Electronic Supplementary Material (ESI) for ChemComm.


# Design and Synthesis of *ortho*-Phthalaldehyde Phosphoramidite for Single-step, Rapid, Efficient and Chemoseletive Coupling of DNA with Proteins under Physiological Conditions

Yanli Ma,[a] Zhehao Lv, [a] Tingyu Li [a], Tian Tian [a], Lianyu Lu [a], Wenli Liu, [a] Zhi Zhu, [a] Chaoyong Yang [a, b]

[a]*MOE Key Laboratory of Spectrochemical Analysis & Instrumentation, Collaborative Innovation Center of Chemistry for Energy Materials, Key Laboratory for Chemical Biology of Fujian Province, State Key Laboratory of Physical Chemistry of Solid Surfaces, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China.*

[b]*Institute of Molecular Medicine, Renji Hospital, Shanghai Jiao Tong University, School of Medicine, Shanghai, 200240, China*

## Materials and reagents

All DNA synthesis reagents were purchased from ChemGenes (Wilmington, MA, USA). Methyl 3-(4-hydroxyphenyl)propanoate, 6-amino-1-hexanol, N,N'-dicyclohexylcarbodiimide (DCC), 1-hydroxybenzotrizole (HOBt) and N,N'-diisopropylethylamine (DIPEA) were purchased from J&K Scientific Ltd. (Shanghai, China). Sodium chloride and potassium chloride were purchased from Jinhuada Chemical Reagents (Guangzhou, China). And other reagents were obtained from Sinopharm Chemical Reagents (Shanghai, China). All solutions were prepared with Milli-Q ultrapure water (grade: 18.2 MΩ). $^{1}$H and $^{13}$C-NMR spectra were recorded on a Bruker AM 500 ($^{1}$H: 500 MHz, $^{31}$P: 202 MHz) at room temperature. Chemical shifts are expressed in parts per million (ppm) and the spectra are calibrated to residual solvent signals of $CDCl_3$ (7.26 ppm ($^{1}$H)). Coupling constants are given in hertz (Hz) and the following notations indicate the multiplicity of the signals: s (singlet), d (doublet), brs (broad singlet), t (triplet), q (quartet), m (multiplet), ap. (apparent).

## Synthesis of OPA phosphoramidite

### (a): 3-(3-Formyl-4-hydroxyphenyl)propanoate (1)

Methyl 3-(4-hydroxyphenyl)propanoate (10 g, 55.6 mmol, 1.0 eq.) and anhydrous magnesium dichloride (7.8 g, 83.4 mmol, 1.5 eq.) were dissolved in 240 mL of $CH_3CN$, and then dry paraformaldehyde (11.24 g, 376 mmol, 6.8 eq.) was added. The reaction mixture was heated to reflux for 8 h. The reaction mixture was cooled to room temperature and poured into 600 mL of 5% HCl. The mixture was extracted with diethyl ether (300 mL × 3). The organic phase was washed with a saturated NaCl solution and dried using anhydrous $Na_2SO_4$. After removing the solvent, the crude product was purified by silica gel column chromatography (hexane / ethyl acetate = 10 : 1) to obtain methyl 3-(3-formyl-4-hydroxyphenyl)propanoate (compound 1). $^1$H NMR (500 MHz, $CDCl_3$):δ = 10.88 (s, 1H), 9.86 (s, 1H), 7.38-7.36 (m, 2H), 6.92 (d, 1H), 3.66 (s, 3H), 2.94 (t, 2H), 2.63 (t, 2H) ppm. $^{13}$C NMR (500 MHz, $CDCl_3$): δ = 196.63, 173.13, 160.31, 137.40, 133.14, 132.12, 120.60, 117.92, 51.85, 35.65, 29.84 ppm. ESI-MS (m/z) $[1 + H]^+$ 208.1, found: 208.1.

