Supporting Information

Efficient construction of stable linear gene based on TNA loop modified primer pair for gene delivery

Xuehe Lu,[ab] Xiaohui Wu,[bc] Tiantian Wu,[bc] Lin Han,[ab] Jianbing Liu*[bc] and Baoquan Ding*[abc]

[a]School of Materials Science and Engineering, Zhengzhou University, Zhengzhou 450001, China.

[b]CAS Key Laboratory of Nanosystem and Hierarchical Fabrication, CAS Center for Excellence in Nanoscience, National Center for Nanoscience and Technology, Beijing 100190, China.

[c]University of Chinese Academy of Sciences, Beijing 100049, China

Corresponding author: liujb@nanoctr.cn, dingbq@nanoctr.cn

TABLE OF CONTENTS

SUPPLEMENTAL MATERIALS AND METHODS

1. General information

Materials and agents: $_{L}$-Threonic acid calcium salt, *N*-(2-oxo-1,2-dihydropyrimidin-4-yl)benzamide, thymine, trimethylsilyl trifluoromethanesulfonate, trimethylsilyl trifluoromethanesulfonate, tetrabutylammoniumfluoride (1.0 M in tetrahydrofuran (THF)), *N*-(9H-purin-6-yl)benzamide, methyl alcohol, and 2-cyanoethyl N,N,N',N'-tetraisopropylphosphoramidite were purchased from Bide Pharmatech Ltd. (Shanghai, China). Benzoyl chloride was purchased from 3A Chemical (Shanghai, China). Tetrazole was purchased from Innochem (Beijing, China). *Para*-toluenesulfonic acid monohydrate and diisobutylaluminum hydride solution (1.0 M in toluene) were purchased from HWRK CHEM (Beijing, China). Imidazole was purchased from HEOWNS (Tianjin, China). Anhydrous oxalic acid, 4, 4′-dimethoxytrityl chloride, anhydrous acetonitrile, acetic anhydride pyridine, 4-dimethylaminopyridine, anhydrous dichloromethane, anhydrous dimethyl ether, and tetrahydrofuran were purchased from Shanghai Aladdin Bio-Chem (Shanghai, China). *N*, *O*-bis(trimethylsilyl)acetamide was purchased from Energy Chemical (Shanghai, China). Stains-All was purchased from OKA^{TM} (Beijing, China). Acrylamide/bis-acrylamide solution (19:1, 40%) was purchased from Bio-Rad, Agaroses were purchased from SIGMA. The La-Taq DNA polymerase was purchased from TaKaRa (China). The T4 DNA Ligase was purchased from NEB (USA). Gel extraction kit was purchased from TaKaRa (China). Trypsin and fetal bovine serum (FBS) were purchased from Sangon Biotech Co., Ltd. (Shanghai, China). HEK-293T cell was purchased from the Cell Center at the Institute of Basic Medical Sciences, Chinese Academy of Medical Science.

Characterizations: NMR spectra were recorded on Bruker Avance III (400 MHz), chemical shifts (δ) are given in ppm relative to TMS, coupling constants (J) in Hz. Fluorescence microscope images were detected by LeiCa. Flow cytometry analysis was detected by ACEA NovoCyte^{TM}.

2. Experiment section

2.1 Synthesis of TNA monomers

Four α-$_L$-threofuranosyl nucleoside monomers **7a-d** were synthesized by the Chaput's approach.[1] The general synthetic route is displayed in **Scheme S1**. TNA monomers (TNA-A, TNA-T, TNA-G, and TNA-C) were characterized by ^{1}H, ^{13}C, ^{31}P NMR and ESI-MS spectroscopies.

2.2 Synthesis and characterization of F^2, R^2, F^3, and R^3

General procedures for solid-phase synthesis of F^2, R^2, F^3, and R^3: F^2, R^2, F^3, and R^3 were synthesized based on the Chaput's approach.[2] Synthesis of TNA sequences of those primers was performed in 1 μmole scale based on the 1000 Å universal CPG solid support. The synthetic program possesses a prolonged detritylation time (3 min×2), coupling time (15 min×2), capping time (1 min×3), and oxidation time (1 min×2). The DNA sequences were synthesized by the general procedure. The crude products of F^2, R^2, F^3, and R^3 were obtained by removing protection groups in concentrated ammonium hydroxide at 55 ℃ for 24 h.

Characterization: The crude products of F^2, R^2, F^3, and R^3 were desalted by SEP-PAK classic C18 column and further purified by 15% denature PAGE. The molecular weights of F^2, R^2, F^3 and R^3 were characterized by MALDI-TOF-MS using aqueous saturated 3-hydroxypicolinc acid solution as the matrix. F^2: m/z calcd 5767, found 5770; R^2: m/z calcd 5643, found 5647; F^3: m/z calcd 15115, found 15182; R^3: m/z calcd 14992, found 15053.

2.3 PCR progress

PCR content: 50.0 ng linear CMV-EGFP template, 0.5 μM primer pair of F^1 and R^1, F^2 and R^2, F^3 and R^3, 0.2 mM each dNTP, 1 U La-Taq DNA polymerase, 2.5 μL 10×PCR reaction buffer and added ddH_2O into 25 μL PCR system.

PCR program: 95 ℃ 5 min - (95 ℃ 30 s - 60 ℃ 30 s - 72 ^{o}C 2 min) × 30 cycles - 72 ℃ 5min - 4 ℃ 1 h.

Purification: After purified by gel extraction Kit, the PCR products were directly employed for the following experiments.

2.4 T4 DNA ligation

Synthesis of terminal-closed linear EGFP gene: after obtaining the TNA loop modified EGFP gene utilizing F^3 and R^3 as primer pair, terminal-closed linear EGFP gene was effectively fabricated by T4 DNA ligation at 25 ℃ for 5.0 h.

2.5 Exonuclease resistance

The 150 ng naked linear EGFP gene (F^1+R^1) or terminal-closed linear EGFP gene (F^3+R^3) was incubated with the 0.0 U, 2.5 U, 5.0 U, 7.5 U, and 10.0 U of lambda exonuclease at 37 ℃ for 1.5 h, respectively.

