# SUPPLEMENTARY INFORMATION

## DNA-Templated Synthesis of Encoded Small Molecules by DNA Self-Assembly

Cheng Cao,[a] Peng Zhao,[a] Ze Li,[a] Zitian Chen,[c] Yanyi Huang,[c] Yu Bai,[a,*] and Xiaoyu Li[a,b,*]

[a] Key Laboratory of Bioorganic Chemistry and Molecular Engineering of the Ministry of Education, Beijing National Laboratory of Molecular Sciences (BNLMS), College of Chemistry and Molecular Engineering, Peking University, Beijing, China 100871

[b] Key Laboratory of Chemical Genomics, School of Chemical Biology and Biotechnology, Peking University Shenzhen Graduate School, Shenzhen, China 518055

[c] Biodynamic Optical Imaging Center (BIOPIC) and College of Engineering, Peking University, Beijing, China 100871

*Email: xiaoyuli@pku.edu.cn*
*yu.bai@pku.edu.cn*

# Table of Contents

1. Abbreviations…….....S3

2. Materials and General Methods.....S4

3. Preparation of Photocleavable Linkers and Derivatives.....S5

4. Preparation of the Dihydropyran Scaffold (Compound 12).....S14

5. DNA Sequences, Structures, and Preparation of Modified DNA Oligonucleotides……S15

6. Oligonucleotide Characterization.....S20

7. Experimental Conditions.....S21

8. PAGE Analysis of Enzymatic Ligation and PCR-Amplification of the Reaction Products.....S23

9. Structures of Control RD4 Reagent DNAs.....S26

10. Model Library Synthesis and *in vitro* Selection against Immobilized Streptavidin (Figure 4)...S26

11. Sanger Sequencing.....S27

12. *N*-acylthiazolidine Model Library Synthesis and *in vitro* Selection against Immobilized Streptavidin (Figure 5)……...S28

13. *N*-acylthiazolidine Model Library Synthesis and LC/MS characterization…….S30

14. References.....S32

## 1. Abbreviations.

$Ac_2O$: acetic anhydride

$Boc_2O$: di-*tert*-butyl dicarbonate

*t*-BuOH: *tert*-butanol

DIPEA: *N*, *N*'-diisopropylethylamine

DMAP: 4-dimethylaminopyridine

DCC: *N*, *N*'-dicyclohexyl carbodiimide

DCM: dichloromethane

DMF: *N*, *N*'-dimethylformamide

DMSO: dimethylsulfoxide

DMT-MM：4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methyl morpholinium chloride

dNTP: deoxy-ribonucleoside triphosphate

DTT: dithiothreitol

EDCI: 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride

3-HPA: 3-hydroxypicolinic acid

HBTU: *O*-benzotriazol-1-yl-*N, N, N', N'*-tetramethyluronium hexafluorophosphate

MeOH: methanol

MES: 2-morpholinoethanesulfonic acid

MOPS: 3-(*N*-morpholino)propanesulfonic acid

NHS: *N*-hydroxysuccinimide

PAGE: polyacrylamide gel electrophoresis

TEA: triethylamine

T4 PNK: T4 polynucleotide kinase

TFAA: trifluoroacetic anhydride

THF: tetrahydrofuran

TLC: thin layer chromatography

$TMSN_3$: azidotrimethylsilane

TMSOTf: trimethylsilyl trifluoromethanesulfonate

## 2. Materials and General Methods.

Unless otherwise noted, all reagents and solvents were purchased from commercial sources and used as received. DNA oligonucleotides were synthesized on an Applied Biosystems 394 DNA synthesizer using standard phosphoramidite protocols and purified by C18 reverse-phase HPLC with aqueous 0.1 M triethylammonium acetate (TEAA)/$CH_3CN$ gradient on Agilent 1200 HPLC systems. For all non-standard phosphoramidites, coupling time was modified to 999 seconds. Non-standard phosphoramidites were either purchased or prepared in laboratory as described below. Oligonucleotides with 3'-amino group were synthesized with the 3'-amino-modifier C7 CPG. 5'-amino groups were incorporated using the 5'-amino-modifier 5. Amine modifications on the DNA backbone were introduced with the amino-modifier dT-phosphoramidite (Wuhu Huaren). Oligonucleotides were quantitated by UV using a BioTek Epoch UV-Vis spectrometer based on extinction coefficient at 260 nm. Oligonucleotides were characterized by either a Bruker APEX IV (for ESI-MS) or a Bruker ultrafleXtreme [MALDI-MS, matrix: 8:1 (50 mg/mL 3-HPA in 1:1 water : acetonitrile) : (50 mg/mL ammonium citrate in water)] mass spectrometer. All DNA sequences are written in the 5'- to 3'- orientation. DNAs in denaturing PAGE analysis were stained with ethidium bromide and visualized on a Tanon-1600 gel image system or a BioRad Chemidoc imaging system. Water was purified with a Thermo Scientific Barnstead Nanopure system. Photocleavage experiments were conducted by a UVP CL-1000L Ultraviolet crosslinker at 365 nm wavelength with an intensity of approximately 100 $\mu J/cm^2$.

## 3. Preparation of Photocleavable Linkers and Derivatives.[S1-S3]

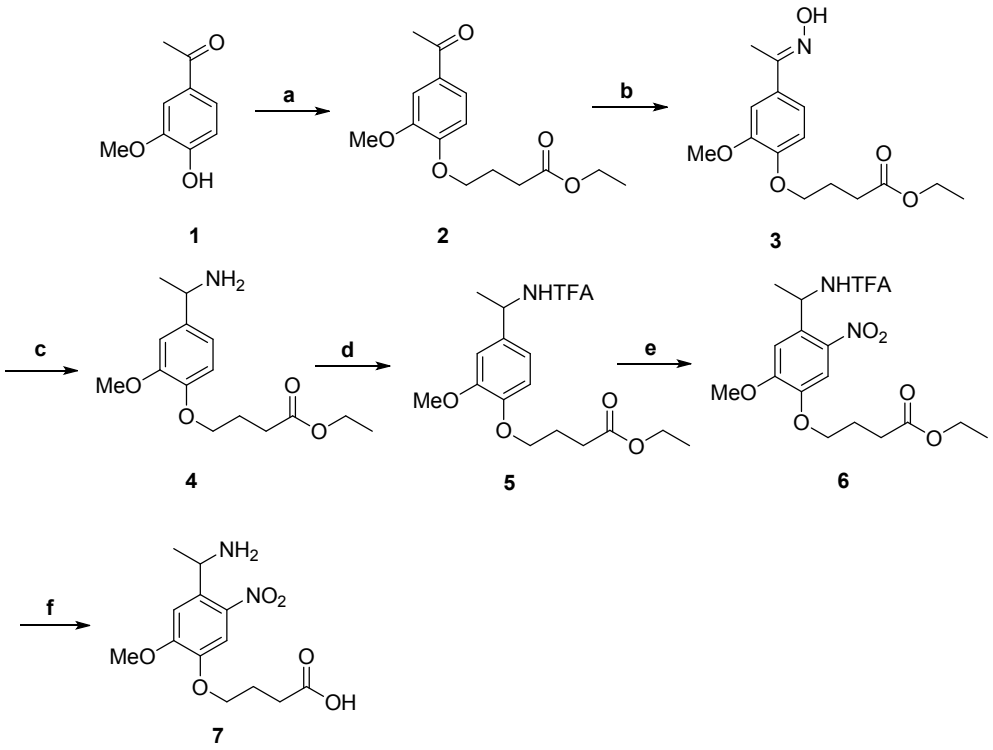
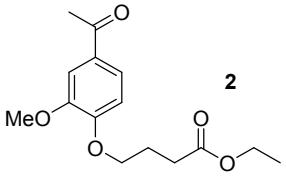


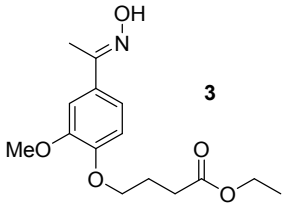

**Figure S1. a**: ethyl 4-bromobutanoate, $K_2CO_3$, DMF, r.t., 89%; **b**: hydroxyl ammonium chloride, pyridine/$H_2O$, r.t., 88%; c: Pd/C, $H_2$, AcOH, r.t., 70%; **d**: TFAA, pyridine, 0 ºC, 91%; **e**: $HNO_3$, $Ac_2O$, 0 ºC, 83%; **f**: NaOH, MeOH, reflux, 94%. See detailed synthesis procedures below.

The preparation of photocleavable linkers followed a report by Holmes and co-workers (Figure S1)[S2] and is also described in details below:



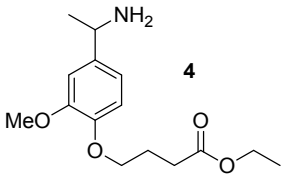

Under nitrogen atmosphere, a mixture of compound **1** (9.06 g, 54.20 mmol) and ethyl 4-bromobutanoate (12.77 g, 65.04 mmol) in 100 mL of DMF was stirred at room temperature overnight. Solvent was removed by rotary evaporation and 400 mL of water was added. The solution was extracted by ethyl acetate (3x 100 mL) and the organic layer was dried over $Na_2SO_4$ and evaporated *in vacuo*. The crude product was purified by flash column chromatography with petroleum ether/ethyl acetate (10:1 to 5:1) to afford ester **2** (13.95 g, 89%) as a white solid. $^1$H NMR (400 MHz, $CDCl_3$) δ 7.55 (dd, *J* = 8.3, 2.0 Hz, 1H), 7.52 (d, *J* = 1.9 Hz, 1H), 6.90 (d, *J* = 8.3 Hz, 1H), 4.21 – 4.06 (m, 4H), 3.91 (s, 3H), 2.62 – 2.47 (m, 5H), 2.19 (p, *J* = 6.8 Hz, 2H), 1.26 (t, *J* = 7.1 Hz, 3H). $^{13}$C NMR

(100 MHz, $CDCl_3$) δ 196.79, 173.01, 152.69, 149.34, 130.57, 123.21, 111.37, 110.58, 67.85, 60.49, 56.01, 30.61, 26.19, 24.35, 14.23. ESI-HRMS m/z 281.13793 $[M+H]^+$; $C_{15}H_{21}O_5$ requires 281.13835.




**3**

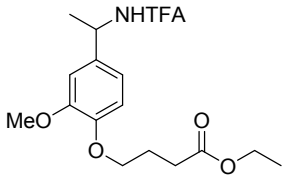
Under nitrogen atmosphere, a mixture of ester **2** (18.80 g, 67.14 mmol) and hydroxyl ammonium chloride (7.00 g, 100.07 mmol) in 100 mL of pyridine and 50 mL of water was stirred at room temperature overnight. The solvent was removed by rotary evaporation and 200 mL of water was added. The solution was extracted by ethyl acetate (3x 100 mL) and the organic layer was washed by brine (100 mL). The organic layer was dried over $Na_2SO_4$ and evaporated *in vacuo*. Then 50 mL of toluene was added and the solution was evaporated to remove pyridine. The crude product was purified by flash column chromatography with petroleum ether/ethyl acetate (7:1 to 4:1) to afford the oxime product **3** (17.49 g, 88%) as a white solid. $^1$H NMR (400 MHz, $CDCl_3$) δ 9.15 (s, 1H), 7.25 (d, $J$ = 2.0 Hz, 1H), 7.12 (dd, $J$ = 8.4, 2.0 Hz, 1H), 6.87 (d, $J$ = 8.4 Hz, 1H), 4.15 (q, $J$ = 7.1 Hz, 2H), 4.08 (t, $J$ = 6.3 Hz, 2H), 3.88 (s, 4H), 2.54 (t, $J$ = 7.3 Hz, 2H), 2.27 (s, 4H), 2.16 (p, $J$ = 6.8 Hz, 2H), 1.25 (t, $J$ = 7.1 Hz, 4H); $^{13}$C NMR (100 MHz, $CDCl_3$) δ 173.25, 155.52, 149.46, 149.33, 129.50, 119.21, 112.52, 109.22, 67.87, 60.46, 55.95, 30.71, 24.47, 14.20, 12.11. ESI-HRMS m/z 296.14897 $[M+H]^+$; $C_{15}H_{22}NO_5$ requires 296.14925.