**(b&c): Methyl 3-(3,4-diformylphenyl)propanoate (2)**

3-(3-formyl-4-hydroxyphenyl)propanoate (10 g, 48 mmol, 1.0 eq.) was dissolved in 100 mL EtOH. Then, formic acid hydrazide (2.8 g, 48.0 mmol, 1.0 eq.) in EtOH (40 mL) was slowly added to the reaction mixture, which was then refluxed for 2 h, followed by cooling in an ice bath. The yellow precipitate was filtered and washed three times with cold EtOH. The precipitate was evaporated under vacuum and dissolved in 250 mL anhydrous THF. Then, lead(IV) acetate (21.2 g, 48.0 mmol, 1.0 eq.) was added in small portions, followed bystirring at room temperature for 2 h. The solvent was then evaporated under vacuum, and the crude product was extracted with EtOAc and water. The combined organic layers were washed with brine and dried with anhydrous $Na_2SO_4$. After removing the solvent, the residue was purified by silica gel column chromatography (hexane / ethyl acetate = 5 : 1) to obtain methyl 3-(3,4-diformylphenyl)propanoate (compound 2). $^1$H NMR (500 MHz, $CDCl_3$):δ = 10.54 (s, 1H), 10.45 (s, 1H), 7.90 (s, 1H), 7.80 (d, 1H), 7.61 (m, 1H), 3.66 (s, 3H), 3.09 (t, 2H), 2.70 (t, 2H) ppm. $^{13}$C NMR (500 MHz, $CDCl_3$): δ =192.44, 192.08, 147.50, 136.76, 134.82, 133.76, 131.96, 130.88, 51.96, 34.74, 30.78 ppm. ESI-MS (m/z) $[1 + H]^+$ 220.1, found: 220.7.

**(d): Methyl 3-(1, 3-dimethoxy-1, 3-dihydroisobenzofuran-5-yl)propanoate (3)**

To the solution of compound 2 (4.2 g, 19.1 mmol, 1.0 eq.) in anhydrous MeOH (50 mL), trimethyl orthoformate (10.4 mL, 95.4 mmol, 5.0 eq.) and lithium tetrafluoroborate (0.5 g, 5.7

mmol, 0.3 eq.) were slowly added, and the reaction mixture was heated to reflux for 2 h. The reaction mixture was cooled to room temperature and quenched by $Et_3N$ (0.8 mL, 5.7 mmol, 0.3 eq.). Then the solvent was evaporated under vacuum, and the reaction residue was partitioned between EtOAc and water. The aqueous layer was extracted with EtOAc (100 mL × 3), and the combined organic layers were washed with brine and dried with anhydrous $Na_2SO_4$. After removing the solvent, the residue was purified by silica gel column chromatography (hexane / ethyl acetate = 4 : 1) to obtain methyl 3-(1, 3-dimethoxy-1,3-dihydroisobenzofuran-5-yl)propanoate (compound 3). $^1$H NMR (500 MHz, $CDCl_3$): δ =7.33-7.27 (m, 2H), 7.23 (s, 1H), 6.29 (t, 1H), 6.03 (d, 1H), 3.66 (s, 3H), 3.43 (m, 6H), 3.00 (m, 2H), 2.64 (m, 2H) ppm. $^{13}$C NMR (500 MHz, $CDCl_3$): δ = 173.17, 142.66, 139.04, 130.25, 130.13, 123.17, 122.80, 106.65, 105.54, 54.48, 54.26, 51.80, 35.73, 30.92 ppm. ESI-MS (m/z)$[1 + Na]^+$ 289.1, found: 288.7.

**(e): 3-(1, 3-Dimethoxy-1,3-dihydroisobenzofuran-5-yl)propanoic acid (4)**

To the solution of compound 3 (3.5 g, 13.2 mmol, 1.0 eq.) in THF (55 mL), 13.2 mL of 1 N NaOH solution was slowly added, and then the reaction mixture was stirred at room temperature for 6 h. After neutralization with 1 N HCl solution, the reaction mixture was extracted twice with EtOAc, and the combined organic layers were washed with brine and dried with anhydrous $Na_2SO_4$. After removing the solvent, the residue was purified by silica gel column chromatography (hexane / ethyl acetate = 1 : 1) to obtain 3-(1,3-dimethoxy-1,3-dihydroisobenzofuran-5-yl)propanoic acid (compound 4). $^1$H NMR (500 MHz, $CDCl_3$): δ = 7.34-7.27 (m, 2H), 7.25 (s, 1H), 6.30–6.29 (m, 1H), 6.04 (d, 1H), 3.45-3.42 (m, 6H), 3.02-2.99 (m, 2H), 2.70-2.69 (m, 2H) ppm. $^{13}$C NMR (500 MHz, $CDCl_3$): δ = 178.08, 142.18, 138.80, 136.50, 130.05, 129.94, 123.04, 122.62, 106.42, 105.31, 54.05, 35.38, 30.38 ppm. ESI-MS (m/z)$[1 + Na]^+$ 275.1, found: 274.6.

