass spectrum of compound **9c**, expected: 5493.8; observed 5494.3.
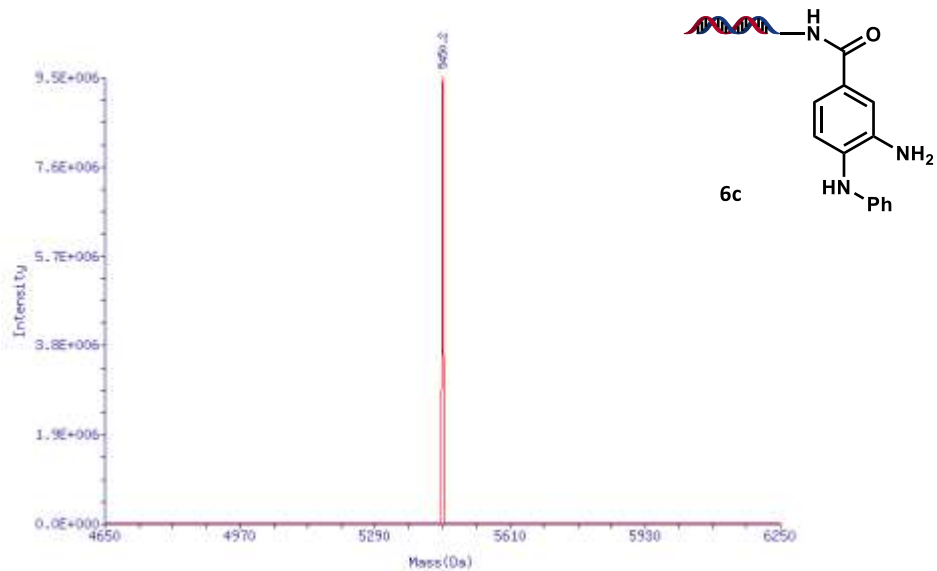
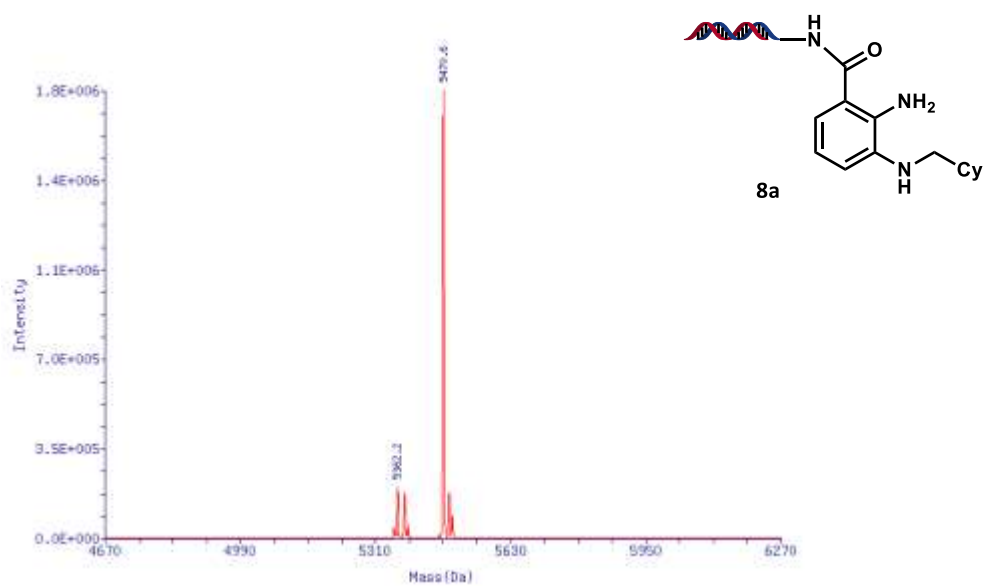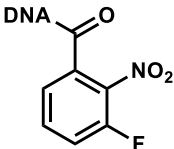**Figure S7.** Deconvoluted mass spectrum of compound **6c**, expected: 5449.8; observed 5450.2.
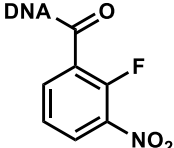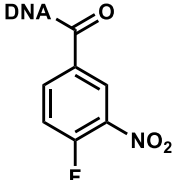

Figure S8. Deconvoluted mass spectrum of compound **8a**, expected: 5469.9; observed 5470.6.