**4**

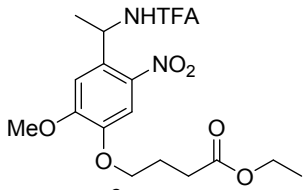
With a hydrogen balloon, a mixture of compound **3** (15.00 g, 50.50 mmol) and Pd/C (5% - 10%, 3.00 g) in 150 mL of dry acetate acid was stirred at room temperature for 48 hours. The solvent was removed by rotary evaporation and 200 mL of water was added. The solution was extracted by ethyl ether (2x 100 mL). Then the inorganic layer was basified to pH > 10 by saturated $Na_2CO_3$ solution and extracted by ethyl acetate (3x 100 mL), the combined organic layer was washed by brine (100 mL), dried over $Na_2SO_4$ and evaporated *in vacuo*. The crude product was purified by flash column chromatography with ethyl acetate/methanol (100 : 1 with 1 % TEA) to

furnish amine **4** (10.39 g, 70%) as a colorless oil. $^1$H NMR (400 MHz, $CDCl_3$) δ 6.92 (s, 1H), 6.84 (d, *J* = 0.9 Hz, 2H), 4.13 (q, *J* = 7.1 Hz, 2H), 4.06 (dt, *J* = 12.8, 6.5 Hz, 3H), 3.87 (s, 3H), 2.53 (t, *J* = 7.3 Hz, 2H), 2.13 (p, *J* = 6.7 Hz, 2H), 1.68 (s, 2H), 1.36 (d, *J* = 6.6 Hz, 3H), 1.25 (t, *J* = 7.1 Hz, 3H). $^{13}$C NMR (100 MHz, $CDCl_3$) δ 173.15, 149.60, 147.11, 140.84, 117.69, 113.48, 109.68, 68.10, 60.30, 55.95, 50.98, 30.73, 25.73, 24.60, 14.19. ESI-HRMS m/z 304.15257 $[M+Na]^+$; $C_{15}H_{23}NNaO_4$ requires 304.15193.




**5**

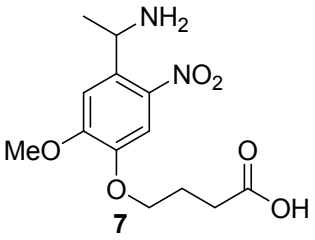
Under nitrogen atmosphere, amine **4** (10.40 g, 37.01 mmol) in 100 mL of pyridine was stirred at 0 °C. TFAA (10.0 mL, 71.43 mmol) was added to the solution drop-wise and the solution was stirred at 0 °C for 2 hours. The solvent was removed by rotary evaporation and 200 mL of water was added. Then the solution was extracted by ethyl acetate (3x 100 mL). The organic layer was washed by brine (100 mL), dried over $Na_2SO_4$ and evaporated *in vacuo*. 50 mL of toluene was added and the solution was evaporated to remove pyridine. The crude product was purified by flash column chromatography with petroleum ether/ethyl acetate (5:1 to 3:1) to furnish the amide product **5** (12.65 g, 91%) as a white solid. $^1$H NMR (400 MHz, $CDCl_3$) $^1$H NMR (400 MHz, $CDCl_3$) δ 6.88 – 6.84 (m, 3H), 5.08 (p, *J* = 7.0 Hz, 1H), 4.17 – 4.07 (m, 2H), 4.07 – 3.99 (m, 2H), 3.84 (s, 3H), 2.52 (t, *J* = 7.3 Hz, 2H), 2.12 (p, *J* = 6.7 Hz, 2H), 1.56 (d, *J* = 6.9 Hz, 2H), 1.25 (t, *J* = 7.1 Hz, 3H). $^{13}$C NMR (100 MHz, $CDCl_3$) δ 173.20, 156.83-155.73(q, J = 36.3 Hz), 149.77, 148.16, 133.94, 120.16-111.27 (q, J = 286 Hz), 118.29, 113.47, 110.54, 68.00, 60.42, 55.99, 49.50, 30.70, 24.50, 20.76, 14.15. ESI-HRMS m/z 378.15257 $[M+H]^+$; $C_{17}H_{23}F_3NO_5$ requires 378.15228.




**6**

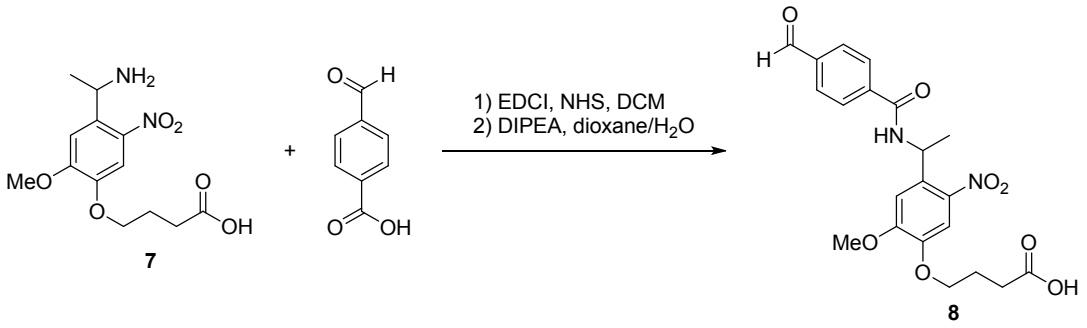
To the solution of 100 mL of $HNO_3$ (70 %) and 30 mL of $Ac_2O$, compound **5** (12.4 g, 32.9 mmol) in 20 mL of $Ac_2O$ was added drop-wise at 0 °C and the solution was stirred at 0 °C for 2 hours. The solution was poured to ice (600 mL) and extracted by ethyl acetate (3x 200 mL). Combined organic layers were washed by brine (100 mL),

dried over $Na_2SO_4$ and evaporated *in vacuo*. The crude product was purified by flash column chromatography with petroleum ether/ethyl acetate (10:1 to 3:1) to furnish the nitrate amide product **6** (12.65 g, 91%) as an off-white solid. $^1$H NMR (400 MHz, $CDCl_3$) δ 7.61 (s, 1H), 7.48 (d, $J$ = 7.5 Hz, 1H), 6.89 (s, 1H), 5.55 (p, $J$ = 7.1 Hz, 1H), 4.21 – 4.08 (m, 4H), 3.96 (d, $J$ = 14.1 Hz, 3H), 2.54 (t, $J$ = 7.2 Hz, 2H), 2.18 (p, $J$ = 6.7 Hz, 2H), 1.62 (d, $J$ = 7.0 Hz, 3H), 1.27 (t, $J$ = 7.1 Hz, 3H). $^{13}$C NMR (100 MHz, $CDCl_3$) δ 172.88, 156.61(q, $J$ = 36.4 Hz), 154.08, 147.67, 140.58, 131.03, 120.17-111.23(q, $J$ = 235 Hz), 110.99, 110.29, 68.38, 60.56, 56.38, 48.36, 30.56, 24.25, 20.16, 14.18. ESI-HRMS m/z 423.13769 $[M+H]^+$; $C_{17}H_{22}F_3N_2O_7$ requires 423.13736.




**7**

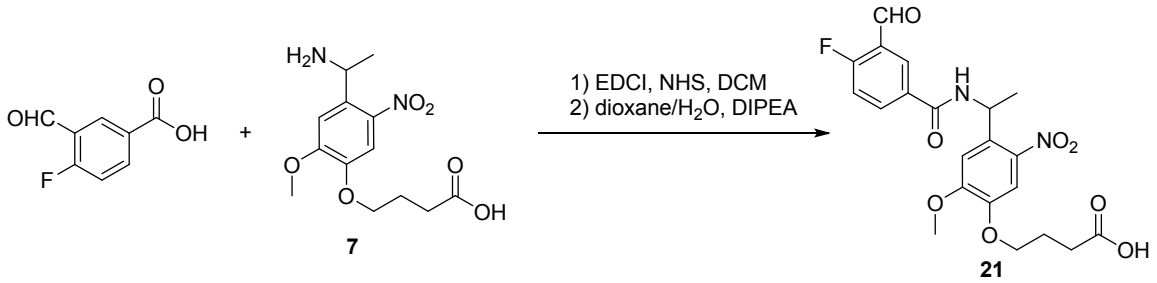
Compound **6** (5.0 g, 11.8 mmol) was dissolved in warm MeOH (100 mL) and aqueous NaOH solution (40 mL, 1 M, 40 mmol) at 40 °C. The resulting solution was heated at reflux for 6 hours, cooled to room temperature, and stirred for 48 hours. The crude mixture was concentrated and poured over a plug of silica (5 cm × 7 cm) and eluted with MeOH. The eluent was concentrated to afford compound **7** (3.30 g, 94%) as an off-white solid. $^1$H NMR (400 MHz, MeOD + $K_2CO_3$) δ 7.50 (s, 1H), 7.29 (s, 1H), 4.63 (q, $J$ = 6.6 Hz, 1H), 4.07 (t, $J$ = 6.6 Hz, 2H), 3.96 (d, $J$ = 3.6 Hz, 3H), 2.35 (t, $J$ = 7.4 Hz, 2H), 2.15 – 2.02 (m, 2H), 1.49 – 1.41 (m, 3H). $^{13}$C NMR (100 MHz, MeOD + $K_2CO_3$) δ 181.82, 155.33, 148.48, 142.12, 137.60, 110.31, 110.27, 70.45, 56.91, 47.23, 35.23, 27.19, 24.28. ESI-HRMS m/z 299.12394 $[M+H]^+$; $C_{13}H_{19}N_2O_6$ requires 299.12376.