**(f): 3-(1,3-Dimethoxy-1,3-dihydroisobenzofuran-5-yl)-N-(6-hydroxyhexyl)propanamide (5)**

In a 50 mL oven-dried round-bottomed flask, compound 4 (724 mg, 2.9 mmol, 1.0 eq.), 6-amino-1-hexanol (680 mg, 5.8 mmol, 2.0 eq.), 1-hydroxybenzotriazole (HOBT, 580 mg, 4.3 mmol, 1.5 eq.), 4-Dimethylaminopyridine (DMAP, 53 mg, 0.43 mmol, 0.15 eq.) and N,N'-dicyclohexylcarbodiimide (DCC, 887 mg, 4.3 mmol, 1.5 eq.) were dissolved in 20 mL of THF

under nitrogen. The reaction mixture was stirred at room temperature for 18 h. The reaction mixture was then filtered, and the residue was purified by silica gel column chromatography (hexane / ethyl acetate = 1 : 3) to obtain 3-(1,3-dimethoxy-1,3-dihydroisobenzofuran-5-yl)-N-(6-hydroxyhexyl)propanamide (compound 5). $^{1}$H NMR (500 MHz, MeOD): δ = 7.34-7.28 (m, 3H), 6.29 (s, 1H), 6.03 (s, 1H), 3.56 (t, 2H), 3.44-3.39 (m, 6H), 3.15 (m, 2H), 3.01 (t, 2H), 2.52 (t, 2H), 1.53 (m, 2H), 1.45 (m, 2H), 1.35 (m, 2H), 1.29 (m, 2H) ppm. $^{13}$C NMR (500 MHz, MeOD): δ = 174.82, 144.31, 140.10, 137.77, 131.20, 124.01, 123.88, 107.85, 106.77, 62.87, 54.72, 54.63, 40.42, 38.81, 33.46, 32.73, 30.31, 27.68, 26.53 ppm. ESI-MS (m/z) $[1 + Na]^+$ 374.2, found: 374.1.

**(g):2-Cyanoethyl(6-(3-(1,3-dimethoxy-1,3-dihydroisobenzofuran-5-yl)propanamido)hexyl)diisopropylphosphoramidite (6)**

Compound 5 (351 mg, 1.0 mmol, 1.0 eq.) was dissolved in 2 mL of anhydrous $CH_2Cl_2$ under dry nitrogen at room temperature, followed by slow addition of N,N'-diisopropylethylamine (DIPEA, 0.6mL ,4.0 mmol ,4.0 eq.). Next, 2-cyanoethyl-N,N'-diisopropylchlorophosphoramidite (0.24 mL, 1.0 mmol, 1.0 eq.) was added dropwise, and the reaction mixture was stirred at room temperature for 4 h. After concentrating by rotary evaporation, the residual oily product was purified by silica gel column chromatography (ethyl acetate : hexane : trimethylamine = 30 : 70 : 2) to obtain 2-cyanoethyl(6-(3-(1,3-dimethoxy-1,3-dihydroisobenzofuran-5-yl)propanamido)hexyl) diisopropylphosphoramidite (OPA phosphoramidite, compound 6). $^{1}$H NMR (500 MHz, MeOD): δ=7.34-7.27 (m, 3H), 6.29 (s, 1H), 6.02 (d, 1H), 3.85-3.83 (m, 2H), 3.76 (t, 1H), 3.68-3.65 (m, 4H), 3.43 (q, 6H), 3.14-3.11 (m, 2H), 3.00 (t, 2H), 2.74 (t, 2H), 2.63 (t, 1H), 2.50 (t, 2H), 2.62-2.59 (m, 2H), 1.43-1.39 (m, 4H), 1.22 (t, 12H). 13 C NMR (500 MHz, MeOD): δ = 174.67, 140.10, 137.77, 131.34, 131.21, 124.00, 123.89, 119.53, 107.83, 106.75, 64.61, 59.74, 59.59, 58.34, 54.73, 54.29, 44.28, 44.18, 40.29, 38.76, 32.74, 32.17, 30.27, 27.54, 26.71, 25.05, 24.99, 21.83 ppm. $^{31}$P NMR (202 MHz, MeOD) δ 147.25 ppm. ESI-MS (m/z)$[1 + Na]^+$ 574.3, found: 574.3.