Figure S9. Deconvoluted mass spectrum of compound **10c**, expected: 5463.8; observed 5464.8.
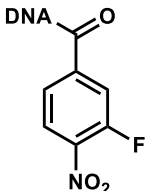

Figure S10. Deconvoluted mass spectrum of compound **19**, expected: 5610.0; observed 5611.0.

4. Preparation of Starting Material

DNA-conjugate substrates (**5, 7, 9, 11**) were prepared using general procedures for acylation and S_NAr reactions. DMT-MM, which can be hydrolyzed in S_NAr reaction, was observed in acylation step. Among them, **5h** and **9b** were unable to obtain due to an unexpected unknown product and low yield, respectively.

Table S1. Data for acylation products

Products	Structure	Expected MW	Observed MW	Conversion
11a		5406.7	5407.4	> 95%
11b		5406.7	5407.4	> 95%
11c		5406.7	5407.4	> 95%
11d		5406.7	5407.0	> 95%
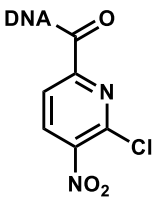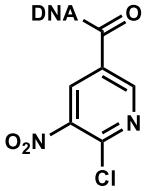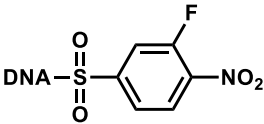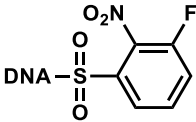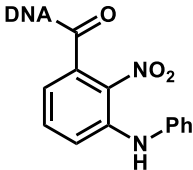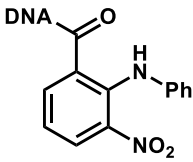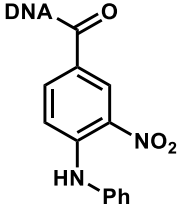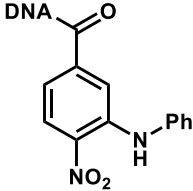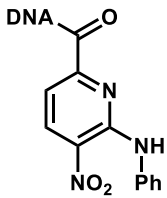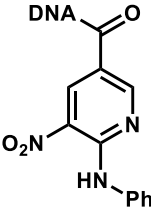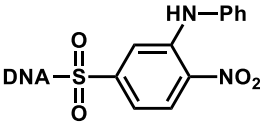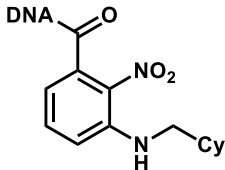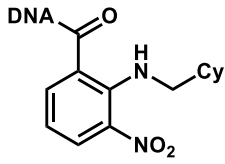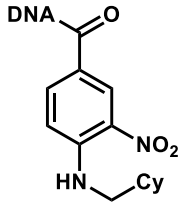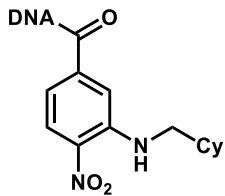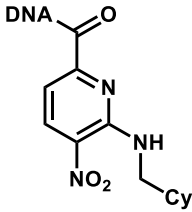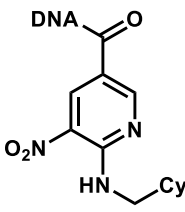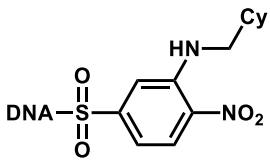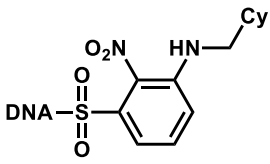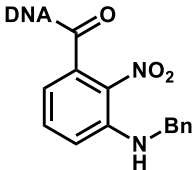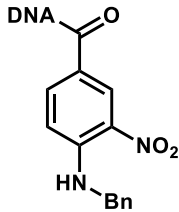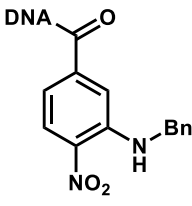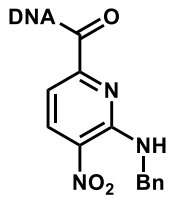
11e		5424.1	5425.0	> 95%
11f		5424.1	5424.9	> 95%
11g		5442.7	5443.1	> 95%
11h		5442.7	5443.3	> 95%
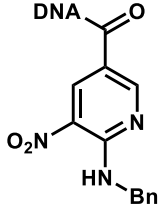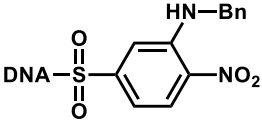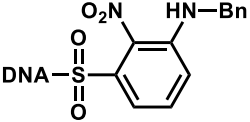
Table S2. Data for S$_N$Ar products