2.6 Serum stability

The 150 ng naked linear EGFP gene (F^1+R^1) or terminal-closed linear EGFP gene (F^3+R^3) was incubated with 10% FBS at 37 ℃ for 0.0 h, 0.5 h, 1.0 h, 2.0 h, 4.0 h, and 6.0 h, respectively.

2.7 Cell culture

HEK-293T cells were cultured in DMEM (Hyclone, Thermo Scientific) complete medium supplemented with 10% FBS (Hyclone, Thermo Scientific), penicillin, and streptomycin (GIBICO, Invitrogen) in an atmosphere of 5% CO_2 at 37 ℃.

2.8 Expression of EGFP protein

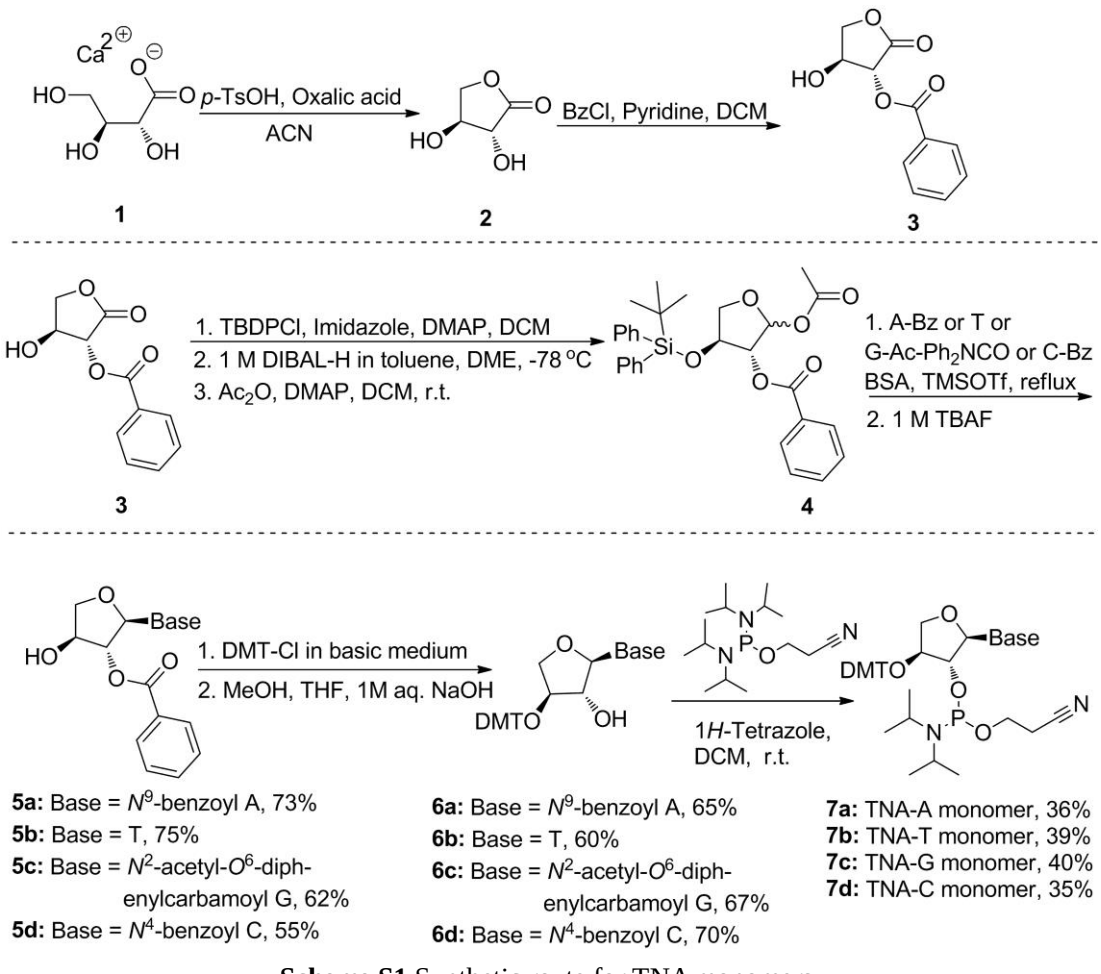
The 0.5 pmol naked linear EGFP gene (F^1+R^1) or terminal-closed linear EGFP gene (F^3+R^3) was transfected to HEK-293T cells by Lipofectamine 2000 reagent (Invitrogen, Thermo Scientific), respectively. After incubation for 12 h, 24 h, 36 h, 48 h, and 60 h, the cells were imaged by a fluorescence microscope with an excitation wavelength of 488 nm, respectively.

2.9 Flow cytometry analysis

The HEK-293T cells with the indicated treatments were harvested after the incubation at 37 ℃ for 60 h. After washing three times with PBS, cells were analyzed by a flow cytometer (ACEA NovoCyte^{TM}). The fluorescence signal was excited at the wavelength of 488 nm.

2.10 Statistical analysis

The error bars represent the standard error of the mean of three independent experiments. Statistical analysis was conducted using Prism 6.0 (GraphPad, San Diego, CA, USA). A value of $P < 0.05$ was considered statistically significant.

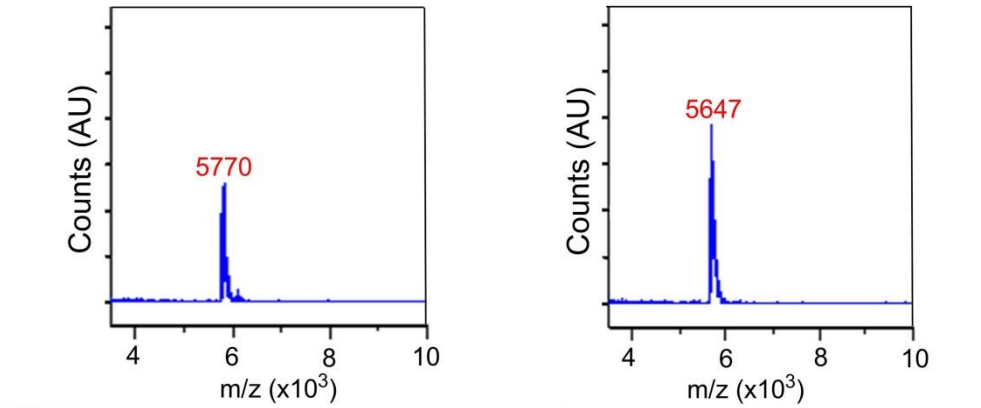


5a: Base = N^9-benzoyl A, 73%
5b: Base = T, 75%
5c: Base = N^2-acetyl-O^6-diphenylcarbamoyl G, 62%
5d: Base = N^4-benzoyl C, 55%