## Synthesis and purification of oligonucleotides

All oligonucleotides used in this work, including biotin-DNA, SH-DNA and OPA-DNA, were synthesized on a 12-column DNA synthesizer according to the standard DNA synthesis protocol. Each DNA was modified with FITC at the 3' end. After DNA synthesis and

modification, the products were cleaved from the solid support, deprotected using aminomethane and ammonium hydroxide (v/v = 1:1) at 65 °C for 30 min, and then purified by HPLC using a reversed-phase C18 column. The DNA products were desalted using NAP 5 and NAP 10 columns. For OPA-DNA, it was necessary to treat with trifluoroacetic acid and $CH_2Cl_2$ (v/v = 1:4) to achieve the deprotection of aldehyde group before use. The DNA was quantified by using a UV-vis spectrometer, and then stored at -20 °C for later use.

Table S1 DNA sequences used for this work

| name | Sequences |
|---|---|
| **Ctrl DNA** | 5' - TTT TTT TT - FITC - 3' |
| **Biotin-DNA** | 5' - biotin - TTT TTT TT - FITC - 3' |
| **OPA-DNA** | 5' - OPA - TTT TTT T - FITC - 3' |
| **SH-DNA** | 5' - SH - TTT TTT TT - FITC - 3' |
| **Mix-OPA-DNA** | 5' - OPA - TTT TTT TTA TCA AGT CAT - FITC - 3' |
| **$NH_2$-DNA** | 5' - TTT TTT TTT T - $NH_2$ - 3' |
| **S-$NH_2$-DNA** | 5' - AAA AAA AA - $NH_2$ - 3' |
| **L-$NH_2$-DNA** | 5' - ATG ACT TGA TAA AAA AAA - $NH_2$ - 3' |

**HPLC purification**

After removing aminomethane and ammonium hydroxide (v/v = 1:1), the crude DNA products were purified by an Agilent HPLC system with a reversed-phase C18 column. Acetonitrile and 0.01 M triethylamine acetate (TEAA) were chosen as mobile phase, and the gradient elution procedure is shown in Table S2.

Table S2 Gradient elution procedure of HPLC purification

| Component (%) / Time (min) | Acetonitrile | 0.01 M TEAA |
|---|---|---|
| 0 | 0 | 100 |
| 3 | 10 | 90 |
| 35 | 50 | 50 |

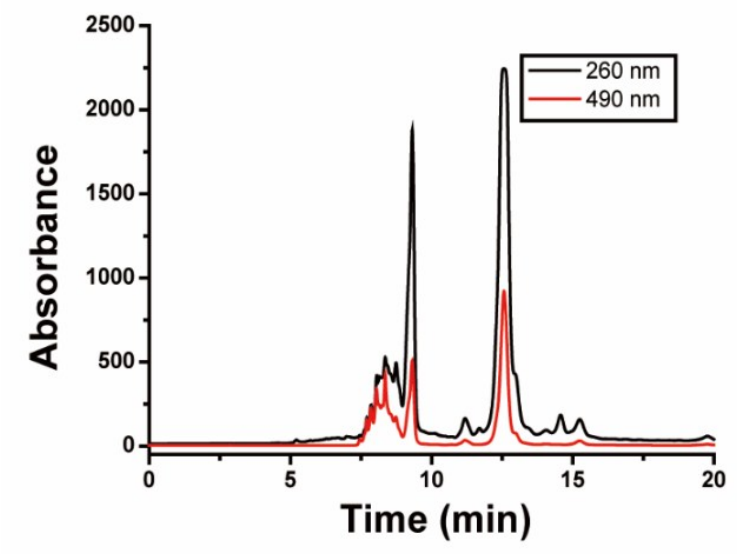



Fig. S1 HPLC chromatogram of OPA-DNA product.