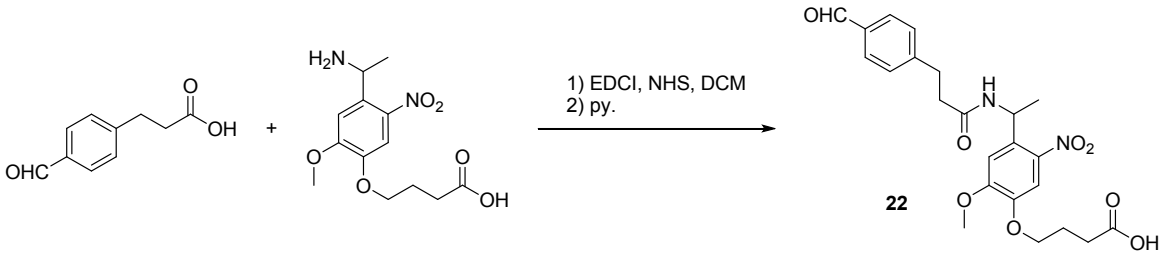

**7** **8**

4-formylbenzoic acid (1.20 g), EDCI (3.04 g), and NHS (1.84 g) were mixed in DCM (50 mL) at room temperature for 3 h, and the *N*-hydroxysuccinimide ester product was purified by flash column chromatography

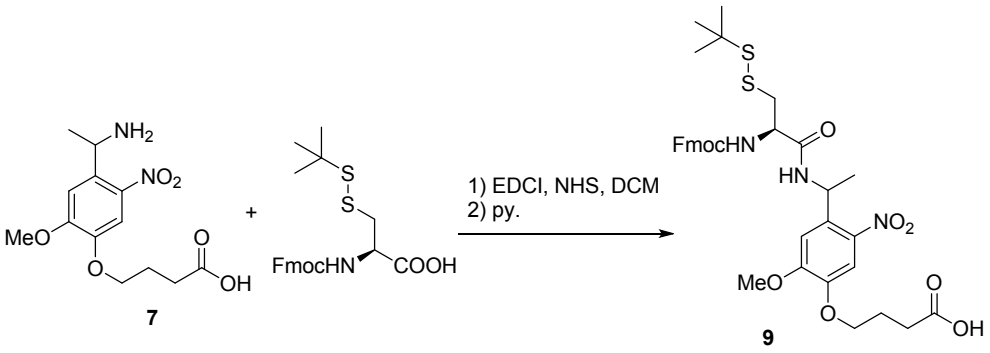
with petroleum ether/ethyl acetate (1:1) (1.95 g, 98 %). Next, the *N*-hydroxysuccinimide ester (50 mg), photocleavable linker **7** (60 mg), and DIPEA(100 mg) were mixed in 5.0 mL of dioxane and 2.5 mL of water at room temperature for 1 h. The solution was acidified to pH < 2 by 1 M HCl and extracted by ethyl acetate three times. The organic layer was dried over $Na_2SO_4$ and then concentrated. The crude product was purified by flash column chromatography with petroleum ether/ethyl acetate (2:1 to 2:3 with 1 % AcOH) to furnish **8** (57 mg). $^1$H NMR (400 MHz, DMSO-*d6*) δ 12.17 (s, 1H), 10.30 – 9.90 (m, 1H), 9.20 (t, *J* = 16.0 Hz, 1H), 8.08 – 7.96 (m, 4H), 7.52 (s, 1H), 7.37 (s, 1H), 5.65 (p, *J* = 6.9 Hz, 1H), 4.11 – 4.02 (m, 2H), 3.88 (s, 3H), 2.39 (t, *J* = 7.3 Hz, 2H), 2.03 – 1.89 (m, 2H), 1.59 (d, *J* = 6.9 Hz, 3H); $^{13}$C NMR (100 MHz, DMSO-*d6*) δ 192.82, 173.94, 165.02, 153.39, 146.34, 140.31, 139.18, 137.85, 134.87, 129.34, 128.07, 109.74, 108.30, 67.92, 56.25, 44.71, 29.92, 24.01, 21.44. ESI-HRMS m/z 431.14529 $[M+H]^+$; $C_{21}H_{23}N_2O_8$ requires 431.14489.




Compound **21** was synthesized with the same procedure as compound **8**, except 4-fluoro-3-formylbenzoic acid was used instead of the 4-formylbenzoic acid. $^1$H NMR (400 MHz, DMSO-*d6*) δ 12.11 (s, 1H), 10.25 (s, 1H), 9.25 (d, *J* = 7.5 Hz, 1H), 8.37 (dd, *J* = 6.7, 2.4 Hz, 1H), 8.23 – 8.16 (m, 1H), 7.55 – 7.49 (m, 1H), 7.34 (s, 1H), 5.62 (p, *J* = 6.9 Hz, 1H), 4.06 (t, *J* = 6.4 Hz, 2H), 3.87 (s, 2H), 2.38 (t, *J* = 7.3 Hz, 2H), 1.99 – 1.89 (m, 2H), 1.59 (d, *J* = 6.9 Hz, 2H). $^{13}$C NMR (101 MHz, DMSO-*d6*) δ 187.50, 173.94, 163.88, 153.34, 146.32, 140.32, 135.87, 135.77, 134.89, 130.92, 128.48, 123.37, 117.10, 116.89, 109.72, 108.24, 67.90, 56.23, 44.74, 29.90, 23.98, 21.40. ESI-HRMS m/z 466.16306 $[M+H]^+$; $C_{21}H_{25}FN_3O_8$ requires 466.16202.

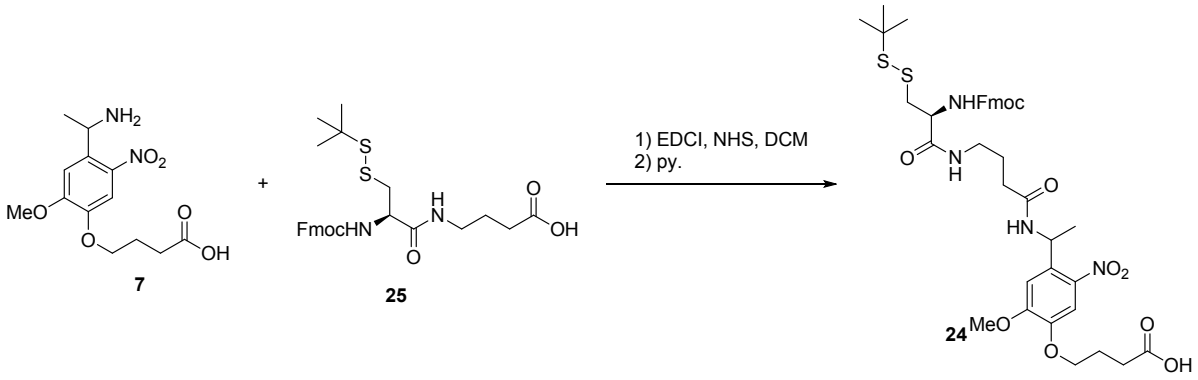



Compound **22** was synthesized with the same procedure as compound **8** except that 3-(4-formylphenyl)propanoic acid was used. $^{1}$H NMR (500 MHz, DMSO-*d6*) δ 12.12 (s, 1H), 9.93 (s, 1H), 8.49 (d, *J* = 7.8 Hz, 1H), 7.73 (t, *J* = 9.1 Hz, 2H), 7.47 (s, 1H), 7.34 (d, *J* = 8.0 Hz, 2H), 7.04 (s, 1H), 5.37 (p, *J* = 6.9 Hz, 1H), 4.05 (t, *J* = 6.5 Hz, 2H), 3.79 (s, 2H), 2.86 (td, *J* = 7.4, 3.3 Hz, 1H), 2.44 (td, *J* = 7.4, 3.2 Hz, 2H), 2.39 (t, *J* = 7.3 Hz, 2H), 1.95 (t, *J* = 6.9 Hz, 1H), 1.37 (d, *J* = 6.9 Hz, 2H). $^{13}$C NMR (126 MHz, DMSO-*d6*) δ 192.47, 173.95, 170.25, 153.30, 148.55, 146.20, 140.05, 135.26, 134.31, 129.40, 128.98, 109.45, 108.27, 67.92, 56.04, 43.73, 36.13, 30.92, 29.93, 24.03, 21.70. ESI-HRMS m/z 459.17720 $[M+H]^{+}$; $C_{23}H_{27}N_2O_8$ requires 459.17619.

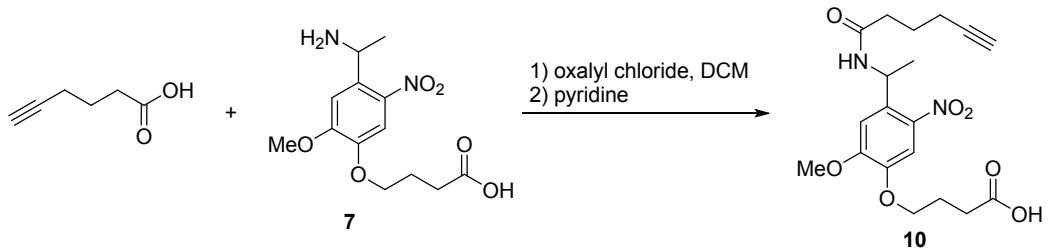



Fmoc-Cys(S-*t*Bu)-OH (215 mg), EDCI (192 mg), and NHS (115 mg) were mixed in DCM (10 mL) at room temperature for 3 h, then the *N*-hydroxysuccinimide ester product was purified by flash column chromatography with petroleum ether/ ethyl acetate (4:1 to 1:1) (233 mg, 88%). The *N*-hydroxysuccinimide ester (106 mg) was dissolved in pyridine (5 mL) and then added to a solution of **7** (60 mg, 0.2 mmol) in pyridine (5 mL). After stirring overnight, the solution was concentrated and partitioned by ethyl acetate (20 mL) and 1 M HCl (30 mL). The solution was extracted by ethyl acetate (3x 10 mL). The organic layer was dried over $Na_2SO_4$ and concentrated by rotary evaporation to give the product **9** (124 mg, 87%). $^{1}$H NMR (major diastereomer, 400 MHz, DMSO-*d6*) δ 8.88 (dd, *J* = 7.0, 4.1 Hz, 1H), 7.89 (t, *J* = 6.8 Hz, 2H), 7.74 (t, *J* = 6.4 Hz, 1H), 7.68 (t, *J* = 7.3 Hz, 1H), 7.51 (d, *J* = 4.9 Hz, 1H), 7.46 – 7.38 (m, 2H), 7.37 – 7.26 (m, 2H), 7.21 (d, *J* = 9.5 Hz, 1H), 6.28 (s, 1H), 5.39 (dd, *J* = 16.1, 7.1 Hz, 1H), 4.32 – 4.14 (m, 4H), 4.05 (t, *J* = 6.5 Hz, 2H), 3.89 (d, *J* = 11.0 Hz, 3H), 3.03 – 2.81 (m, 2H), 2.68 (q, *J*

= 7.2 Hz, 1H), 2.37 (t, *J* = 7.1 Hz, 2H), 1.95 (dt, *J* = 12.2, 6.3 Hz, 2H), 1.43 (dd, *J* = 8.9, 7.2 Hz, 2H), 1.29-1.26 (m, 9H), 1.02 (t, *J* = 7.2 Hz, 1H); $^{13}$C NMR (100 MHz, DMSO-*d6*) δ 174.06, 169.26, 155.81, 153.36, 146.21, 143.72, 142.55, 140.67, 139.93, 137.40, 128.88, 127.58, 125.34, 121.33, 120.04, 109.64, 108.29, 67.93, 56.04, 47.72, 46.58, 45.49, 30.08, 29.52, 24.09, 21.56, 10.40. ESI-HRMS m/z 712.23772 $[M+H]^+$; $C_{35}H_{42}N_3O_9S_2$ requires 712.23570.