6a: Base = N^9-benzoyl A, 65%
6b: Base = T, 60%
6c: Base = N^2-acetyl-O^6-diphenylcarbamoyl G, 67%
6d: Base = N^4-benzoyl C, 70%

7a: TNA-A monomer, 36%
7b: TNA-T monomer, 39%
7c: TNA-G monomer, 40%
7d: TNA-C monomer, 35%

Scheme S1 Synthetic route for TNA monomers.



F^2(TNA) : m/z calcd 5767; found 5770

R^2(TNA) : m/z calcd 5643; found 5647

Fig.S1 MALDI-TOF-MS analysis of F^2 and R^2.

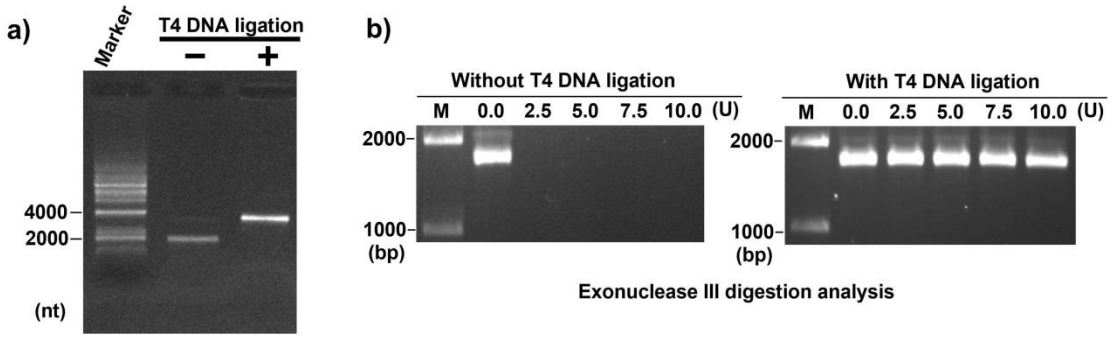


Fig.S2 Evaluation of the T4 DNA ligation efficiency. (a) 1% denatured agarose gel electrophoresis analysis of the PCR products based on the primer pair (F^3 and R^3) without or with T4 DNA ligation. (b) 1% agarose gel electrophoresis analysis for the exonuclease III digestion of the PCR products based on the primer pair (F^3 and R^3) without or with T4 DNA ligation. (M: DNA marker).

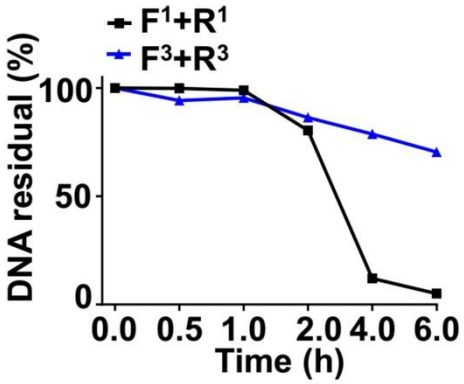


Fig.S3 Statistic results of serum stability of naked linear EGFP gene (F^1+R^1) and terminal-closed linear EGFP gene (F^3+R^3) by Image J analysis.

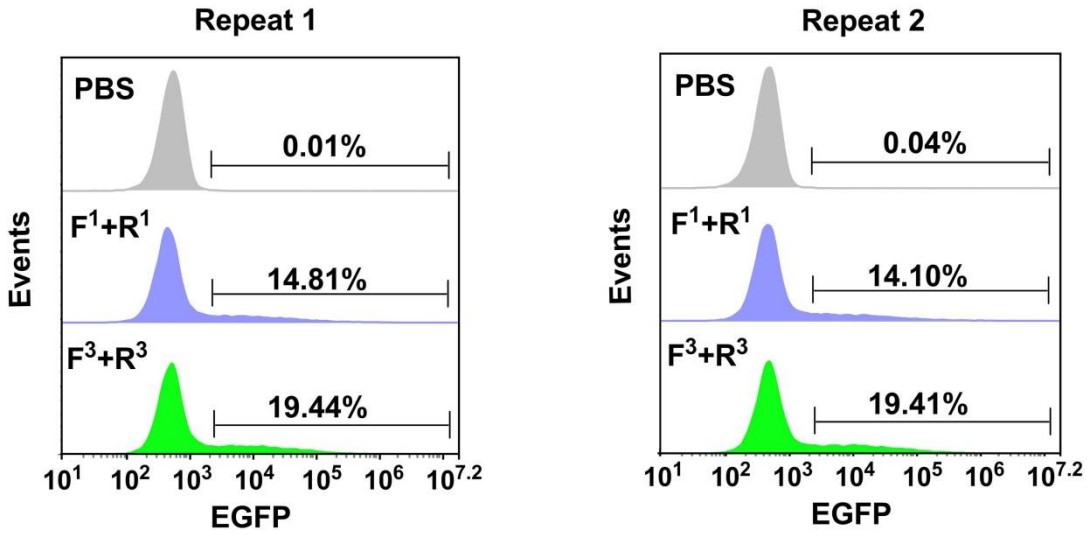


Fig.S4 The flow cytometry analysis of EGFP expression level in HEK-293T cells transfected with PBS, naked linear EGFP gene (F^1+R^1), and terminal-closed linear EGFP gene (F^3+R^3).

