# Supplemental Information

## Synthesis and Polymerase Activity of a Fluorescent Cytidine TNA Triphosphate Analogue

Hui Mei[#], Changhua Shi[#], Randi M. Jimenez, Yajun Wang,

Miramar Kardouh, and John C. Chaput*

Department of Pharmaceutical Sciences, University of California, Irvine, CA 92697-3958.

* To whom correspondence should be addressed. E-mail: jchaput@uci.edu

[#] These authors contributed equally to the manuscript

Downloaded from https://academic.oup.com/nar/article/45/10/5629/3796333 by guest on 25 April 2024

# Table of contents



Downloaded from https://academic.oup.com/nar/article/45/10/5629/3796333 by guest on 25 April 2024

**Expression and Purification of Kod-RI.**

Kod-RI was generated as previously described. Briefly, the Kod-RI gene was sub-cloned into a pET21 vector and transformed into an Acella® cell (Edge BioSystems, MD) strain for protein expression. A single colony from the transformation was used for 10 ml LB overnight culture and then the overnight culture was diluted with 1:100 into the fresh LB for the large scale expression. The cells were induced with 1 mM of IPTG for 20 hat 18 ℃. Once the cell density reached (OD600) 0.8, the cell pelleted, re-suspended in a lysis buffer (10 mM Tris.Cl pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol) and lysed by sonication. Then the cell lysate was clarified by centrifugation, and the clear lysate incubated for 20 min at 70 ℃. The lysate was cooled to 24 ℃ and clarified by centrifugation to remove the denatured *E. coli* proteins. The clarified cell lysate was purified using a HiTrap Q-HiTrap heparin coupled column system with FPLC (GE, Marlborough, MA). Using this system, the Q column captures the nucleic acid from the cell lysate and the heparin column captures the polymerase. After sample loading, the Q column was decoupled from the heparin column and the polymerase was eluted from heparin column using a mobile phase of (10 mM Tris.Cl pH 8.5, 1M NaCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol). The elution fractions were pooled together and then dialyzed into the protein storage buffer (10 mM Tris.Cl pH 8.5, 200 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol). The purified protein was analyzed by SDS PAGE and UV quantified.

**X-ray crystallography of compound 5**

An analytical sample of **5** was crystallized from methanol and a trace amout of water. The yellow crystal of dimensions 0.212 x 0.240 x 0.380 mm was mounted in a cryoloop and transferred to a Bruker SMART APEX II diffractometer. The APEX2[1] program package was used to determine the unit-cell parameters and for data collection (10 sec/frame scan time for a sphere of diffraction data). The raw frame data was processed using SAINT[2] and SADABS[3] to yield the reflection data file. Subsequent calculations were carried out using the SHELXTL[4] program. The diffraction symmetry was *mmm* and the systematic absences were consistent with the orthorhombic space group $P2_12_12_1$ that was later determined to be correct.

The structure was solved by direct methods and refined on $F^2$ by full-matrix least-squares techniques. The analytical scattering factors[5] for neutral atoms were used throughout the



Downloaded from https://academic.oup.com/nar/article/45/10/5629/3796333 by guest on 25 April 2024

analysis. Hydrogen atoms were located from a difference-Fourier map and refined (x,y,z, riding $U_{iso}$). There was one molecule of water present per formula-unit.

At convergence, wR2 = 0.0683 and Goof = 1.066 for 253 variables refined against 3368 data (0.74Å), R1 = 0.0265 for those _ data with I > 2.0σ(I). The absolute structure was assigned by refinement of the Flack parameter[6].