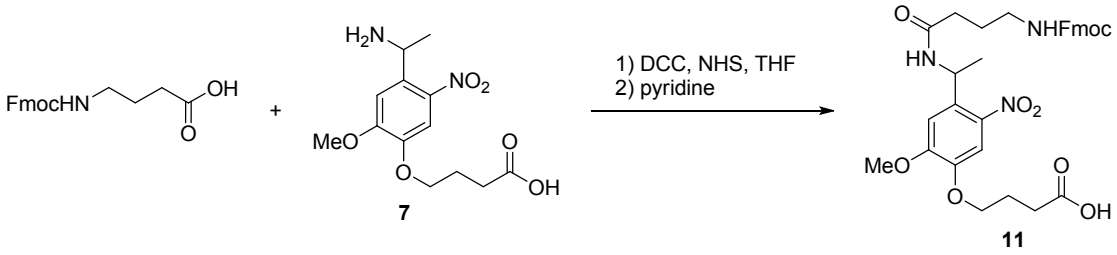



Compound **24** (diastereomers) was synthesized with the same procedure as compound **9** but with compound **25** as the carboxylic acid. $^1$H NMR (400 MHz, DMSO-*d6*) δ 8.47 (d, *J* = 7.3 Hz, 1H), 7.89 (d, *J* = 7.5 Hz, 1H), 7.73 (d, *J* = 7.3 Hz, 2H), 7.65 (d, *J* = 8.3 Hz, 1H), 7.49 (s, 1H), 7.41 (t, *J* = 7.3 Hz, 2H), 7.32 (t, *J* = 7.1 Hz, 2H), 7.16-7.10 (m, 1H), 5.45 – 5.12 (m, 1H), 4.26 – 4.17 (m, 4H), 4.07 – 4.01 (m, 2H), 3.88 (d, *J* = 3.4 Hz, 3H), 3.16 (d, *J* = 5.2 Hz, 2H), 3.09 – 2.96 (m, 4H), 2.37 (ddd, *J* = 16.2, 8.6, 2.8 Hz, 4H), 2.14 – 2.03 (m, 3H), 1.39 (d, *J* = 6.9 Hz, 2H), 1.36 (d, *J* = 6.9 Hz, 2H), 1.29 (t, *J* = 2.2 Hz, 9H), 1.17 (dd, *J* = 14.0, 7.0 Hz, 2H).ESI-HRMS m/z 797.28664 $[M+H]^+$; $C_{39}H_{49}N_4O_{10}S_2$ requires 797.28846.

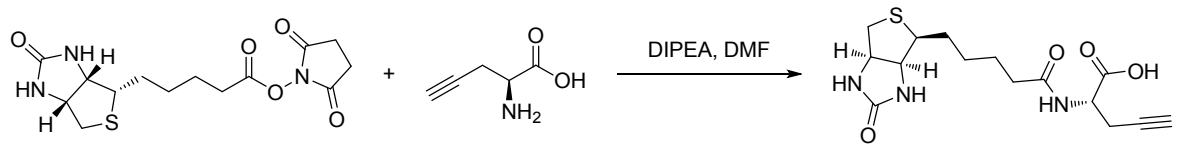



Hex-5-ynoic acid (342 mg, 3.0 mmol) and oxalyl chloride (1.0 mL, excess) were stirred with a drop of DMF in DCM (10 mL) at 40 ºC for 30 minutes. Then the solution was concentrated to obtain the hex-5-ynoyl chloride by rotary evaporation. The above product was dissolved with DCM before a solution of **7** (160 mg, 0.5 mmol) in pyridine (15 mL) was added. After stirring overnight, the solution was concentrated and partitioned by ethyl acetate (20 mL) and 1 M HCl (30 mL). The organic layer was washed by brine (100 mL), dried over $Na_2SO_4$ and

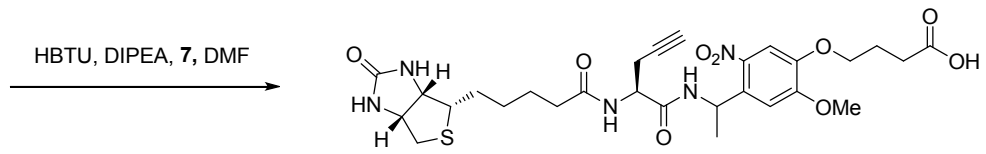
evaporated *in vacuo*. The crude product was purified by flash column chromatography with petroleum ether/ethyl acetate (1:1 to 1:4) to furnish the product **10** (200 mg, 95%) as a white solid.[1]H NMR (400 MHz, DMSO-*d6*) δ 12.22 (s, 1H), 8.55 (d, *J* = 7.6 Hz, 1H), 7.56 – 7.39 (m, 1H), 7.25 – 7.02 (m, 1H), 5.36 (p, *J* = 6.9 Hz, 1H), 4.10 – 3.99 (m, 2H), 3.89 (s, 3H), 2.83 – 2.72 (m, 1H), 2.38 (t, *J* = 7.3 Hz, 2H), 2.25 – 2.14 (m, 2H), 2.13 – 2.03 (m, 2H), 1.99 – 1.89 (m, 2H), 1.67 – 1.55 (m, 2H), 1.39 (t, *J* = 8.1 Hz, 3H). [13]C NMR (100 MHz, DMSO-*d6*) δ 173.97, 170.79, 153.33, 146.18, 139.99, 135.50, 109.26, 108.30, 83.94, 71.48, 67.87, 56.13, 43.88, 33.92, 29.93, 24.07, 24.00, 21.73, 17.26. ESI-HRMS m/z 393.16545 $[M+H]^+$; $C_{19}H_{25}N_2O_7$ requires 393.16563.




Fmoc-4-amino butanoic acid (325 mg, 1.0 mmol), DCC (300 mg, 1.5 mmol) and NHS (200 mg, 1.8 mmol) were stirred in THF (20 mL) at room temperature for 4 hours. Then the solution was filtrated and concentrated. The resulting mixture was purified by flash column chromatography with petroleum ether/ethyl acetate (3:1 to 2:3) to obtain the *N*-hydroxysuccinimide ester. The ester product was dissolved in pyridine (5 mL) and then was added to a solution of **7** (298 mg, 1.0 mmol) in pyridine (15 mL). After stirring overnight, the solution was concentrated. Then ethyl acetate (20 mL) was added. After filtration, the residual powder was washed by ethyl acetate (20 mL) to furnish **11** (430 mg, 71%) as a yellow powder. [1]H NMR (400 MHz, DMSO) δ 8.75 (d, *J* = 4.6 Hz, 1H), 8.61 (d, *J* = 7.7 Hz, 1H), 8.16 (t, *J* = 7.7 Hz, 1H), 7.89 (d, *J* = 7.5 Hz, 2H), 7.71 – 7.66 (m, 2H), 7.49 (s, 1H), 7.41 (t, *J* = 7.4 Hz, 2H), 7.33 (t, *J* = 7.4 Hz, 2H), 7.21 (s, 1H), 5.37 (p, *J* = 6.7 Hz, 1H), 4.29 (d, *J* = 7.0 Hz, 2H), 4.21 (t, *J* = 6.7 Hz, 1H), 4.05 (t, *J* = 6.3 Hz, 2H), 3.88 (s, 3H), 2.93 (dd, *J* = 12.6, 6.5 Hz, 2H), 2.39 (t, *J* = 7.3 Hz, 2H), 2.11 (t, *J* = 7.3 Hz, 2H), 1.99 – 1.90 (m, 2H), 1.56 (dt, *J* = 20.2, 6.9 Hz, 2H), 1.41 (d, *J* = 6.9 Hz, 3H). [13]C NMR (100 MHz, DMSO) δ173.96, 171.09, 156.06, 153.36, 146.24, 146.15, 143.87, 140.69, 140.58, 139.93, 135.58, 127.56, 127.01, 125.35, 125.10, 120.08, 109.38, 108.25, 67.84, 65.17, 56.14, 46.71, 43.88, 39.79, 32.56, 29.91, 25.59, 23.99, 21.80. ESI-HRMS m/z 606.24426 $[M+H]^+$; $C_{32}H_{36}N_3O_9$ requires 606.24461.




**19**




**20**

*L*-Propargylglycine (113 mg, 1.0 mmol), biotinyl-*N*-hydroxysuccinimide ester (341 mg, 1.0 mmol) and DIPEA (258 mg, 2.0 mmol) were stirred in DMF (5 mL) at room temperature for 2 hours. Then the solution was concentrated. The resulting mixture was added with ethyl acetate to precipitate the product **19**, which was collected by filtration (233.4 mg, 68.8%). **19** (34 mg, 0.1 mmol) was added to a solution of HBTU (30 mg, 0.1 mmol) and DIPEA (25.8 mg, 0.2 mmol) in DMF (3 mL). After stirring for 15 minutes, **7** (30 mg, 0.1 mmol) was added. The reaction was stirred for another 2 hours before ethyl acetate (20 mL) was added. After filtration, the residual powder was washed by ethyl acetate (20 mL) and methanol (1 mL) to furnish the product **20** (60.5 mg) as a yellow powder. The product was used for DNA conjugation without further characterization except ESI-HRMS. ESI-HRMS m/z 642.21874 $[M+Na]^+$; $C_{28}H_{37}N_5NaO_9S$ requires 642.22042.

## 4. Preparation of the Dihydropyran Scaffold (Compound 12).[S5]

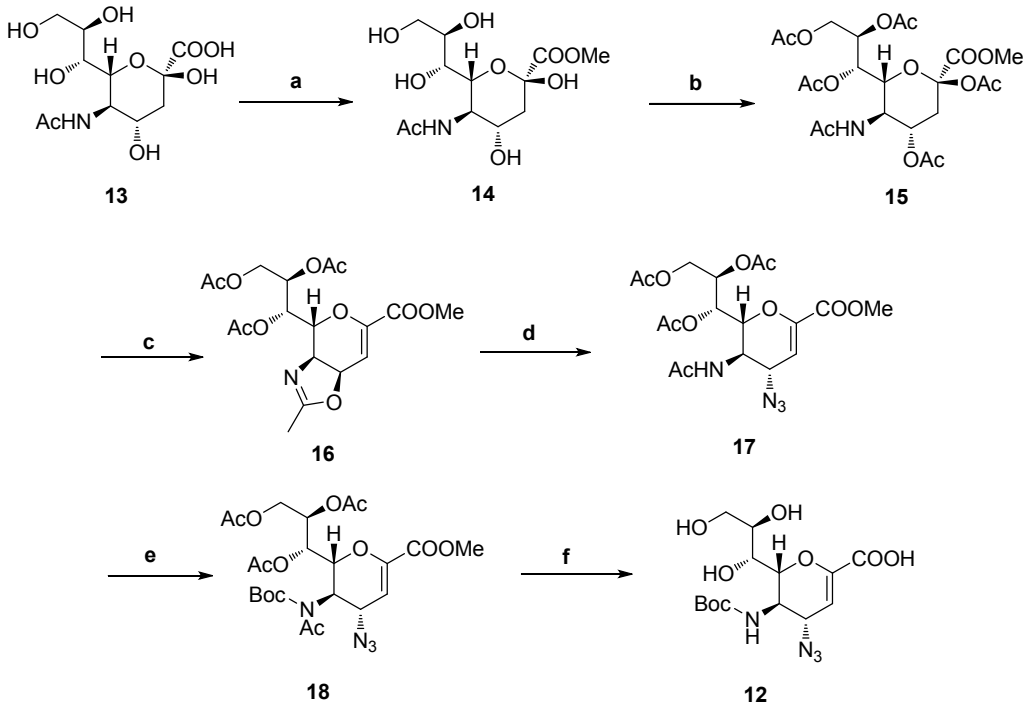



**Figure S2.** a) Dowex, MeOH, r.t., 24 h, 85%;[S5a] b) $Ac_2O$, pyridine, r.t., 12 h, 71%; c) TMSOTf, EtOAc, 0 ºC, 6 h, 70%;[S5b] d) $TMSN_3$, *t*-BuOH, 37 ºC , 11.5 h, 60%; [S5b] e) $Boc_2O$, DMAP, dioxane, r.t., 12 h, 50%; [S5c] f) 0.2 M NaOH, r.t., 24 h, 50%. [S5c]

Preparation of the dihydropyran scaffold (**12**) followed previous reports as shown above in Figure S2. Compound **12** was conjugated to DNA directly via the carboxylic acid handle. The aldehyde group was generated by $NaIO_4$-mediated diol-cleavage immediately before DTS. See details in Section 5(3).