**Feasibility of OPA-DNA coupling with protein**

To demonstrate the conjugation feasibility of OPA-DNA and protein, the conjugation of OPA-DNA and streptavidin sepharose beads (SA beads, 34 μm) was chosen. Three kinds of DNA (OPA-DNA, Biotin-DNA and Ctrl DNA) labelled with FITC were synthesized. For all experiments, DNA was dissolved in PBS buffer. In detail, 1 μM Ctrl DNA, 200 nM and 400 nM Biotin-DNA were each incubated with 200 nM streptavidin on beads. In addition, six concentrations (100 nM, 200 nM, 400 nM, 600 nM, 800 nM and 1 μM) of OPA-DNA were incubated with 200 nM streptavidin on beads, respectively. Then, each sample was incubated in the dark at 25 $^{0}$C for 30 min. After incubation, the samples were washed to remove unreacted DNA for flow cytometry characterization.

**Optimization of the reaction time of phthalimidine formation from OPA-DNA and cytochrome c**

Model proteins, such as cytochrome c from equine heart (purchased from Sigma-Aldrich), were dissolved in PBS buffer (pH 7.4) with final concentration of 20 μM. For all experiments, DNA was dissolved in PBS buffer, pH 7.4 with final reaction concentration of 20 μM in 20 μL reaction system. The reaction mixture was incubated at 25 $^{0}$C for 0 min, 2.5 min, 5 min, 10 min, 20 min, 30 min, and 40 min. After reaction, the reaction was quenched by adding 1 μL of saturated glycine solution. Then, the reaction efficiency was analyzed by 15% SDS-PAGE. The

SDS-PAGE results were visualized using a condensed imaging instrument (fluorescence analysis) and by Coomassie Brilliant Blue staining.

**OPA-DNA coupling with cytochrome c using different OPA-DNA concentrations**

In 20 μL of reaction system, cytochrome c (20 μM) was incubated with different concentrations of OPA-DNA (10 μM, 20 μM, 40 μM, 60 μM, 80 μM, and 100 μM) at 25 $^{0}$C for 30 min in PBS buffer (pH 7.4). Then the reaction was quenched by adding 1 μL of saturated glycine solution. The control DNA was the same sequence as OPA-DNA without modification with OPA. Then, the reaction system was analyzed by 15% SDS-PAGE. The SDS-PAGE results were visualized using a condensed imaging instrument and by Coomassie Brilliant Blue staining.

**OPA-DNA coupling with lysozyme using different OPA-DNA concentrations**

Lysozyme from chicken egg white (purchased from Sigma-Aldrich) was also used as a model protein. In 20 μL of reaction system, lysozyme (20 μM) was incubated with different concentrations of OPA-DNA (10 μM, 20 μM, 40 μM, 60 μM, 80 μM, and 100 μM) at 25 $^{0}$C for 30 min in PBS buffer (pH 7.4). The control DNA was 40 μM. Then the reaction was quenched by adding 1 μL of saturated glycine solution. The reaction system was analyzed by 15% SDS-PAGE (Fig. S2). The SDS-PAGE results were visualized using a condensed imaging instrument and by Coomassie Brilliant Blue staining.