**Pre-steady-state equations for kinetic analysis:**

$$primer \xrightarrow{k_{obs1}} n+1 \xrightarrow{k_{obs2}} n+2$$

$$[primer] = A0e^{-k_{obs1}t} \qquad \textbf{Equation 1}$$

$$[n+1] = \frac{k_{obs1}A0}{k_{obs2}-k_{obs1}}(e^{-k_{obs1}t} - e^{-k_{obs2}t}) \qquad \textbf{Equation 2}$$

$$[n+2] = A0 - [primer] - [n+2] \qquad \textbf{Equation 3}$$

$$[n+2] = A0(1 - \frac{k_{obs2}}{k_{obs2}-k_{obs1}}e^{-k_{obs1}t} + \frac{k_{obs1}}{k_{obs2}-k_{obs1}}e^{-k_{obs2}t}) \qquad \textbf{Equation 4}$$

Downloaded from https://academic.oup.com/nar/article/45/10/5629/3796333 by guest on 25 April 2024

**Table S1. Crystal data and structure refinement for 5**

| | | |
|---|---|---|
| Temperature | 133(2) K | |
| Wavelength | 0.71073 Å | |
| Crystal system | Orthorhombic | |
| Space group | $P2_12_12_1$ | |
| Unit cell dimensions | a = 5.2252(3) Å | α= 90 °. |
| | b = 10.9309(7) Å | β= 90 °. |
| | c = 24.2885(16) Å | γ = 90 °. |
| Volume | 1387.27(15) $Å^3$ | |
| Crystal color | yellow | |
| Crystal size | 0.380 x 0.240 x 0.212 $mm^3$ | |
| Theta range for data collection | 1.677 to 28.633 ° | |
| Index ranges | $-7 \leq h \leq 6$, $-14 \leq k \leq 14$, $-31 \leq l \leq 32$ | |
| Reflections collected | 16618 | |
| Independent reflections | 3368 [R(int) = 0.0302] | |
| Completeness to theta = 25.500 ° | 99.9 % | |
| Absorption correction | Semi-empirical from equivalents | |
| Max. and min. transmission | 0.8621 and 0.8216 | |
| Refinement method | Full-matrix least-squares on $F^2$ | |
| Data / restraints / parameters | 3368 / 0 / 253 | |
| Goodness-of-fit on $F^2$ | 1.066 | |
| Final R indices [I>2sigma(I) = 3278 data] | R1 = 0.0265, wR2 = 0.0676 | |
| R indices (all data, 0.74 Å) | R1 = 0.0273, wR2 = 0.0683 | |
| Absolute structure parameter | -0.01(2) | |
| Largest diff. peak and hole | 0.283 and -0.206 e.$Å^{-3}$ | |

Downloaded from https://academic.oup.com/nar/article/45/10/5629/3796333 by guest on 25 April 2024

**Table S2. Oligonucleotides used in this study**

| Name | Sequence (5'-3') | **Mol. Wt. (Calc.)** | **Mol. Wt. (Found)** |
|---|---|---|---|
| PBS2-1G | ACTATTCAACTTACAATCGTATCAACCTTATAAT CCACTTGGCTACTGCATACGAGTGTC | 18286.9 | 18286.8 [a] |
| PBS2-2G | ACTATTCAACTTACAATGGTATCAACCTTATAAT CCACTTGGCTACTGCATACGAGTGTC | 18326.9 | 18326.6 [a] |
| PBS2-3G | ACTATTCAACTTACAATGGGATCAACCTTATAAT CCACTTGGCTACTGCATACGAGTGTC | 18352.0 | 18351.4 [a] |
| PSB7-1G | TTTGCCGTCTCAATGCACTGACGATCC | 8186.3 | 8185.6 [a] |
| PBS7-2G | TTTGCCGGTCTCAATGCACTGACGATCC | 8515.6 | 8515.2 [a] |
| 4NT9G | GGATCGTCAGTGCATTGAGATTAAGACTCGCCA TGTTACGATCTGCCAAGTACAGCCTTGAATCGT CACTGGTGGTATCCCCTTGGGGA/3'-ddC | 27768.0 | 27768.2 [a] |
| PBS8_Extra | FAM/CTTTTAAGAACCGGACGAACGTCC CCTTGGGGATACCACC | 12763.4 | 12763.9 [a] |
| PBS7 | GGATCGTCAGTGCATTGAGA | 6197.1 | 6196.3 [a] |
| Extra | CTTTTAAGAACCGGACGAAC | 6110.0 | 6110.2 [a] |
| PBS7-TNA5-1G | TTTTGCCGCCTATTCTCAATGCACTGACGATCC | 9990.5 | 9989.6 [a] |
| PBS2-IR800 | IR800/spacer/GACACTCGTATGCAGTAGCC | 7875.8 | 7872.8 [b] |
| PBS7-IR700 | IR700/GGATCGTCAGTGCATTGAGA | 6951.0 | 6950.9 [a] |
| PBS7-TNA5-IR800 | IR800/GGATCGTCAGTGCATTGAGAatagg (low case is TNA) | 9145.6 | 9142.7 [b] |

[a] Results provided by Integrated DNA Technologies (IDT). [b] Measured in the linear negative mode..