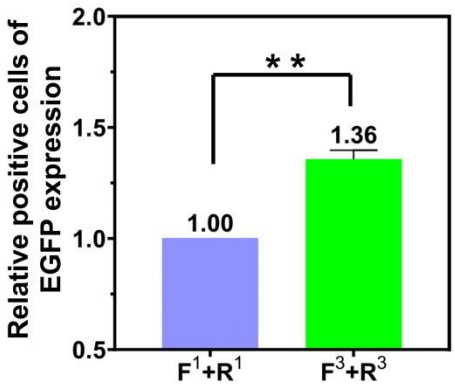


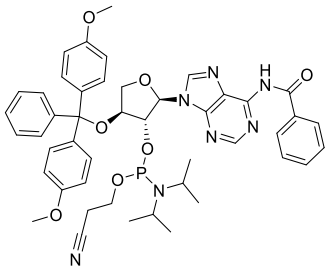
Fig.S5 The flow cytometry analysis of relative EGFP positive cells (**$P < 0.01$).

Table S1 DNA sequences of the three groups of primers and CMV-EGFP gene.

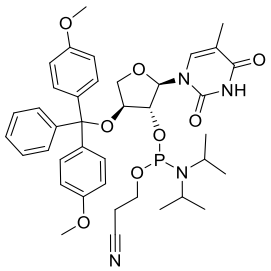
Name	Sequence (5'-3')
F^1 (DNA)	**CCTCCCAAGATGGCCCGATA**
R^1 (DNA)	**CCTCCTCTTCCCTGTCCAAA**
F^2 (TNA)	**CCTCCCAAGATGGCCCGATA**
R^2 (TNA)	**CCTCCTCTTCCCTGTCCAAA**
F^3 (TNA loop)	**/Phos/GCGC**GGCGCTCAGTTGGGAGAGCGCCGCGCCCTCCCAAGATGGCCCGATA
R^3 (TNA loop)	**/Phos/GCGC**GGCGCTCAGTTGGGAGAGCGCCGCGCCCTCCTCTTCCCTGTCCAAA
EGFP template	CCTCCCAAGATGGCCCGATAATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCA TATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACG ACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTT CCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTG TATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATT ATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCAT CGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGAC TCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAA ATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAG GCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAG CGCTACCGGACTCAGATCTCGAGCTCAAGCTTCGAATTCTGCAGTCGACGGTACCGCGGGC CCGGGATCCACCGGTCGCCACC**ATG**GTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTG CCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGG GCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCT GCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGC TACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCC AGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTT CGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGC AACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCG ACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAG CGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTG CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCG ATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCT GTACAAG**TAA**AGCGGCCGCGACTCTAGATCATAATCAGCCATACCACATTTGTAGAGGTTT TACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAAT TGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACA AATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCA ATGTATCTTAAG**TTTGGACAGGGAAGAGGAGG**

Note: TNA sequences are highlighted in red color.

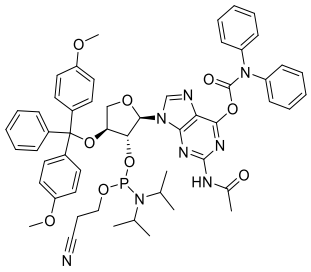
^{1}H NMR, ^{13}C NMR, ^{31}P NMR and MS of TNA monomers (7a-d)



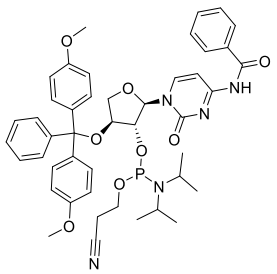
(2R, 3R, 4S)-2-(6-benzamido-9H-purin-9-yl)-4-(bis(4-methoxyphenyl)(phenyl)methoxy)tetrahydrofuran-3-yl (2-cyanoethyl) diisopropylphosphoramidite(**7a**): ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.11 (s, 1H), 8.81 (s, 1H), 8.35 (s, 1H), 8.04 (d, *J* = 7.5 Hz, 2H), 7.67 – 7.58 (m, 1H), 7.57 – 7.50 (m, 2H), 7.25 – 7.17 (m, 5H), 7.12 (dd, *J* = 18.9, 8.4 Hz, 4H), 6.82 – 6.76 (m, 4H), 6.16 (s, 1H), 4.95 (d, *J* = 9.8 Hz, 1H), 4.36 (s, 1H), 3.85 (dd, *J* = 9.9, 3.9 Hz, 1H), 3.76 (s, 6H), 3.68 – 3.64 (m, 2H), 3.63 – 3.51 (m, 2H), 3.44 (d, *J* = 10.0 Hz, 1H), 2.49 – 2.31 (m, 2H), 1.20 (s, 12H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 164.52, 158.78, 158.73, 152.52, 151.36, 149.25, 144.36, 141.72, 135.71, 135.45, 133.81, 132.75, 129.99, 129.77, 128.89, 128.11, 127.93, 127.81, 127.13, 123.30, 117.38, 113.48, 113.45, 90.23, 90.16, 88.10, 81.50, 81.33, 77.84, 77.81, 75.04, 58.46, 58.27, 55.24, 55.22, 43.52, 43.39, 24.68, 24.61, 24.46, 24.39, 20.03, 19.95. ^{31}P NMR (162 MHz, $CDCl_3$) δ 151.53. HRMS (ESI) calcd for $(C_{46}H_{50}N_7NaO_7P)^+$ 866.3402, found 866.3399.