## 5. DNA Sequences, Structures, and Preparation of Modified DNA Oligonucleotides.

**1) DNA sequences.**

- **RD1** and **RD4**: 5'-GCA CGC ACT TAA GGT CTA GTG ATC CGA GCC AGT CTC TAG AC-3'
- **RD2, RD5**, and **RD7-1**: 5'-CTT AAG TGC GTG C**T***C CTG GGC TTG CT-3'
- **RD3, RD3-1** and **RD6**: 5'-TGA CTG TTT CAG TCA AGC AAG CCC AGG **T***AG TGA CTC TGC GAA TAT CG-3'
- **RD7-2**: 5'-CTT AAG TGC GTG D**T***D DTG GGC TTG CT-3'
- **RD1-1:** 5'-CAC GCA CTT AAG GTC TAG TGA TCC GAG CCA GTC TCT AGA C -3'
- **RD1-2:** 5'-CAC GCA CTT AAG GTC AGC TGA TCC GAG CCA GTC TGC TGA C -3'
- **RD1-3:** 5'-CAC GCA CTT AAG GTC CTA TGA TCC GAG CCA GTC TTA GGA C -3'
- **RD2-1:** 5'-CTT AAG TGC GTG TC**T*** ACC TGG GCT TGC T -3'
- **RD2-2:** 5'-CTT AAG TGC GTG GT**T*** CCC TGG GCT TGC T -3'
- **RD3-1:** 5'-TGA CTG TTT CAG TCA AGC AAG CCC AGG **T***AG TGA CTC TGC GAA TAT CG-3'
- **RD3-2:** 5'-TGA GAC TTT GTC TCA AGC AAG CCC AGG **T***AG TGA CTC TGC GAA TAT CG-3'
- **RD3-3:** 5'-TGA TGA TTT TCA TCA AGC AAG CCC AGG **T***AG TGA CTC TGC GAA TAT CG -3'
- Mismatched **RD1** and **RD4** (**misRD1** and **misRD4**): 5'-GCT CCC TCA TTA CGA CAA GAG TTG CCA CCG TCT GTA TTG TC-3'
- Mismatched **RD3** and **RD6** (**misRD3** and **misRD6**): 5'-TGA GAC TAT GAC TGA TGT ACT GCG TCC **T***AC ACT ATC ACC CTA ATT CC-3'

Note: **T*** indicates the thymine base with small molecule modifications (Figure S3). D denotes either A, G or T.

## 2) Reagent DNA structures.

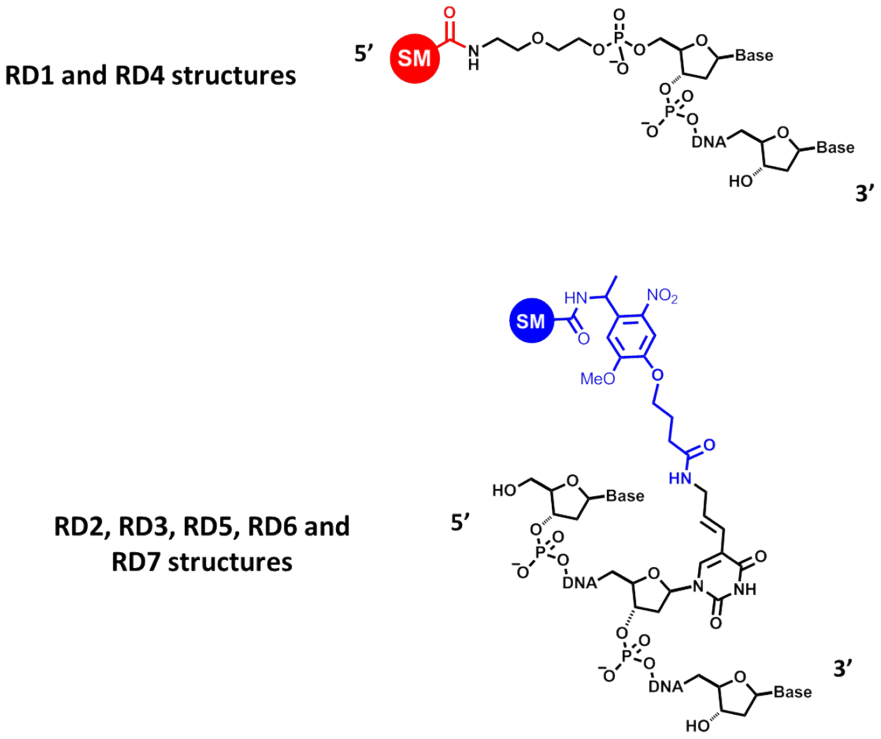



**Figure S3.** Structures of reagent DNAs used in this study.

## 3) Reagent DNA preparations.

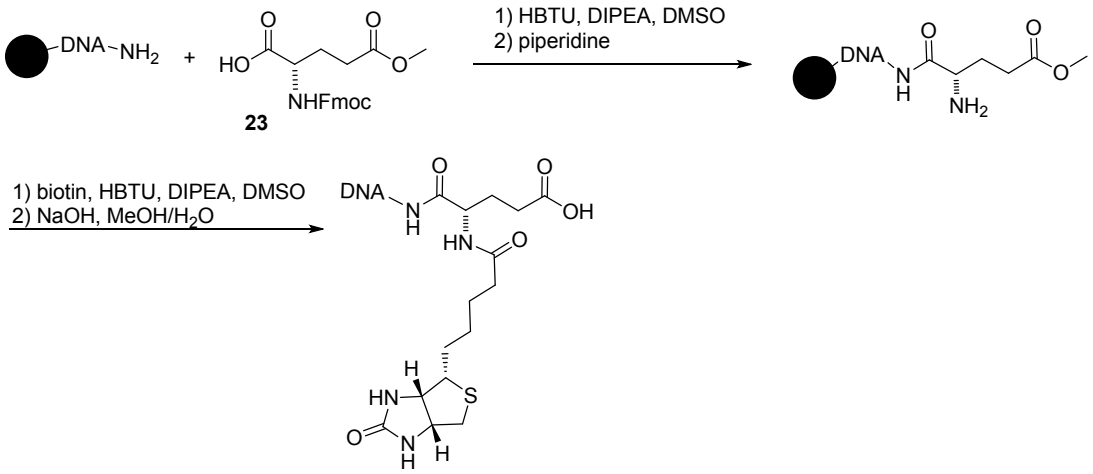
- Preparation of **RD1**, **misRD1** and **RD1-2**:

Bis(2,5-dioxopyrrolidin-1-yl) glutarate (7.0 mg) was dissolved in 100 μL of dry DMF; the solution was treated with 70 μL of amine-modified DNA (**RD1**, **misRD1**, or **RD1-2**) and 70 μL of 1.0 M phosphate buffer (pH 7.2). The mixture was maintained at 37 ºC for 4 hours. The reaction mixture was purified by gel filtration (NAP5 column, GE Pharmacia) and reverse phase HPLC. The product was characterized by denaturing PAGE and MALDI-TOF or ESI-MS.

- Preparation of **RD1-1**:

Bis(2,5-dioxopyrrolidin-1-yl) adipate (7.0 mg) was dissolved in 100 μL of dry DMF; the solution was treated with 70 μL of amine-modified **RD1-1** DNA and 70 μL of 1.0 M phosphate buffer (pH 7.2). The mixture was maintained at 37 ºC for 4 hours. The reaction mixture was purified by gel filtration (NAP5 column, GE Pharmacia) and reverse phase HPLC. The product was characterized by denaturing PAGE and MALDI-TOF or ESI-MS.

- Preparation of **RD1-3**:




CPG with 5'-amine modified oligonucleotides was treated with a solution of compound **23** (1 mmol), HBTU (0.9 mmol), and DIPEA (2 mmol) in DMSO (0.9 mL) at 37 ºC for 12 h with vortexing. Then the CPG beads were treated with 20% piperidine in DMF for 10 min at 37 ºC. After three washes with DMF and acetonitrile, the beads was treated with a solution of biotin (1 mmol), HBTU (0.9 mmol), and DIPEA (2 mmol) in DMSO (0.9 mL) at 37 ºC for 12 h while vortexing. Then the CPG was cleaved by 500 µL of 0.4 M NaOH in MeOH/$H_2O$ (4:1) at 25 ºC for 17 h. The product was quenched by 300 µL of 2 M TEAA, concentrated by speedvac to remove MeOH, purified by reverse phase HPLC, and then characterized by denaturing PAGE and MALDI-TOF or ESI-MS.

- Preparation of **RD2, RD2-1,** and **RD2-3**:

EDCI (5.0 mg) and NHS (4.0 mg) were added to a solution of compound **9** (14.0 mg) in DCM (1.5 mL) at room temperature for 6 hours. The solution was washed three times with a $NaHSO_4$ solution (5%, w/v). Then the organic layer was dried and concentrated. The crude product was re-dissolved in dry DMSO (400 µL) and was treated with amine-modified DNA (**RD2, RD2-1,** or **RD2-3;** 50 µL) and TEA/HCl buffer (0.5 M, pH 7.0, 50 µL) at 37 ºC for 6 hours. The modified DNA was purified by ethanol precipitation and reverse phase HPLC. The product solution was lyophilized and treated with 200 µL of 0.1 M NaOH solution at 37 ºC for 4 hours, and then the reaction was quenched by 50 µL of 2 M TEAA. The product was purified by reverse phase HPLC and characterized by denaturing PAGE and MALDI-TOF or ESI-MS.

The *t*BuS-protected DNA was deprotected by 100 mM DTT in a total volume of 100 µL of 200 mM MOPS buffer (pH 8.5) at 25 ºC for 2 hours. The resulting oligonucleotide was purified by gel filtration and lyophilization with degassed 0.1 M TEAA buffer.

- Preparation of **RD2-2**:

**RD2-2** was prepared with the amine-modified **RD2-2** DNA and compound **24** with the same procedure as for **RD2**.

- Preparation of **RD3, misRD3,** and **RD3-1**:

EDCI (3.0 mg) and NHS (4.0 mg) were added to a solution of compound **8** (8.0 mg) in DCM (1.5 mL) at room temperature for 6 hours. The solution was washed three times with a $NaHSO_4$ solution (5%, w/v). Then the organic layer was dried and concentrated. The crude product was re-dissolved in dry DMSO (400 μL) and treated with DNA (**RD3, misRD3,** or **RD3-1;** 50 μL) and TEA/HCl buffer (0.5 M, pH 10.0, 50 μL) at 37 ºC for 6 hours. The final product was purified by ethanol precipitation and reverse phase HPLC and was characterized by denaturing PAGE and MALDI-TOF mass spectrometry.