Products	Structure	Expected MW	Observed MW	Conversion
5a		5479.8	5480.6	45%

5b		5479.8	5480.2	67%
5c		5479.8	5480.4	88%
5d		5479.8	5480.5	74%
5e		5480.8	5481.6	77%
5f		5480.8	5481.6	83%
5g		5515.8	5516.5	77%

7a		5499.9	5500.5	> 95%
7b		5499.9	5500.5	> 95%
7c		5499.9	5500.5	> 95%
7d		5499.9	5500.5	> 95%
7e		5500.9	5501.2	> 95%
7f		5500.9	5501.3	> 95%

7g		5535.9	5536.3	> 95%
7h		5535.9	5536.4	> 95%
9a		5493.8	5494.3	> 95%
9c		5493.8	5494.3	> 95%
9d		5493.8	5494.3	> 95%
9e		5494.8	5495.0	> 95%

9f		5494.8	5495.0	> 95%
9g		5529.9	5530.1	90%
9h		5529.9	5530.1	77%

5. Analytical Data for Scheme 4

Figure S11. LC-MS spectrum of compound **38** (via one-pot synthesis)

Figure S12. Deconvoluted mass spectrum of compound **38** (via one-pot synthesis), expected: 43916.9; observed 43925.7.
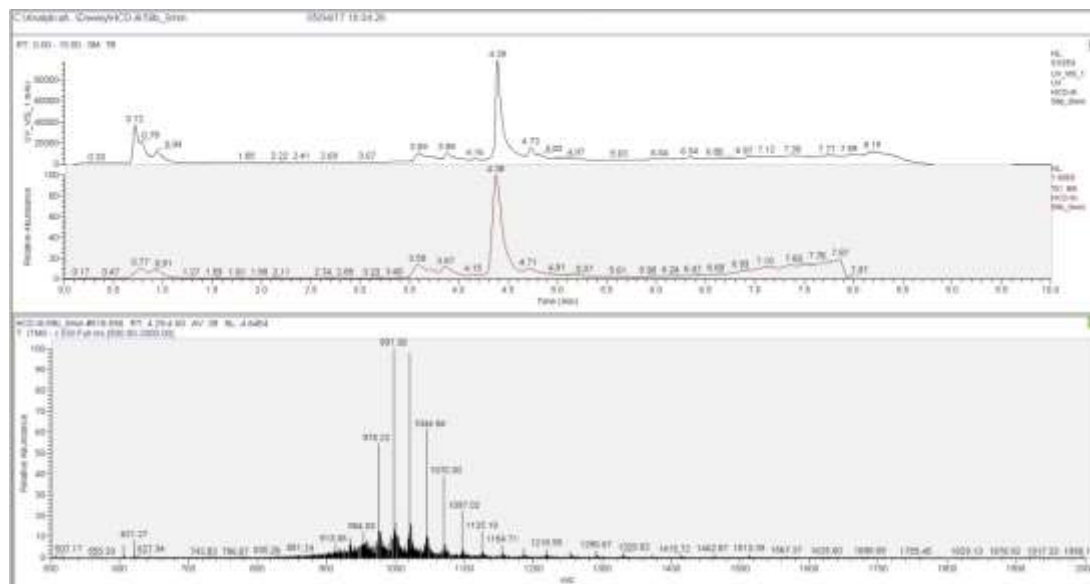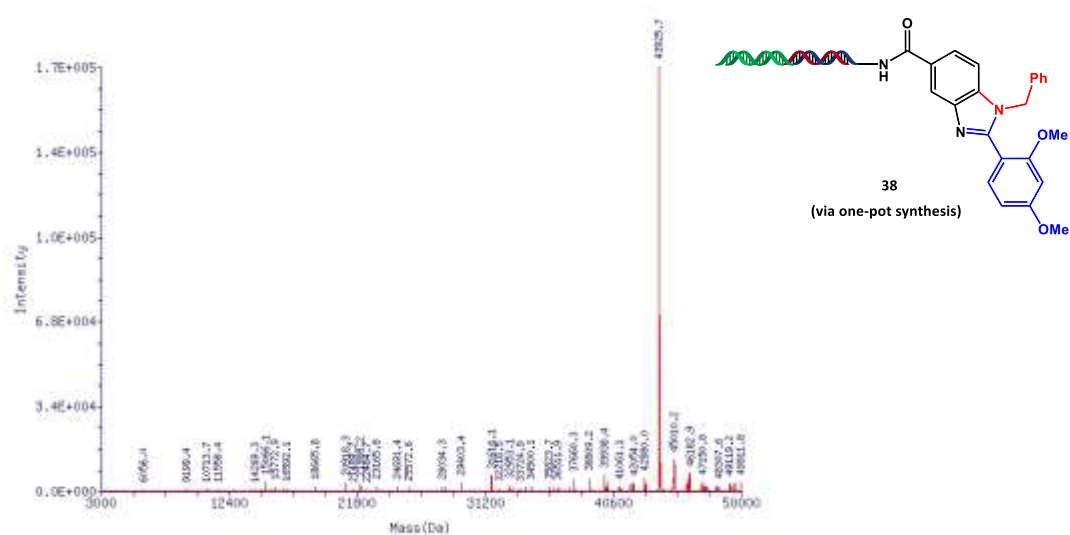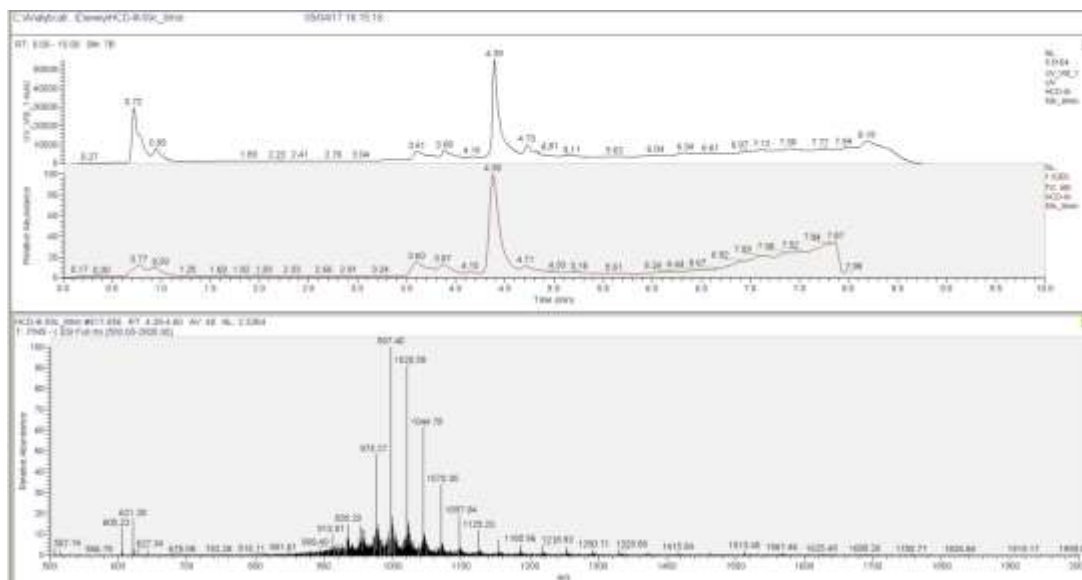
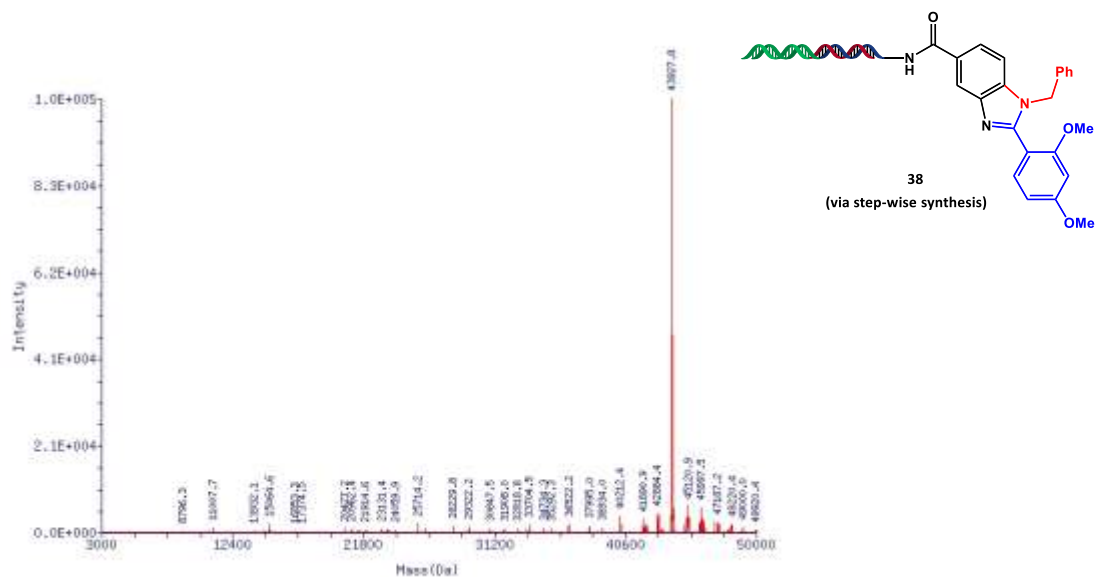**Figure S13.** LC-MS spectrum of compound **38** (via step-wise synthesis)