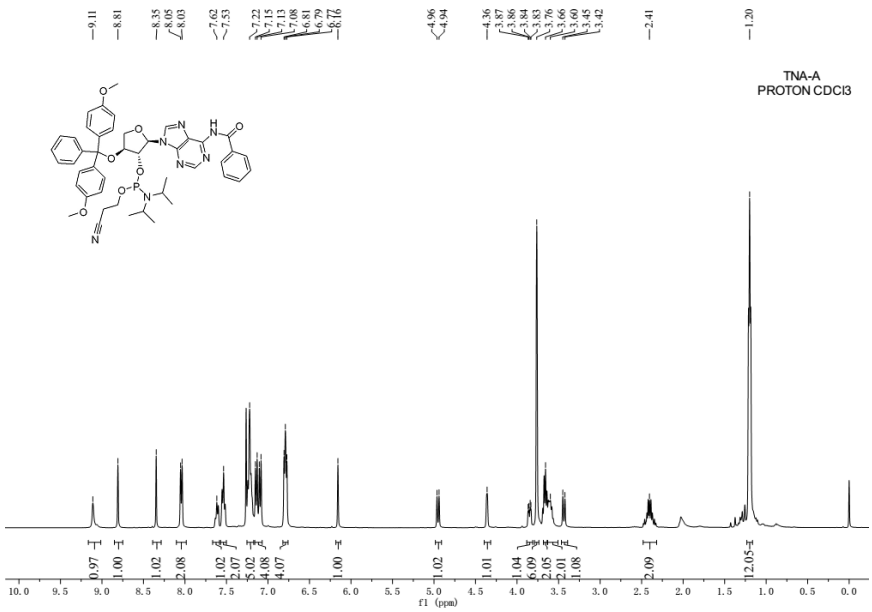
(2R, 3R, 4S)-4-(bis(4-methoxyphenyl)(phenyl)methoxy)-2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)tetrahydrofuran-3-yl (2-cyanoethyl) diisopropylphosphoramidite(**7b**): ^{1}H NMR (400 MHz, $CDCl_3$) δ 9.05 (s, 1H), 7.47 (s, 1H), 7.37 (d, *J* = 7.6 Hz, 2H), 7.32 – 7.24 (m, 7H), 6.85 (d, *J* = 8.4 Hz, 4H), 5.86 (s, 1H), 4.27 (d, *J* = 8.1 Hz, 1H), 4.16 (s, 1H), 3.97 – 3.88 (m, 1H), 3.81 (s, 6H), 3.80 – 3.73 (m, 2H), 3.63 – 3.51 (m, 2H), 3.35 (d, *J* = 9.9 Hz, 1H), 2.69 (t, *J* = 6.1 Hz, 2H), 1.81 (s, 3H), 1.16 (d, *J* = 6.7 Hz, 6H), 1.03 (d, *J* = 6.7 Hz, 6H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 164.06, 158.82, 150.48, 144.54, 137.03, 135.59, 135.49, 129.78, 128.25, 127.66, 127.17, 118.09, 113.72, 113.65, 109.39, 91.52, 91.49, 88.29, 82.23, 82.03, 77.75, 77.69, 75.01, 58.56, 58.37, 55.30, 43.61, 43.48, 24.53, 24.47, 24.45, 24.40, 20.27, 20.20, 12.68. ^{31}P NMR (162 MHz, $CDCl_3$) δ 152.22. HRMS (ESI) calcd for $(C_{39}H_{47}N_4NaO_8P)^+$ 753.3024, found 753.3028.