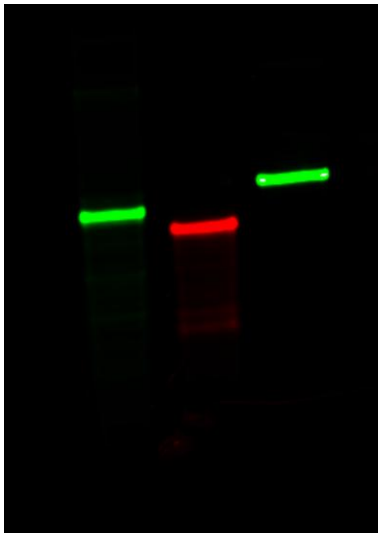

**Figure S1. Purity analysis of IR-dye labelled primers (PBS2-IR800, PBS7-IR700 and PBS7-TNA5-IR800).** Analyzed by 20% denaturing urea PAGE and scanned by an ODYSSEY scanner.



Downloaded from https://academic.oup.com/nar/article/45/10/5629/3796333 by guest on 25 April 2024

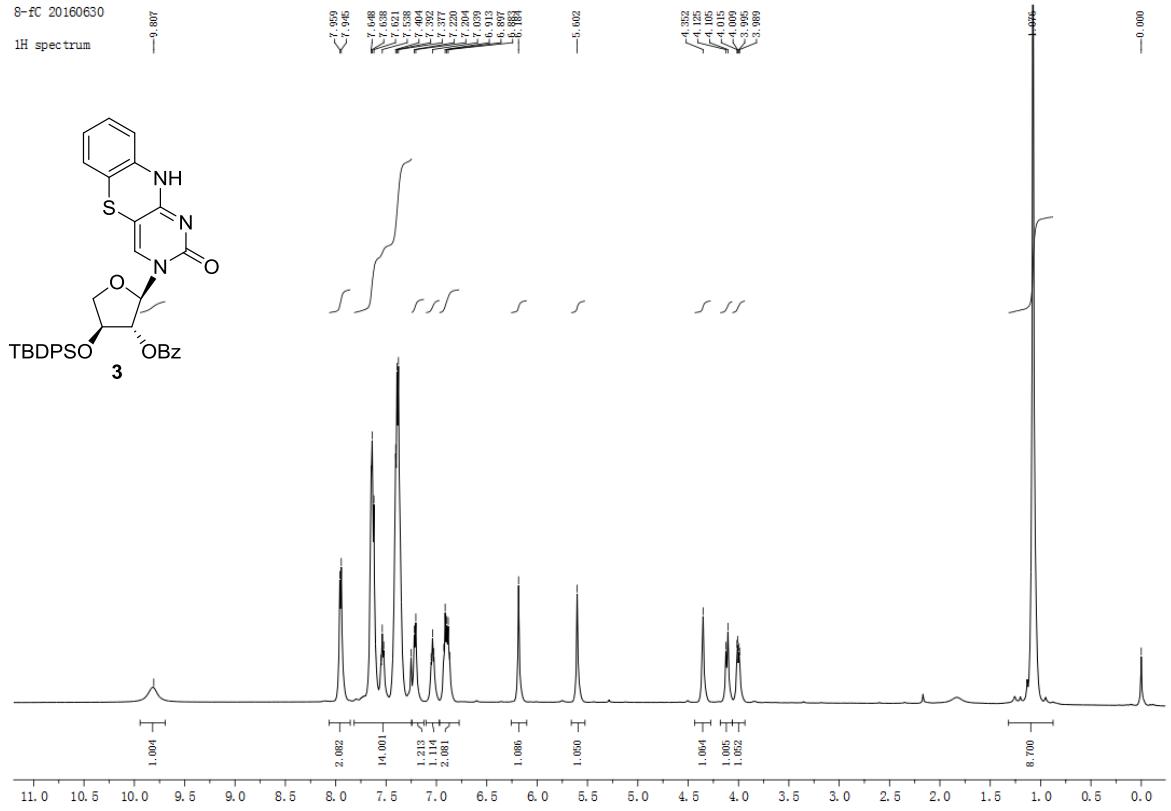


**Figure S2**. $^{1}$H NMR spectrum of compound **3** (500 MHz, $CDCl_3$).

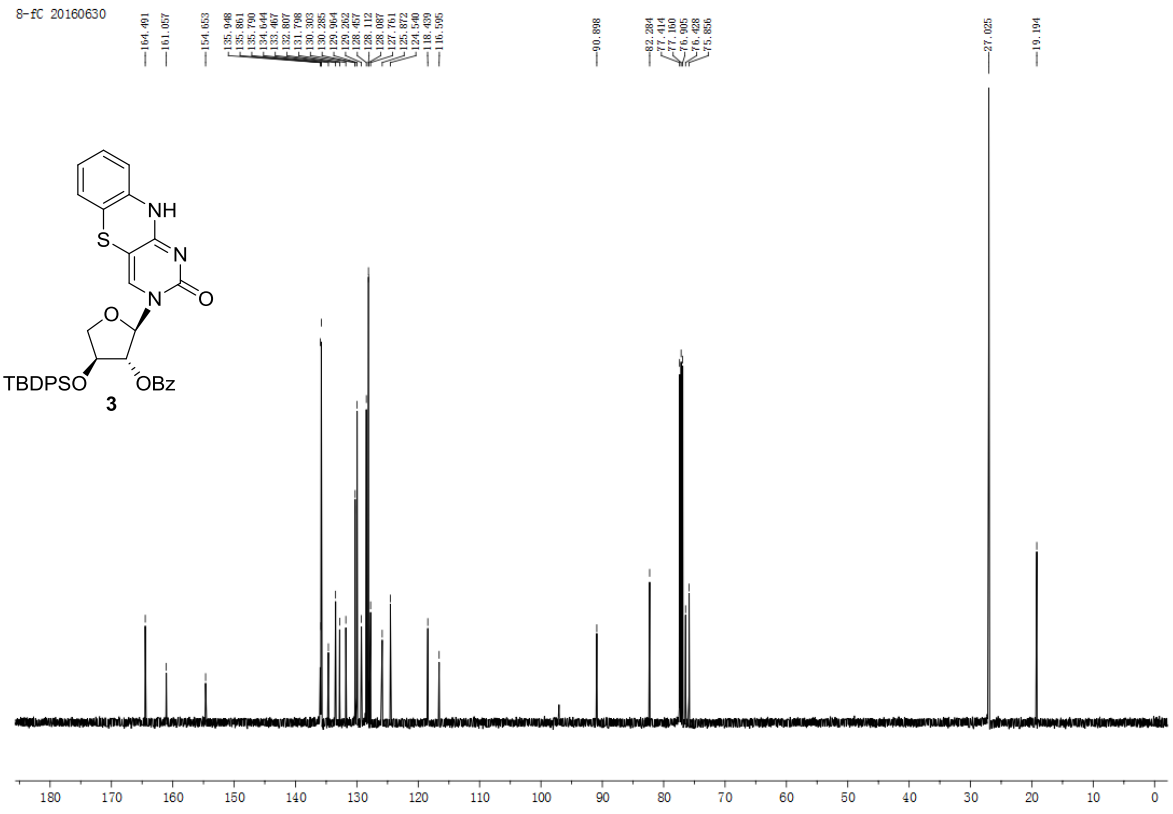


**Figure S3**. $^{13}$C NMR spectrum of compound **3** (125.8 MHz, $CDCl_3$).



Downloaded from https://academic.oup.com/nar/article/45/10/5629/3796333 by guest on 25 April 2024