Figure S14. Deconvoluted mass spectrum of compound **38** (via step-wise synthesis), expected: 43916.9; observed 43927.8.
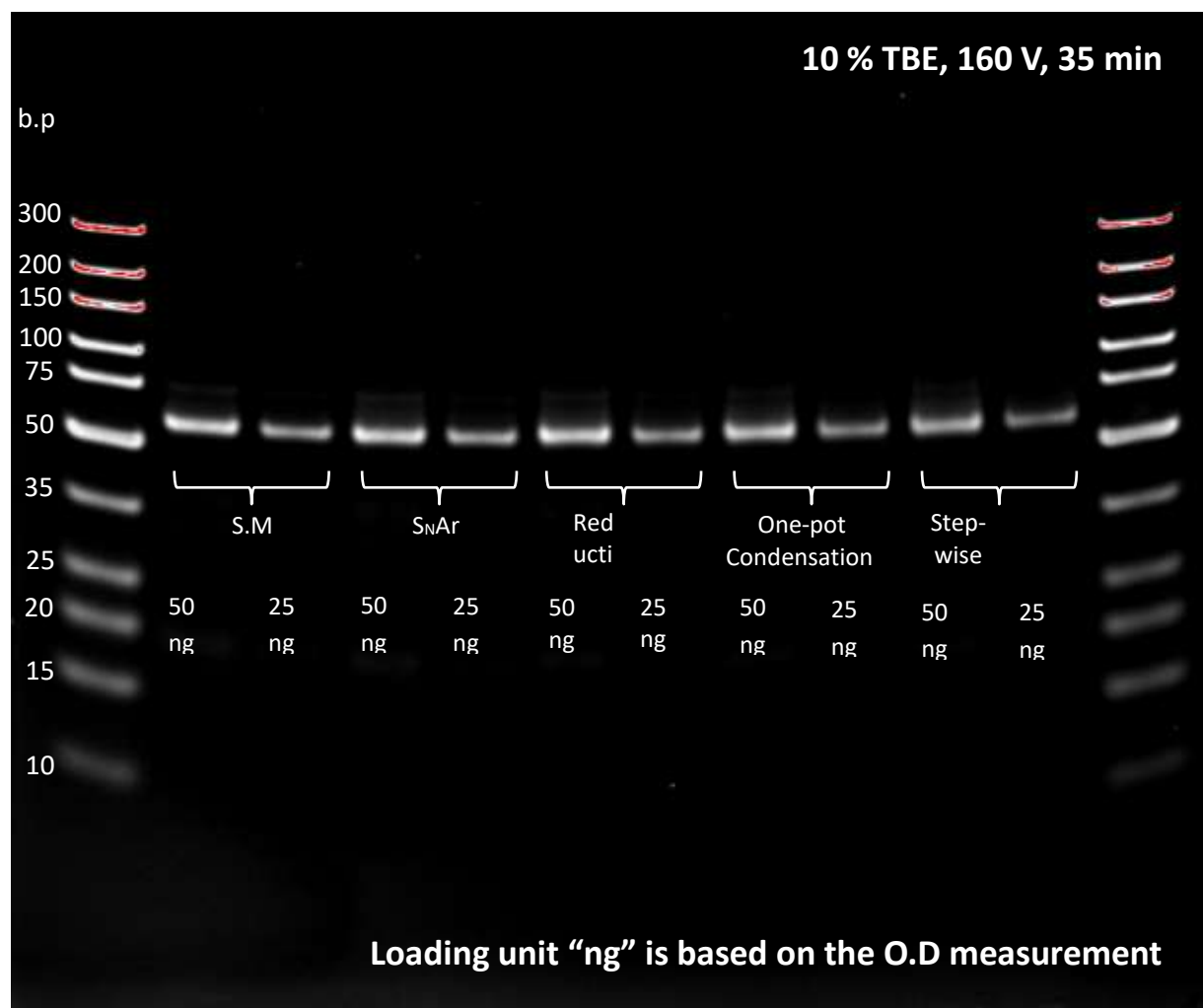
Figure S15. Gel electrophoresis image from the samples of scheme 4
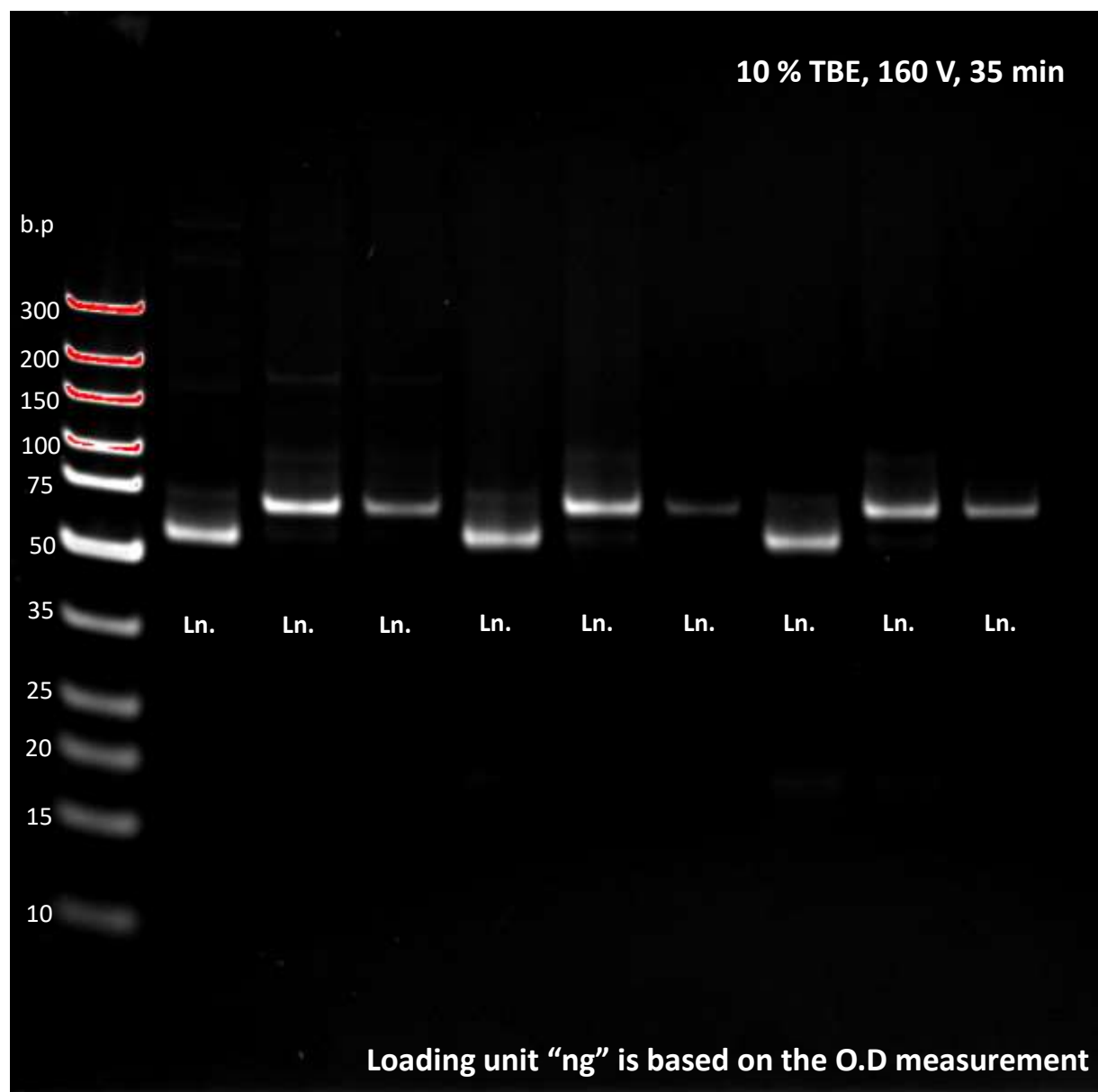
Figure S16. Gel electrophoresis image from the samples of scheme 4 (ligation validation). Lane 1. Naïve DNA (58-bp) before ligation, 50 ng; Lane 2. Naïve DNA (58-bp) after ligation, 50 ng; Lane 3. Naïve DNA (58-bp) after ligation, 25 ng; Lane 4. Compound **37** (via one-pot synthesis), 50 ng; Lane 5. Compound **38**, 50 ng; Lane 6. Compound **38**, 25 ng; Lane 7. Compound **37** (via step-wise synthesis), 50 ng; Lane 8. Compound **38**, 50 ng; Lane 9. Compound **38**, 25 ng