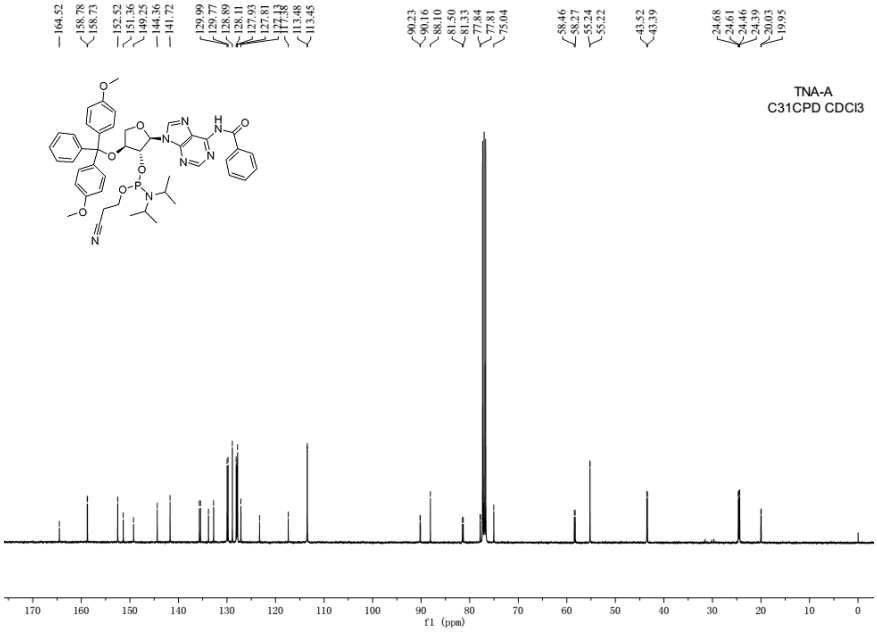


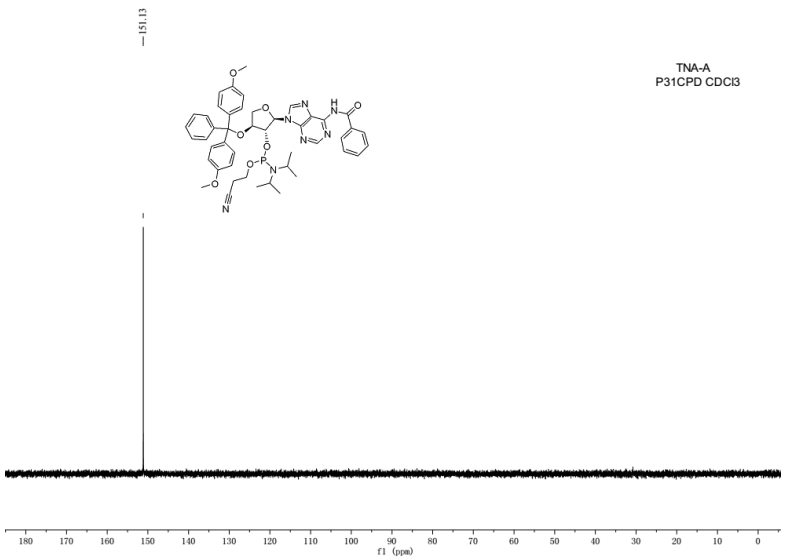
2-acetamido-9-((2R,3R,4S)-4-(bis(4-methoxyphenyl)(phenyl)methoxy)-3-(((2-cyanoethoxy)(diisopropylamino)phosphanyl)oxy)tetrahydrofuran-2-yl)-9H-purin-6-yl diphenylcarbamate(**7c**): ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.26 (d, *J* = 5.3 Hz, 1H), 7.99 (d, *J* = 48.6 Hz, 1H), 7.45 (s, 4H), 7.37 (t, *J* = 7.4 Hz, 4H), 7.21 (s, 10H), 6.79 (d, *J* = 7.3 Hz, 4H), 6.02 (d, *J* = 27.4 Hz, 1H), 4.76 (dd, *J* = 34.5, 8.2 Hz, 1H), 4.38 (d, *J* = 48.2 Hz, 1H), 3.84 (dd, *J* = 9.6, 3.6 Hz, 1H), 3.74 – 3.65 (m, 6H), 3.64 – 3.51 (m, 4H), 3.39 (d, *J* = 9.8 Hz, 1H), 2.58 (d, *J* = 34.4 Hz, 3H), 2.45 (d, *J* = 5.7 Hz, 1H), 2.24 (d, *J* = 37.2 Hz, 1H), 1.20 – 1.08 (m, 12H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 171.48, 170.66, 158.82, 158.78, 158.73, 156.01, 154.51, 154.22, 151.99, 151.97, 150.43, 150.38, 144.37, 144.32, 143.17, 142.78, 141.81, 135.69, 135.63, 135.49, 135.39, 130.01, 129.91, 129.78, 129.69, 129.17, 128.13, 127.95, 127.84, 127.10, 121.37, 117.58, 117.30, 113.52, 113.50, 90.55, 90.47, 89.73, 88.10, 87.95, 81.63, 81.45, 81.41, 81.26, 77.74, 77.68, 75.26, 74.38, 58.03, 57.94, 57.85, 57.76, 55.23, 55.20, 43.47, 43.35, 25.48, 25.45, 25.29, 24.55, 24.49, 20.12, 20.05, 19.80, 19.72. ^{31}P NMR (162 MHz, $CDCl_3$) δ 151.00. HRMS (ESI) calcd for $(C_{39}H_{47}N_4NaO_8P)^+$ 1015.3878, found 1015.3872.

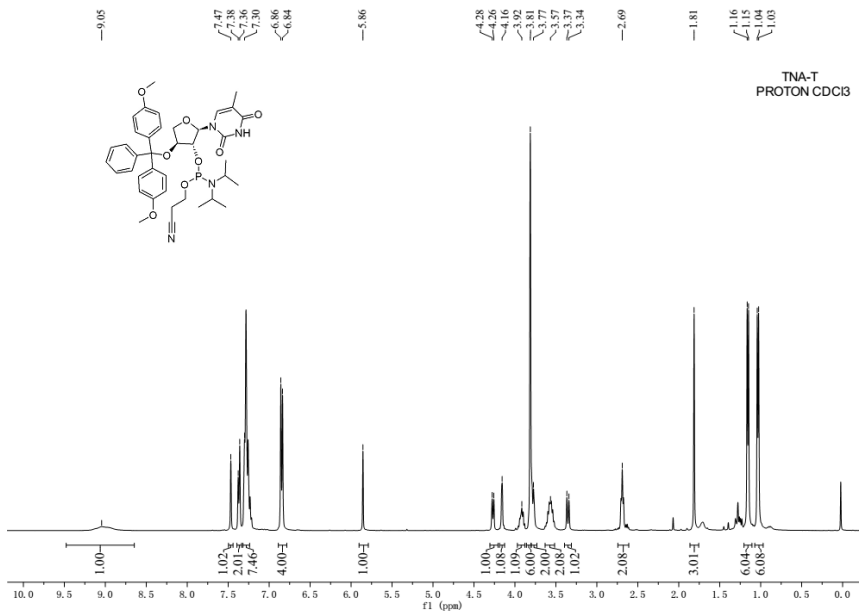


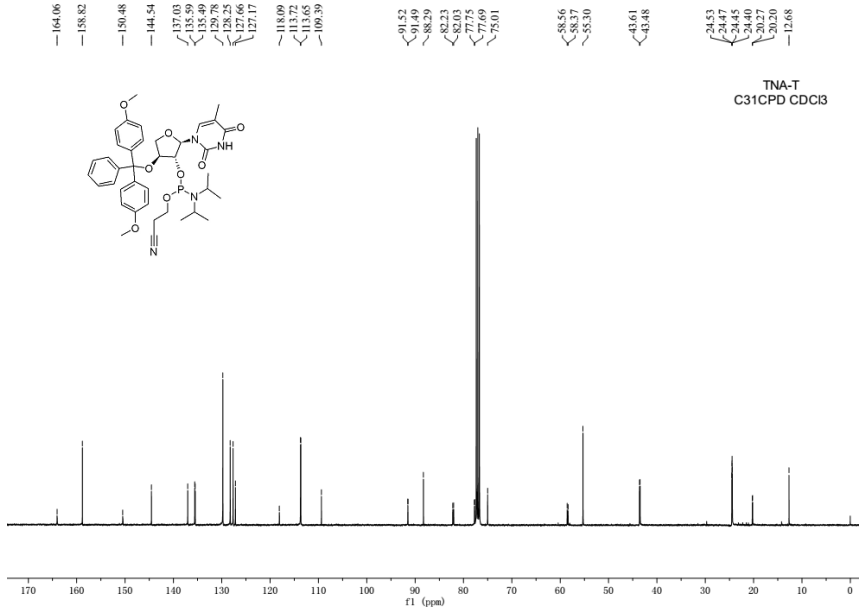
(2R, 3R, 4S)-2-(4-benzamido-2-oxopyrimidin-1(2H)-yl)-4-(bis(4-methoxyphenyl)(phenyl)methoxy)tetrahydrofuran-3-yl (2-cyanoethyl) diisopropylphosphoramidite (**7d**):^{1}H NMR (400 MHz, $CDCl_3$) δ 8.71 (s, 1H), 8.05 (d, *J* = 7.3 Hz, 1H), 7.94 (d, *J* = 7.3 Hz, 2H), 7.68 – 7.62 (m, 1H), 7.60 – 7.48 (m, 3H), 7.28 – 7.14 (m, 8H), 6.82 (d, *J* = 8.3 Hz, 4H), 5.88 (s, 1H), 4.57 (d, *J* = 8.3 Hz, 1H), 4.13 (s, 1H), 4.11 – 4.03 (m, 1H), 3.95 – 3.86 (m, 2H), 3.77 (s, 6H), 3.62 – 3.53 (m, 2H), 3.31 (d, *J* = 9.9 Hz, 1H), 2.88 – 2.78 (m, 1H), 2.77 – 2.69 (m, 1H), 1.16 (d, *J* = 6.6 Hz, 6H), 1.01 (d, *J* = 6.6 Hz, 6H). ^{13}C NMR (100 MHz, $CDCl_3$) δ 162.18, 158.80, 145.70, 144.37, 135.64, 135.50, 133.17, 129.83, 129.79, 129.10, 128.12, 127.78, 127.47, 127.17, 118.30, 113.57, 113.51, 93.16, 93.12, 88.33, 82.08, 81.87, 77.64, 77.59, 75.78, 58.88, 58.69, 55.22, 43.71, 43.58, 24.47, 24.44, 24.39, 24.37, 20.35, 20.27. ^{31}P NMR (162 MHz, $CDCl_3$) δ 152.63. HRMS (ESI) calcd for $(C_{45}H_{50}N_5NaO_8P)^+$ 842.3289, found 842.3283.