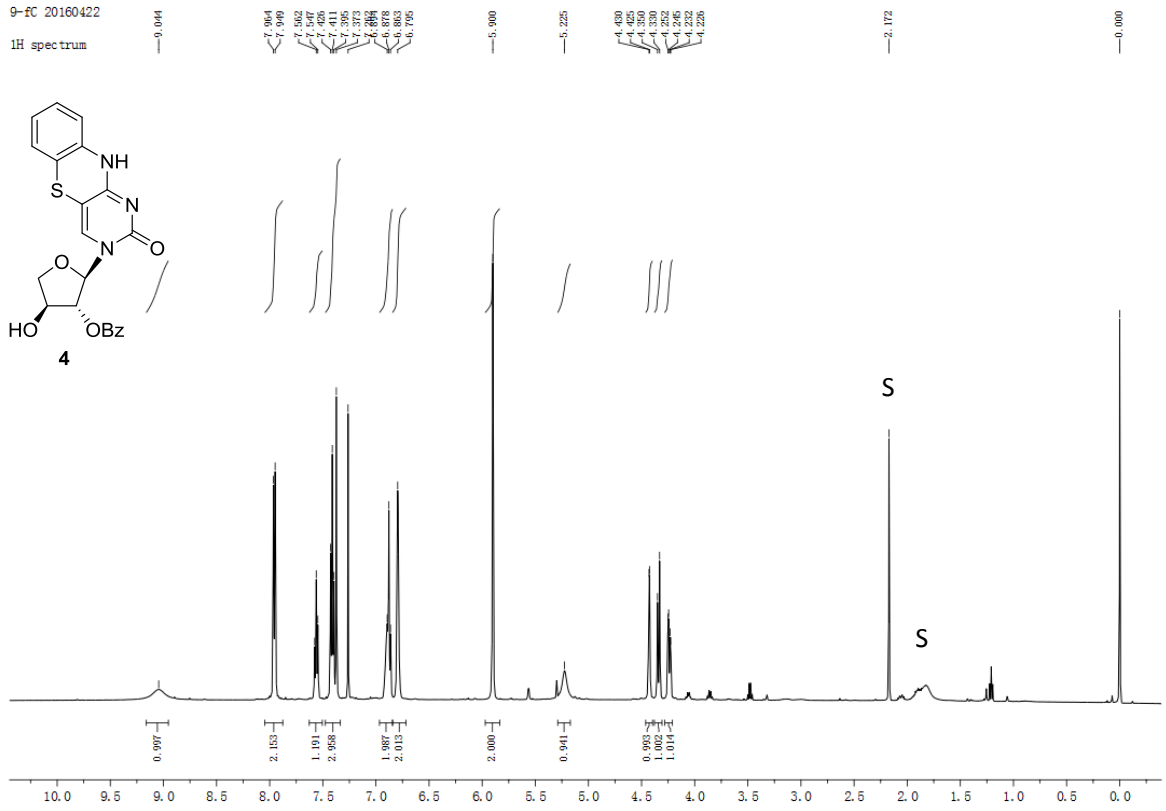


**Figure S4**. $^{1}$H NMR spectrum of compound **4** (500 MHz, $CDCl_3$).

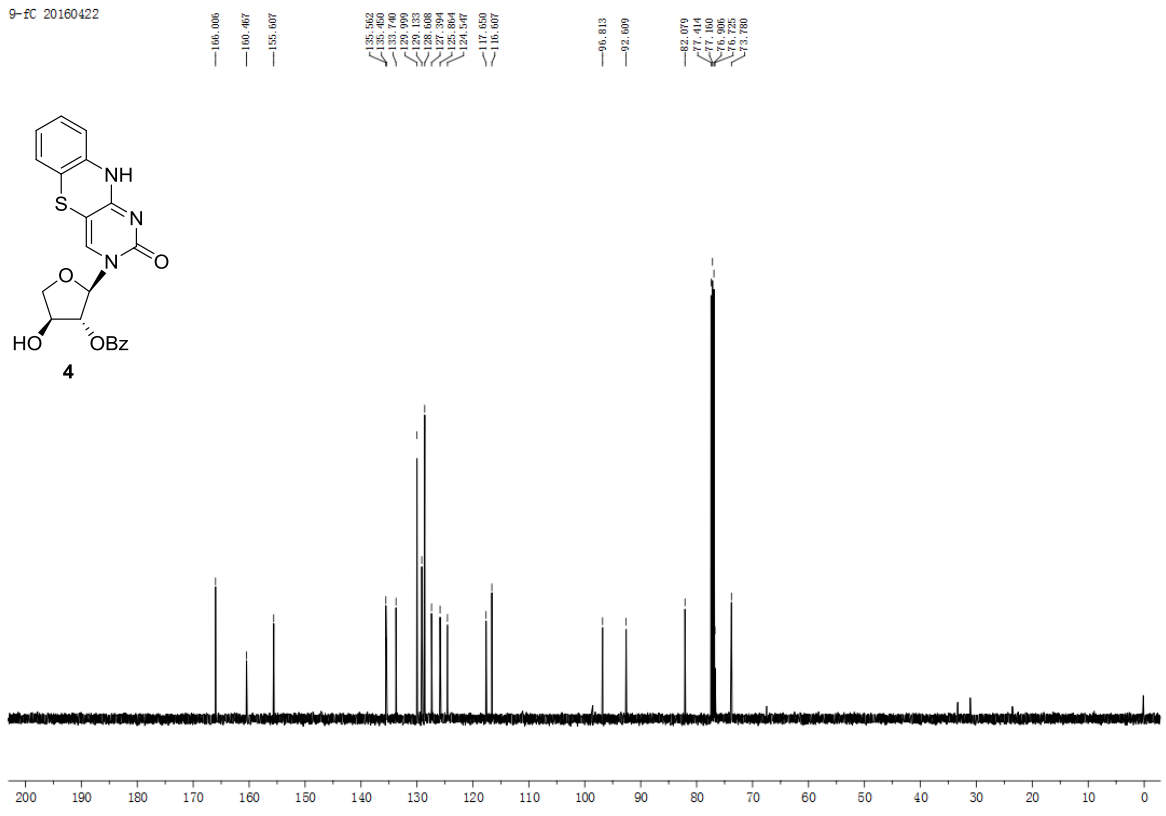


**Figure S5**. $^{13}$C NMR spectrum of compound **4** (125.8 MHz, $CDCl_3$).



Downloaded from https://academic.oup.com/nar/article/45/10/5629/3796333 by guest on 25 April 2024