- Preparation of **RD3-2**:

**RD3-2** was prepared with the amine-modified **RD3-2** DNA and compound **21** with the same procedure as for **RD3-1**.

- Preparation of **RD3-3**:

**RD3-3** was prepared with the amine-modified **RD3-3** DNA and compound **22** by similar procedure as for **RD3-1**.

- Preparation of **RD4** (and **misRD4**):

DCC (3.0 mg), NHS (4.0 mg) and compound **12** (4.0 mg) were dissolved in DMF (100 μL). The solution was maintained at 37 ºC for 2 hours and then centrifuged. 70 μL of the supernatant was treated with amine-modified **RD4** DNA (70 μL) and phosphate buffer (0.5 M, pH 7.0, 70 μL) at 37 ºC for 2 hours. The modified DNA was purified by gel filtration and reverse phase HPLC, and then was characterized by MALDI-TOF mass spectrometry. The product solution was lyophilized and treated with 200 μL of 50 mM $NaIO_4$ at 37 ºC for 30 minutes. The resulting DNA product was desalted by gel filtration to obtain **RD4** (or **misRD4**).

- Preparation of **RD5** (and **RD7-2**):

EDCI (3.0 mg) and NHS (4.0 mg) were added to a solution of compound **10** (10.0 mg) in DCM (1.0 mL). The

solution was maintained at 37 ºC for 2 hours before being washed three times with a $NaHSO_4$ solution (5 %). Then the organic layer was dried and concentrated. The crude product was re-dissolved in dry DMSO (400 μL) and treated with amine-modified **RD5** (or **RD7-2**) DNA (50 μL) and TEA/HCl buffer (0.5 M, pH 10.0, 50 μL) at 37 ºC for 2 hours. The modified DNA product was purified by ethanol precipitation and reverse phase HPLC and was characterized by denaturing PAGE and MALDI-TOF.

- Preparation of **RD6** (and **misRD6**):

EDCI (5.0 mg) and NHS (4.0 mg) were added to a solution of compound **11** (14.0 mg) in DCM (1.5 mL). The solution was maintained at 45 ºC for 2 hours and then was concentrated. The crude product was re-dissolved in dry DMSO (400 μL) and then treated with amine-modified **RD6** DNA (50 μL) and TEA/HCl buffer (0.5 M, pH 10.0, 50 μL) at 37 ºC for 2 hours. The modified DNA product was purified by ethanol precipitation and reverse phase HPLC.

The product solution was lyophilized and treated with 200 μL of 0.1 M NaOH solution at 37 ºC for 4 hours. Then the reaction was quenched by 50 μL of 2 M TEAA and was purified by reverse phase HPLC and characterized by denaturing PAGE and MALDI-TOF mass spectrometry.

- Preparation of **RD7-1**:

DCC (100 mM), NHS (100 mM) and compound **20** (100 mM) in DMSO (100 μL) was maintained at 45 ºC for 2 hours. The crude product was treated with amine-modified **RD7-1** DNA (50 μL), DMSO (200μL) and TEA/HCl buffer (0.5 M, pH 10.0, 50 μL) at 37 ºC for 2 hours. The modified DNA product was purified by ethanol precipitation and reverse phase HPLC and characterized by denaturing PAGE and MALDI-TOF mass spectrometry.

## 6. Oligonucleotide Characterization.

| oligonucleotide | expected mass (Da) | observed mass (Da) | note |
|---|---|---|---|
| **RD1** | 12845.2 | 12845.8 | ESI |
| **RD2**, with -S(*t*Bu) | 8409.5 | 8407.2 | MALDI |
| **RD3** or **RD3-1** | 14938.2 | 14930.2 | MALDI |
| mismatched **RD1** | 12768.2 | 12767.0 | MALDI |
| mismatched **RD3** | 14785.6 | 14778.8 | MALDI |
| **RD4** (before oxidation) | 13094.5 | 13094.1 | MALDI |
| **RD5** | 8390.5 | 8389.3 | MALDI |
| **RD6** | 14899.7 | 14903.5 | MALDI |
| mismatch **RD4** (before oxidation) | 13010.5 | 13011.9 | MALDI |
| mismatch **RD6** | 14728.6 | 14722.3 | MALDI |
| **RD7-1** | 8617.8 | 8620.7 | MALDI |
| **RD1-1** | 12532.1 | 12530.7 | MALDI |
| **RD1-2** | 12523.1 | 12525.6 | MALDI |
| **RD1-3** | 12763.4 | 12766.8 | MALDI |
| **RD2-1,** with -S(*t*Bu) | 8815.0 | 8815.2 | MALDI |
| **RD2-3,** with -S(*t*Bu) | 8197.6 | 8199.6 | MALDI |
| **RD3-2** | 14963.7 | 14961.1 | MALDI |
| **RD3-3** | 14972.8 | 14974.7 | MALDI |

## 7. Experimental Conditions.[S4]

- **DNA-templated reactions:**

**RD2** (400 pmol) and **RD3** (400 pmol) were combined in a total volume of 400 μL of 100 mM MOPS buffer (pH 7.0) and 1 M NaCl at 25 ºC for 6 hours before **RD1** (400 pmol) was added. DMT-MM (5.9 mg, 50 mM final concentration) was added to the mixture, which was maintained at 25 ºC for 12 hours. The reaction product was purified by ethanol precipitation and denaturing PAGE.

**RD4** (400 pmol), **RD5** (400 pmol), $Cu(OAc)_2$ (500 μM) and sodium ascorbate (500 μM) were maintained in a total volume of 400 μL of 100 mM MES buffer (pH 6.0) and 1 M NaCl at 25 ºC for 3 hours. Then **RD6** (400 pmol) and $NaBH_3CN$ (1.3 mg, 50 mM final concentration) were added. The reaction mixture was maintained at 25 ºC for 4 hours, and the reaction product was purified by ethanol precipitation and denaturing PAGE.

- **Gel purification:**

Gel bands were chopped into fine particles and added with 10 volumes of 0.1 M TEAA. The sample was frozen for 30 minutes at -80 ºC and then was quickly thawed in a hot water bath. After soaking for 5 minutes at 90 ºC, it was placed on a rotary shaker overnight at 37 ºC. The supernatant was collected and concentrated by extracting against an equal volume of *n*-butanol. The upper butanol layer was removed to repeat the extraction procedure until the lower aqueous volume is convenient for precipitation.

- **Ethanol precipitation:**

The sample's pH was adjusted with 0.1 volume of 3 M NaOAc (pH 5.0); then 0.1 volume of 7.5 mg/mL glycogen (Aldrich, final glycogen concentration: 200 μg/mL) and 2.5 volume of absolute ethanol were added. The solution was maintained at -80 °C for 2 hours and then centrifuged at 17,000 g for 15 minutes at 4 °C. The supernatant was discarded and the pellet was rinsed once with cold 70% ethanol. After centrifuge at 17,000 g for another 5 minutes at 4 °C, the supernatant was discarded and the pellet was dried by a speedvac. Recovered sample was dissolved in appropriate buffer for subsequent analysis or experiments.

- **Enzyme-mediated ligation:**[S6]

A 50 μL solution containing 10 units of T4 PNK and 40 pmol of gel-purified oligonucleotide in 1x T4 DNA

ligase buffer and 1 mM ATP was maintained at 37 ºC for 4 hour, before 350 units of T4 DNA ligase were added. The reaction solution was maintained at 16 ºC for 12 hours. The ligation product was purified by gel filtration, ethanol precipitation, and denaturing PAGE.

- **Photocleavage:**

Oligonucleotides were dissolved in 0.1 M TEAA buffer (pH 7.0) to a concentration of 1.0 μM. The solution was irradiated under 365 nm at 0 ºC for 15 minutes. The cleavage product was purified by ethanol precipitation or lyophilization and analyzed by denaturing PAGE.

- ***Afl*II digestion:**

50 μL of 100 pmol oligonucleotide was mixed with 20 units of *Afl*II and 5 μL 0.1% BSA in 1x *Afl*II buffer at 37 ºC for 4 hour. The digestion product was purified by ethanol precipitation.

- **PCR amplification:**

Typically, the PCR solution contains 0.6 nM template, 1.0 mM primers (forward and reverse), 8 mM $MgSO_4$, 200 μM dNTP, 0.5 M betaine, 10 μg/mL BSA and 20 units/mL of *Vent* (exo$^-$) polymerase (NEB). The mixture was incubated for 2 minutes at 92 ºC, followed by 30 cycles of 30 seconds at 92 ºC, 15 seconds at 55 ºC and 1 minutes at 70 ºC, which was followed by 2 minutes' final incubation at 70 ºC. The PCR product was purified by ethanol precipitation.