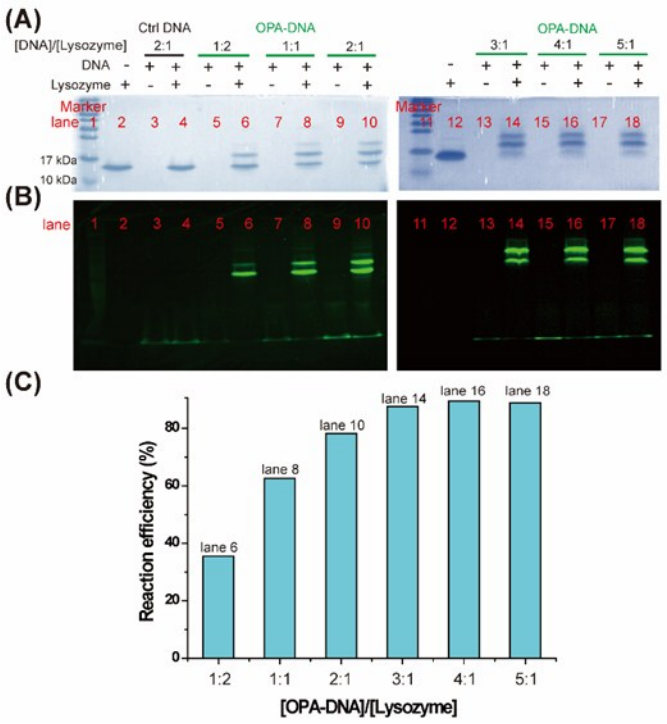



Fig. S2 15% SDS-PAGE analysis of different concentrations of OPA-DNA incubated with lysozyme. The Coomassie Brilliant Blue staining result (A) and fluorescence result (B) of 15% SDS-PAGE. (C) The reaction efficiency of lysozyme was analyzed by image J.

**Mix-OPA-DNA coupling with cytochrome c and lysozyme**

For all experiments, DNA was dissolved in PBS buffer. In detail, 1 μM Ctrl DNA, 800 nM OPA-DNA, 200 nM and 400 nM Biotin-DNA were each incubated with 200 nM streptavidin on beads. Moreover, six concentrations (100 nM, 200 nM, 400 nM, 600 nM, 800 nM and 1 μM) of mix-OPA-DNA were incubated with 200 nM streptavidin on beads. Then, each sample was incubated in the dark at 25 $^{0}$C for 30 min. After incubation, the samples were washed to remove unreacted DNA for flow cytometry characterization.

**Mix-OPA-DNA coupling with cytochrome c and lysozyme**

Mix-OPA-DNA was used for bioconjugation with cytochrome c. In 20 μL of reaction system, Mix-OPA-DNA (10 μM and 20 μM) was incubated with 20 μM cytochrome c and lysozyme at 25 $^{0}$C for 30 min in PBS buffer (pH 7.4). Then the reaction was quenched by adding 1 μL of saturated glycine solution. The coupling result was analyzed by 15% SDS-PAGE. The SDS-PAGE results were visualized using a condensed imaging instrument and by Coomassie Brilliant Blue staining. The results indicated that the coupling efficiency of Mix-

OPA-DNA was obviously lower than the coupling efficiency of OPA-DNA with cytochrome c or lysozyme (Fig. S3, ESI†). The possible cause of this result is that both the DNA base sequence and the length of OPA-DNA influenced the coupling efficiency of OPA-DNA to cytochrome c and lysozyme.

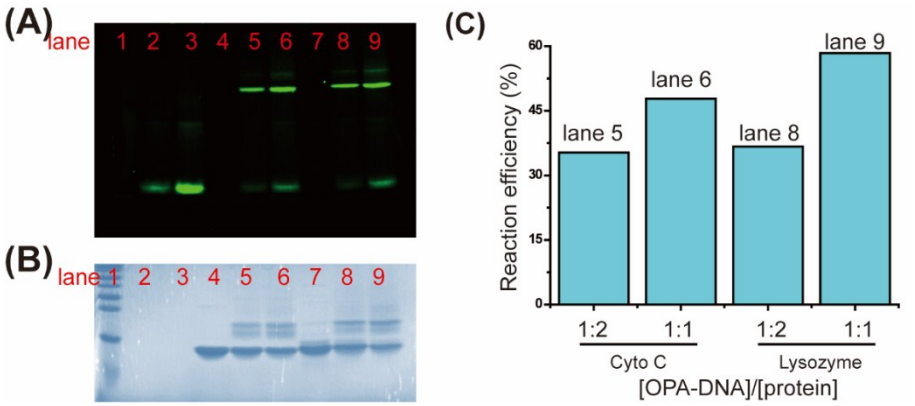
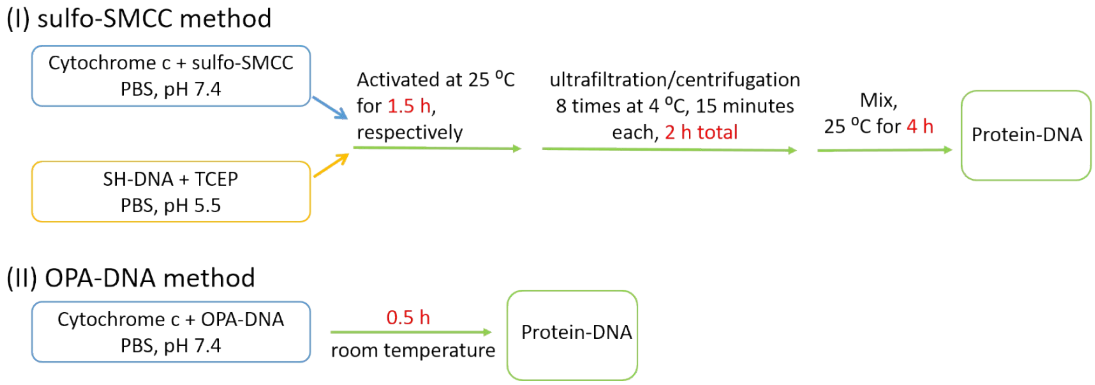