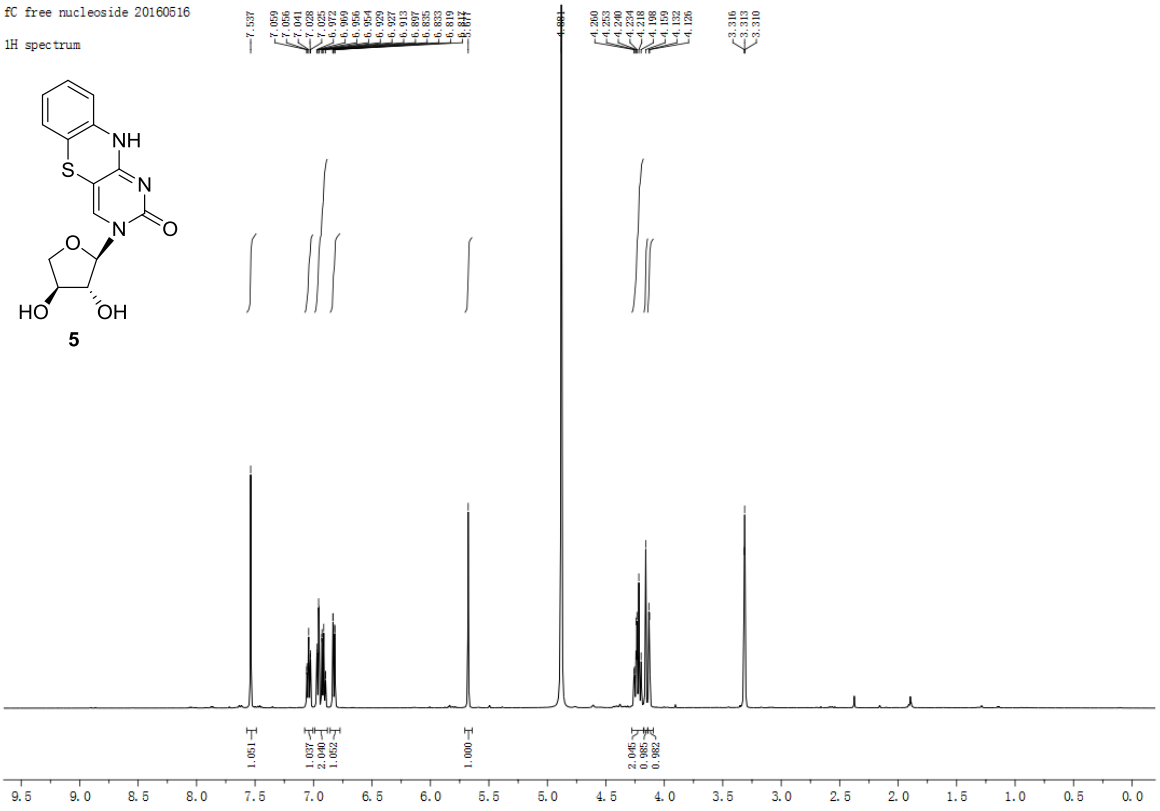


**Figure S6**. $^{1}$H NMR spectrum of compound **5** (500 MHz, $CD_3OD$).

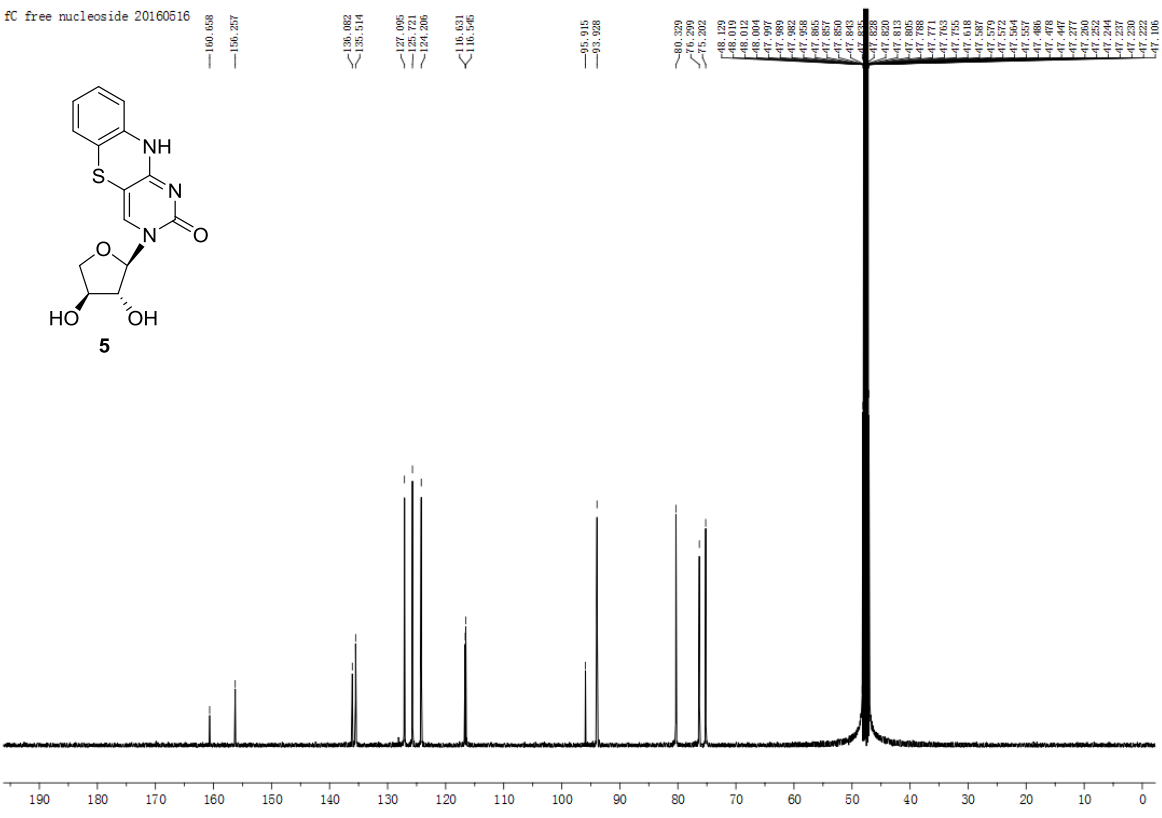


**Figure S7**. $^{13}$C NMR spectrum of compound **5** (125.8 MHz, $CD_3OD$).