## 8. PAGE Analysis of Enzymatic Ligation and PCR-Amplification of the Products.

### 1) *N*-acylthiazolidine.

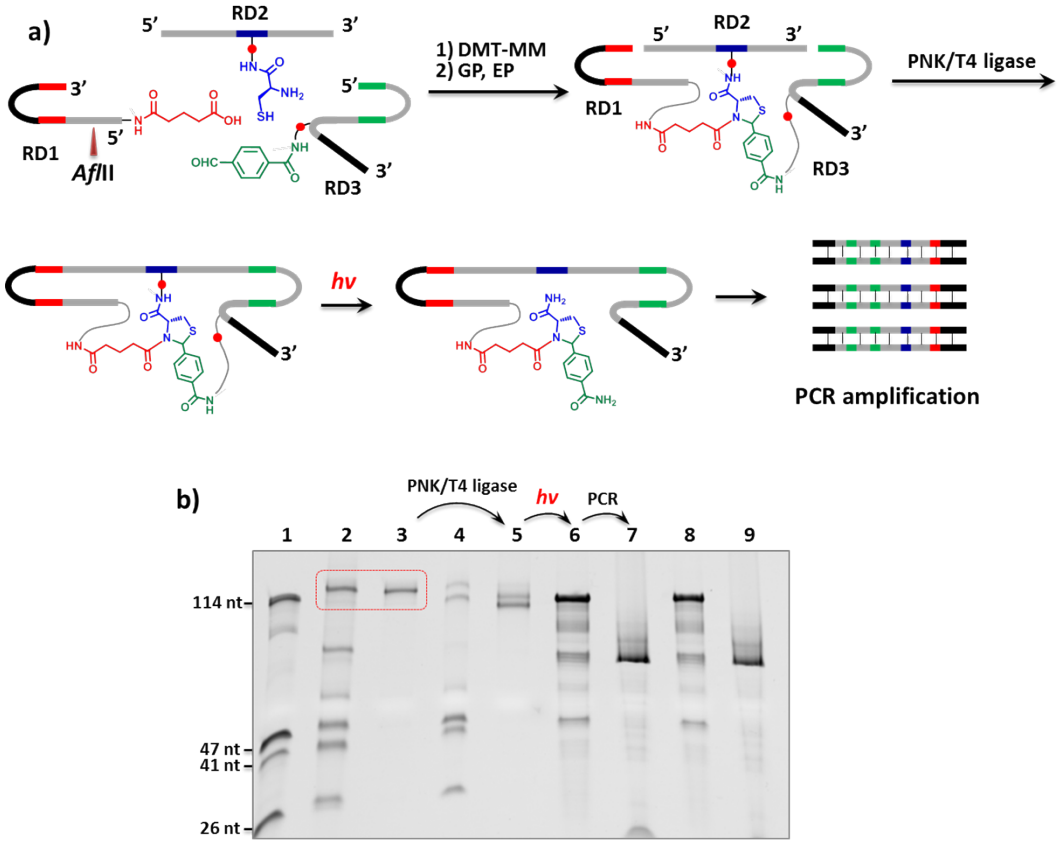



**Figure S4.** a) Scheme of DNA-templated chemical reaction, enzyme-mediated ligation, photo-cleavage, and PCR amplification; b) Denaturing PAGE analysis. Lane 1: DNA standards; lane 2: **RD2** and **RD3** were incubated for 6 h, then **RD1** was added with DMT-MM (50 mM) for 12 h; lane 3: after gel purification of the 114-nt product from lane 2; lane 4: lane 2 after irradiation; lane 5: lane 3 after PNK/T4 DNA ligase-mediated ligation; lane 6: lane 5 after irradiation; lane 7: PCR amplification of the lane 6 ligation product; lane 8: **RD1**, **RD2** and **RD3** were directly ligated by PNK/T4 DNA ligase without subjecting to chemical reactions (used as a reference for the lane 6 product); lane 9: PCR product standard sample. GP: gel purification; EP: ethanol precipitation. In b), red rectangle shows the 114-nt chemical reaction product before (lane 2) and after (lane 3) gel purification.

As shown in Figure S4a, after chemical reactions between **RD1**, **RD2** and **RD3** by DNA assembly, the reaction product was gel-purified and subjected to PNK/T4 DNA ligase-mediated ligation to connect DNA strands. After photo-cleavage, the product was subjected to *Afl*II digestion and MALDI-MS analysis (Figure 2 of the main text). Additionally, we also performed PCR amplification to ensure that the ligated DNA, which is a duplex DNA encoding the final product, can be efficiently amplified. The whole process analyzed by denaturing PAGE is shown above in Figure S4b.

PCR amplification of the final product DNA is expected to generate a 95-bp DNA duplex, which is too stable

to be denatured under PAGE condition (15% polyacrylamide, 25% formamide, 7 M urea, 1x TBE) due to its length; therefore, it appeared at a position slightly lower than its expected length (lane 7). We have validated this with a standard sample (lane 9, the PCR amplification product from a standard DNA).

- Reverse primer for PCR: 5'-TGATCCGAGCCAGTCT-3'
- Forward primer for PCR: 5'-CGATATTCGCAGAGTCAC-3'

## 2) Dihydropyran scaffold decoration.

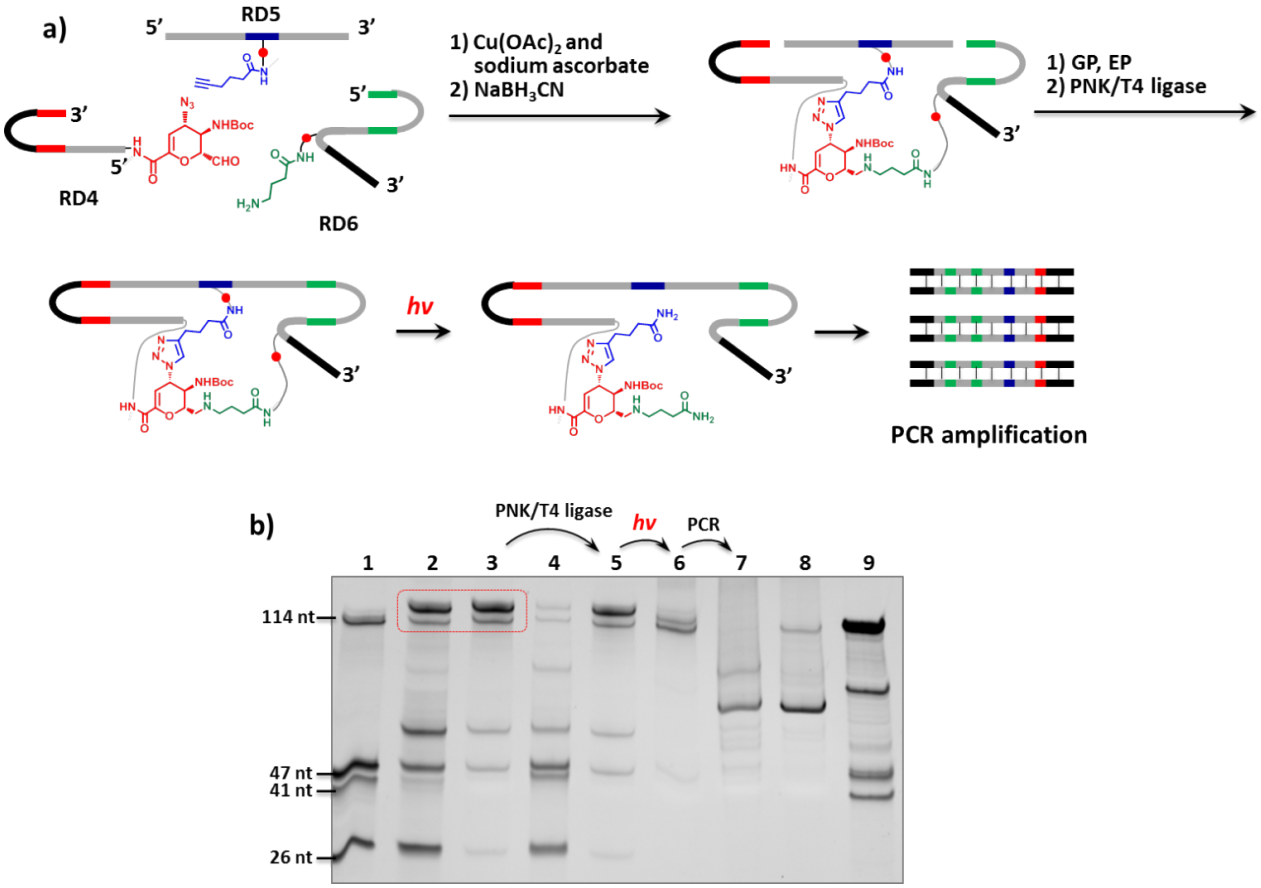



**Figure S5.** a) Scheme of DNA-templated chemical reaction, enzyme-mediated ligation, photo-cleavage, and PCR amplification; b) Denaturing PAGE analysis. Lane 1: DNA standards; lane 2: **RD4** and **RD5** were incubated with $Cu(OAc)_2$ (500 μM) and sodium ascorbate (500 μM) for 3 hours, then **RD6** was added along with $NaBH_3CN$ (50 mM) for 4 hours; lane 3: reaction product of lane 2 after gel purification; lane 4: lane 2 after irradiation; lane 5: lane 3 after enzyme-mediated ligation; lane 6: lane 5 after irradiation; lane 7: PCR amplification of the lane 6 ligation product; lane 8: PCR product standard; lane 9: **RD4**, **RD5** and **RD6** were directly ligated by PNK/T4 DNA ligase without subjecting to chemical reactions (used as a reference for the lane 6 product). GP: gel purification; EP: ethanol precipitation. In b), red rectangle shows the 114-nt chemical reaction product before (lane 2) and after (lane 3) gel purification.

The rationale of the experiments shown above in Figure S5 is the same as in Figure S4. Similar results were obtained.

## 9. Structures of Control RD4 Reagent DNAs.

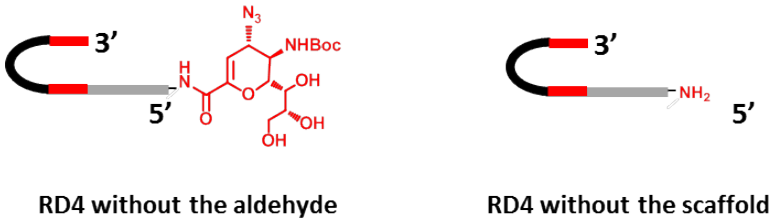



**Figure S6.** Negative control reagent DNA **RD4**'s used in Figure 3 of the main text..

## 10. Model Library Synthesis and *in vitro* Selection against Immobilized Streptavidin (Figure 4).

### 1) Model library synthesis.

The model library in Figure 4 was synthesized with **RD4**, **RD6**, and **RD7** following the same procedure as described above in section 7. **RD7** is a mixture of **RD7-1** and **RD7-2** at 1:100 molar ratio.

### 2) Selection against immobilized streptavidin.

The model library (~1 nmol) was dissolved in 100 µL 1x PBS and then was incubated with 30 µL high capacity streptavidin agarose resin (Thermo Scientific) for 2 hours at 4 ºC. Resin were washed with 0.5 M NaCl for 5 times, 4 M guanidine hydrochloride for 5 times, 0.5 M NaCl for 5 times, 4 M guanidine hydrochloride for 5 time, and water for 2 times. Selected library members were eluted by incubating the resin in 95% formamide + 10 mM EDTA (pH 8.2) at 95 ºC for 20 minutes. Eluted library members were collected by ethanol precipitation.

### 3) PCR amplification (for Sanger sequencing).

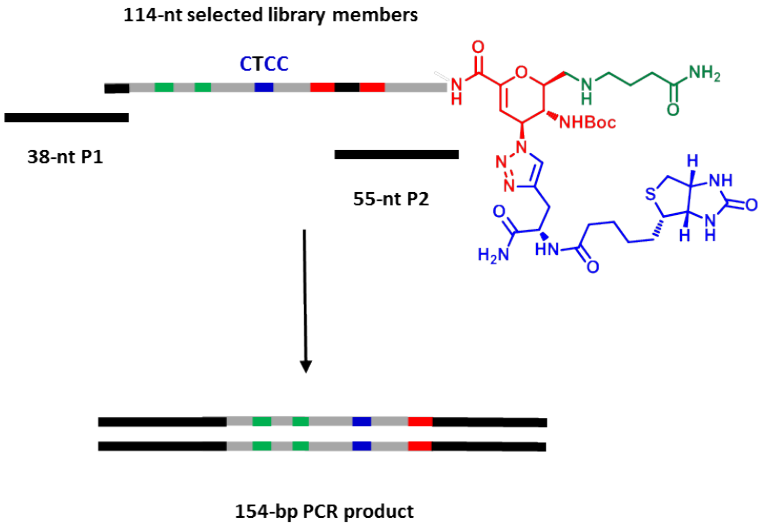



**Figure S7.** PCR amplification of selected library members with primers **P1** and **P2**.

PCR reactions followed the same procedure as described above in section 7.