Figure S17. Bioanalyzer data from the samples of scheme 4

6. Sanger Sequencing for Compound 40

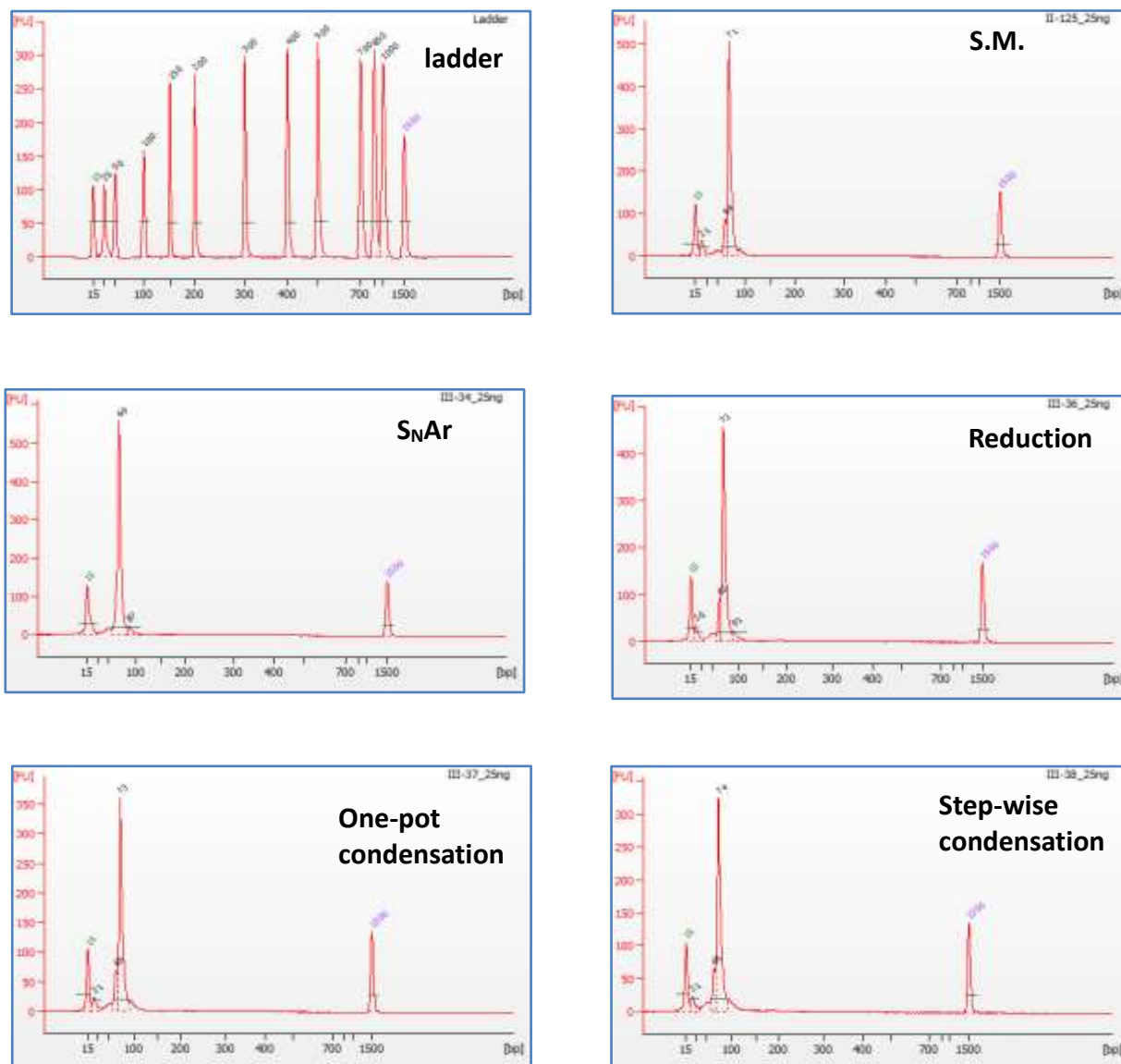**Sample preparation:** The Ampure purified product was diluted to the concentration at 1.0E7 molecules/µL. To a PCR tube was added diluted sample (1 µL), 10X highfidelity PCR buffer (10 µL), 50 mM MgSO4 (4µL), 10 mM dNTP mix (2 µL), Platinum Taq DNA Polymerase (1 µL), 10 µM forward primer (8 µL), 10 µM reverse primer (8 µL), and nuclease-free water (66 µL). The PCR was performed for 24 cycles before Ampure purification. The purified sample was quantified by O.D measurement to appear at 9 ng/µL. 5 µL of PCR product and 1 µL of primer (10 pmol/µL) were provided for Sanger sequencing, which was performed on a 3130XL genetic analyzer of applied biosystems.

18

Results:

Expected sequence: 5'-TAT GAT ACT AAA GTA AGT CAC ACA CAA TTG GAG CAG TCC TGA GTG AAT ACC TGC AT-3'

Observed sequence (red color: misread, □: missing): 5'-AAT GAT ACT AAA GTA AGTT C□C ACA CAA TTG GAG CAG TCC TGA GTTG A□T CCC TGC AT-3'