Fig. S3 15% SDS-PAGE analysis of Mix-OPA-DNA coupling with cytochrome c and lysozyme. The results of SDS-PAGE analysis were visualized using a condensed imaging instrument (A) and Coomassie Blue (B). (C) The reaction efficiency of cytochrome c was analyzed by image J.

**Method comparison**




We used sulfo-SMCC for bioconjugation of SH-DNA and protein for method comparison. Briefly, cytochrome c was mixed with the appropriate sulfo-SMCC in PBS and incubated at 25 $^0$C for 1.5 h. The excess sulfo-SMCC was removed by ultrafiltration centrifugation 8 times using phosphate buffer ($Na_2HPO_4$ 8 mM, NaCl 137 mM, $NaH_2PO_4$ 2 mM, pH 7.4). Meanwhile, SH-DNA was activated by tris (2-chloroethyl) phosphate (TCEP) at 25 $^0$C in 50 mM sodium phosphate buffer (pH 5.5) for 1.5 h, and then was treated by ultrafiltration centrifugation 8

times to remove TCEP at 4 $^0$C. The above solutions of sulfo-SMCC-activated cytochrome c and TCEP-activated SH-DNA were mixed together and incubated at 25 $^0$C for 4 h. On the other hand, OPA-DNA (20 μM) was incubated with cytochrome c (20 μM) in PBS (pH 7.4) for 30 min. Then, the reaction result was analyzed by 15% SDS-PAGE, which was visualized using a condensed imaging instrument and by Coomassie Brilliant Blue staining.

**OPA-DNA coupling with amino modified DNA ($NH_2$-DNA) in different concentrations of sodium chloride**

OPA-DNA reacted with $NH_2$-DNA in different concentrations of sodium chloride. Specifically, 20 μM OPA-DNA was incubated with 100 μM $NH_2$-DNA in 20 μL reaction system with different concentrations of sodium chloride (137 mM, 500 mM, 750 mM, 1 M, 1.5 M and 2 M) in the dark at 25 $^0$C for 30 min. Then the reaction was quenched by adding 1 μL of saturated glycine solution. The samples were analyzed by 15% denatured PAGE. The results were visualized using a condensed imaging instrument.

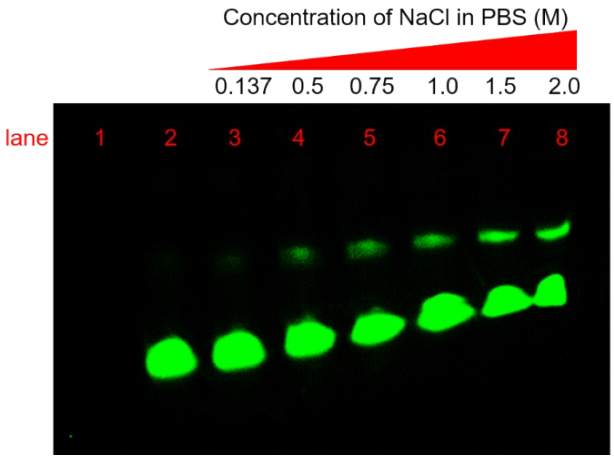



Fig. S4 15% denatured PAGE analysis of OPA-DNA coupling with $NH_2$-DNA. The results were visualized using a condensed imaging instrument for fluorescent analysis. Lane 1: $NH_2$-DNA, lane 2: OPA-DNA, lane 3-8: $NH_2$-DNA + OPA-DNA with different concentrations of NaCl.