Primer sequences for PCR amplification:

- **P1:** 5'-GGG CAA GGA GCA CAT CAA GTC GAT ATT CGC AGA GTC AC-3'
- **P2:** 5'-CCG GAA CAT CAA GGA CCC TAT ACT TTA AGT GCC AAG GAC TGA TCC GAG CCA GTC T-3'

## 11. Sanger Sequencing.

- **Predicted PCR products before selection: 5'-**CCG GAA CAT CAA GGA CCC TAT ACT TTA AGT GCC AAG GAC TGA TCC GAG CCA GTC TCT AGA CCT TAA GTG CGT G**DTDD**T GGG CTT GCT TGA CTG TTT CAG TCA AGC AAG CCC AGG TAG TGA CTC TGC GAA TAT CGA CTT GAT GTG CTC CTT GCC C-3' (and its complementary sequence; D = A, G or T)

- **Predicted PCR products after selection: 5'-**CCG GAA CAT CAA GGA CCC TAT ACT TTA AGT GCC AAG GAC TGA TCC GAG CCA GTC TCT AGA CCT TAA GTG CGT G**CTCC**T GGG CTT GCT TGA CTG TTT CAG TCA AGC AAG CCC AGG TAG TGA CTC TGC GAA TAT CGA CTT GAT GTG CTC CTT GCC C-3' (and its complementary sequence)

- **Primer sequences for sequencing:**

  **P3:** 5'-GGG CAA GGA GCA CAT CAA GT-3'

  **P4:** 5'-CCG GAA CAT CAA GGA CCC TA-3' (for sequencing of the opposite strand).

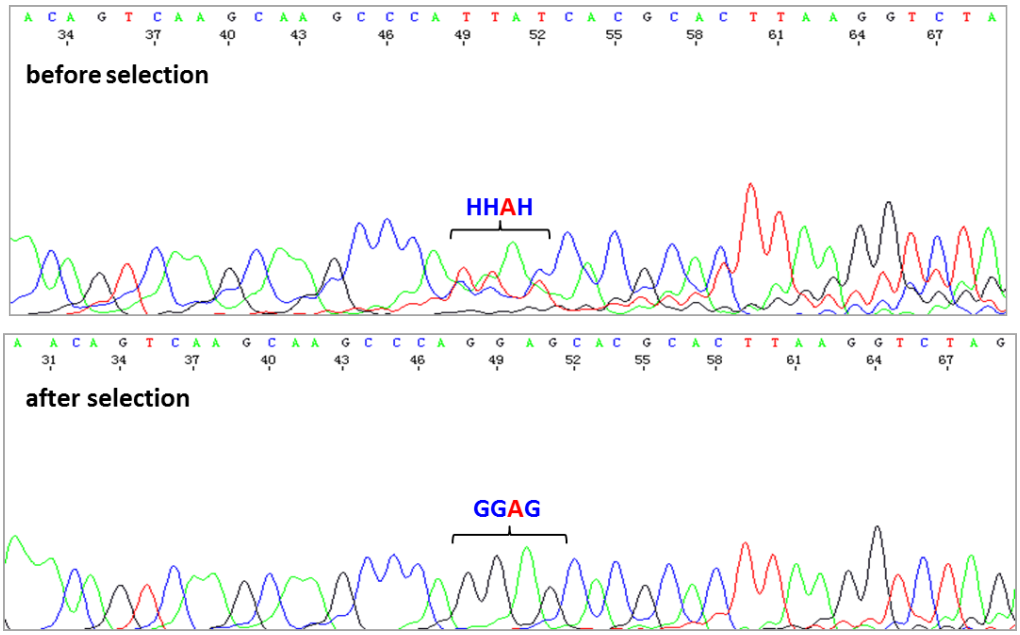



**Figure S8.** Sequencing results of the opposite strand of the model library before and after selection against immobilized streptavidin. The encoding sites before and after the selection were compared as marked in the figure. H: A, C or T.

## 12. *N*-acylthiazolidine Model Library Synthesis and *in vitro* Selection against Immobilized Streptavidin (Figure 5).

### 1) Library synthesis.

The model library was synthesized with **RD1-1, RD1-2, RD1-3**, **RD2-1**, **RD2-2**, **RD3-1, RD3-2** and **RD3-3** following the same procedure as described above in section 7. Molar ratio of **RD1-1**, **RD1-2** and **RD1-3** is 200:200:1.

### 2) Selection against immobilized streptavidin.

Selection procedure followed the same procedure as described above in Section 10.

**3) PCR and Illumina® sequencing.**

High-throughput sequencing was performed on an Illumina HiSeq 2500 sequencer using the standard 2 x 100 bp paired-end sequencing procedure. For each sample we obtained at least $10^6$ paired-end reads for accurate assessment of the enrichment fold changes.

- **PCR amplification, gel extraction and product quantitation.**

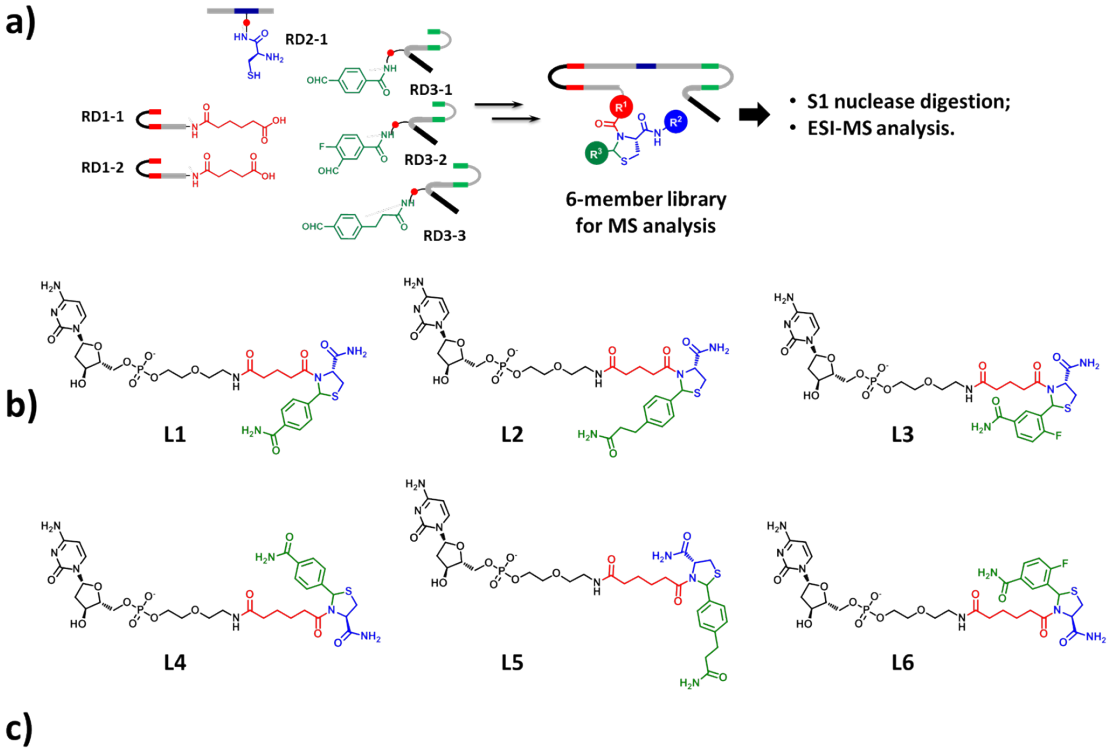
The eluted library was amplified by PCR with the procedure as described above. After PCR amplification, the products were recovered by gel extraction with the Gel Extraction Kit (CWBIO) following manufacturer's protocol.

**Forward Primer:** 5'-
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTGATCCGAGC
CAGTCT-3'

**Reverse Primers:** 5'-
*CAAGCAGAAGACGGCATACGAGAT*GCTGTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGATA
TTCGCAGAGTCAC-3'
5'-
*CAAGCAGAAGACGGCATACGAGAT*TCGGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCGATA
TTCGCAGAGTCAC-3'

Note: Italic fonts indicate Illumina sequencing adaptors. Sequences in read are primer binding sites for PCR amplification.

## 13. *N*-acylthiazolidine Model Library Synthesis and LC/MS characterization.




c)

| Library compounds | Expected mass (Da) | Observed mass (Da) | LC retention time (min) |
|---|---|---|---|
| L1 | 740.2121 | 740.2107 | 4.64 |
| L2 | 768.2434 | 768.2436 | 4.92 |
| L3 | 758.2026 | 758.2061 | 5.08 |
| L4 | 754.2277 | 754.2277 | 4.75 |
| L5 | 782.2590 | 782.2589 | 5.05 |
| L6 | 772.2183 | 772.2186 | 5.27 |

**Figure S9**: a) Preparation of a 6-member *N*-acylthiazolidine library for MS characterization. The library was assembled with the same procedure as in Figure 2 of the main text. All reagent DNAs are at equal ratio. The library was treated with S1 nuclease and then analyzed by ESI-MS. b) Expected structures of library compounds after S1 nuclease digestion. c) UPLC ESI-MS analysis results.

### 1) Model library synthesis.

The model library was synthesized with **RD1-1, RD1-2**, **RD2-3**,, **RD3-1, RD3-2** and **RD3-3** following the same procedure as described above in section 7.

**2) S1 nuclease digestion.**

500 pmol of library was digested by 1000U of S1 nuclease (Fermentas) in 50 μL of 1x buffer at 37 ºC for 2 h and 45 ºC for 1 h. Then the sample was lyophilized, dissolved in 50 μL of water, and directly used for LC-MS analysis.

**3) UPLC-ESI-MS characterization.**

Digestion product was analyzed with an ACQUITY UPLC system (WATERS) using a Waters T3 column and a 12 min linear gradient of acetonitrile (0% -100%) in 0.1% (v:v) aqueous formic acid. The eluent was directly injected into a Xevo G2 Q-TOF mass spectrometer (WATERS). Data was acquired and processed with MassLynx 4.1 software (Waters). Mass spectra were recorded in the negative mode within an m/z range of 100-1000.

## 14. References:


S1. N. Hammond, P. Koumi, G. J. Langley, A. Lowe, T. Brown, *Org. Biomol. Chem.* **2007**, *5*, 1878.

S2. C. P. Holmes, *J. Org. Chem.* **1997**, *62*, 2370.

S3. R. C. Clevenger, J. M. Raible, A. M. Peck, B. S. J. Blagg, *J. Org. Chem.* **2004**, *69*, 4375.

S4. a) X. Li, Z. J. Gartner, B. N. Tse, D. R. Liu, *J. Am. Chem. Soc.* **2004**, *126*, 5090. b) C. T. Calderone, D. R. Liu, *Angew. Chem. Int. Ed.* **2005**, *44*, 7383.

S5. a) J. Deng, Z. Sheng, K. Zhou, M. Duan, C. Yu, L. Jiang, *Bioconjugate Chem.* **2009**, 20, 533. b) M. Chandler, M.J. Bamford, R. Conroy, B. Lamont, B. Patel, V. K. Patel, I. P. Steeples, R. Storer, N. G. Weir, M. Wright, C. Williamson, *J. Chem. Soc. Perkin Trans*. **1995**, 1173. c) P. W. Smith, I. D. Starkeyl, P. D. Howesl, S. L. Sollisl, S. P. Keelingl, P. C. Cherryl, M. Itzsteinz, W. Y. Wu, B. Jin, *Eur. J. Med. Chem.* **1996**, *31*,143.

S6. M. H. Hansen, P. Blakskjaer, L. K. Petersen, T. H. Hansen, J. W. Højfeldt, K. V. Gothelf, N. J. Hansen, *J. Am. Chem. Soc.* **2009**, *131*, 1322.