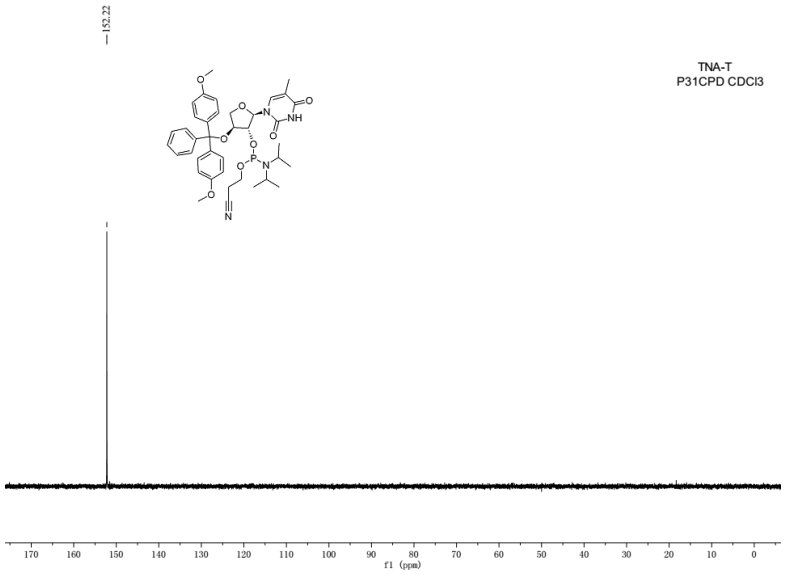

TNA-A
PROTON CDCl3
f1 (ppm)

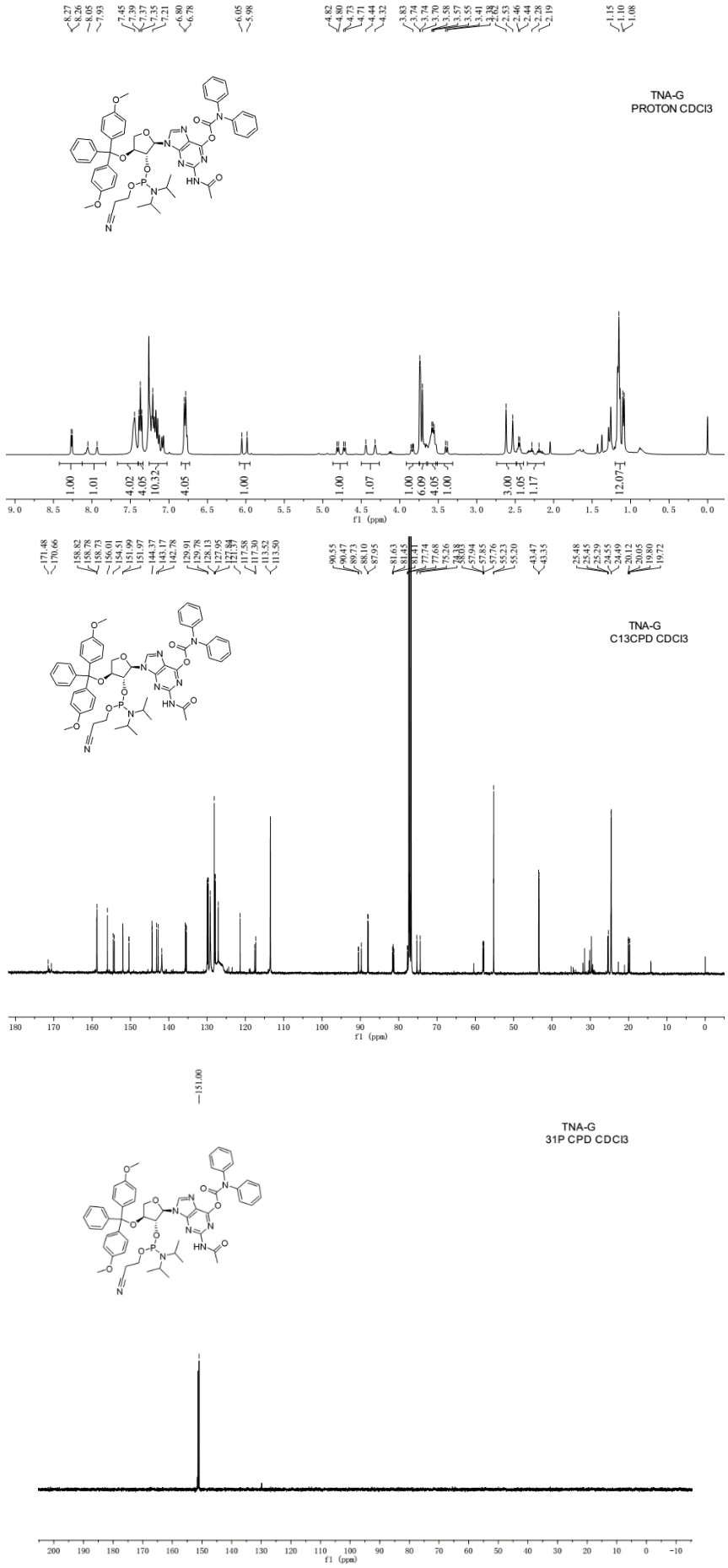

TNA-A
C31CPD CDCl3
f1 (ppm)