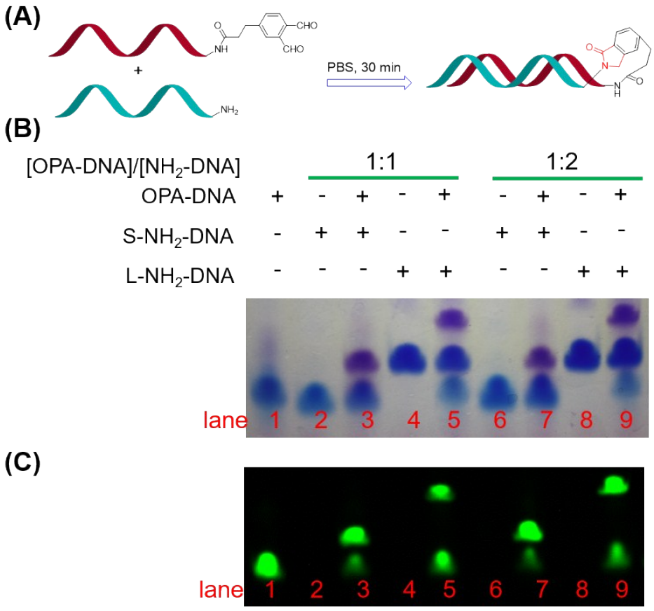



Fig. S5 (A) Schematic illustration of OPA-DNA coupling with amino-modified cDNA. 15% denatured PAGE analysis of the coupling efficiency of OPA-DNA (labelled with FITC) with two different concentrations of $NH_2$-DNA. The PAGE results were characterized by Stains-All (B) and condensed imaging (C).

**OPA-DNA and mix-OPA-DNA coupling with amino modified cDNA (S-$NH_2$-DNA and L-$NH_2$-DNA)**

OPA-DNA was used to react with two different $NH_2$-DNA sequences (S-$NH_2$-DNA and L-$NH_2$-DNA). Specifically, in 20 μL reaction system, 20 μM OPA-DNA was incubated with S-$NH_2$-DNA or L-$NH_2$-DNA (both 20 μM and 40 μM) in the dark at 25 $^0$C for 30 min.. Then the reaction was quenched by adding 1 μL of saturated glycine solution. The samples were analyzed by 15% denatured PAGE visualized using a condensed imaging instrument and with Stains-All. The coupling reaction of Mix-OPA-DNA and two different $NH_2$-DNA sequences (S-$NH_2$-DNA and L-$NH_2$-DNA) was the same as the reaction of OPA-DNA.

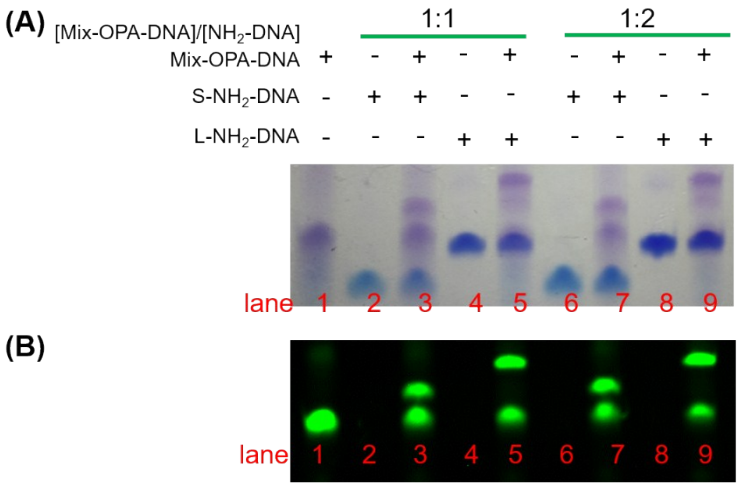



Fig. S6 15% denatured PAGE analysis of Mix-OPA-DNA coupling with $NH_2$-DNA. The PAGE results were visualized using Stains-All (A) and using a condensed imaging instrument (B).