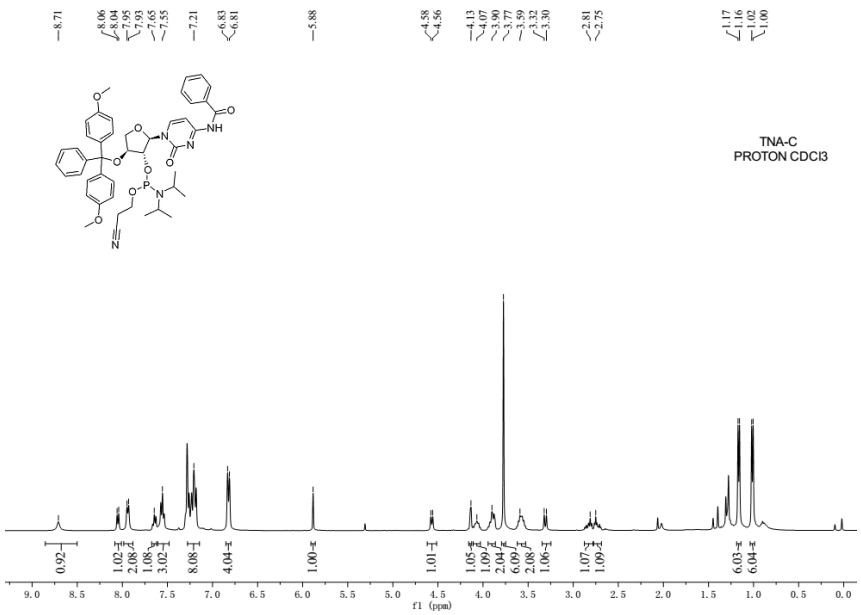

TNA-A
P31CPD CDCl3
f1 (ppm)

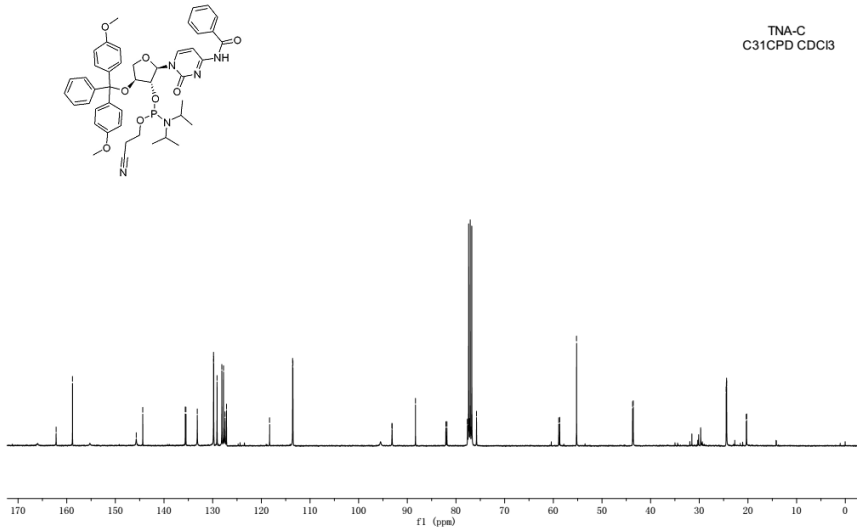

TNA-T
PROTON CDCl3
f1 (ppm)

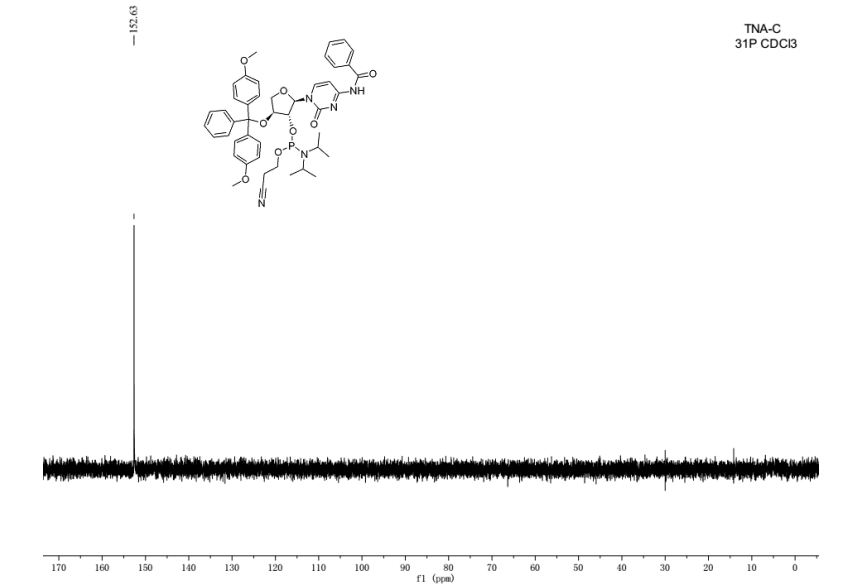

TNA-T
C31CPD CDCl3
f1 (ppm)


TNA-T
P31CPD CDCl3
f1 (ppm)


TNA-G
PROTON CDCl3
TNA-G
C13CPD CDCl3
TNA-G
31P CPD CDCl3


TNA-C
PROTON CDCl3
f1 (ppm)


TNA-C
C31CPD CDCl3
f1 (ppm)


TNA-C
31P CDCl3
f1 (ppm)

SUPPLEMENTAL REFERENCES

[1] S. P. Sau, N. E. Fahmi, J. Y. Liao, S. Bala and J. C. Chaput, *J. Org. Chem.*, 2016, **81**, 2302.

[2] S. Zhang, J. C. Chaput, *Current Protocols in Nucleic Acid Chemistry*, 2012, **4**, 4.51.1.