Downloaded from https://academic.oup.com/nar/article/45/10/5629/3796333 by guest on 25 April 2024

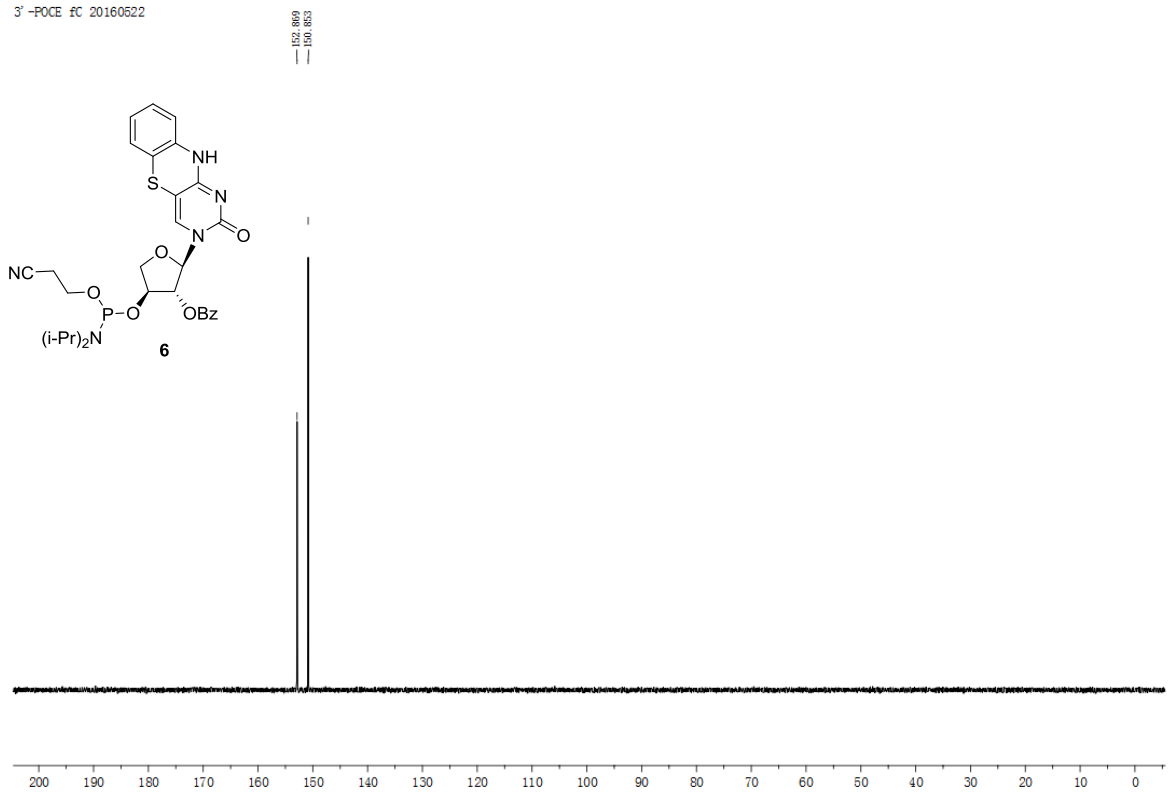


**Figure S8**. $^{31}$P NMR spectrum of compound **6** (162 MHz, $CDCl_3$).

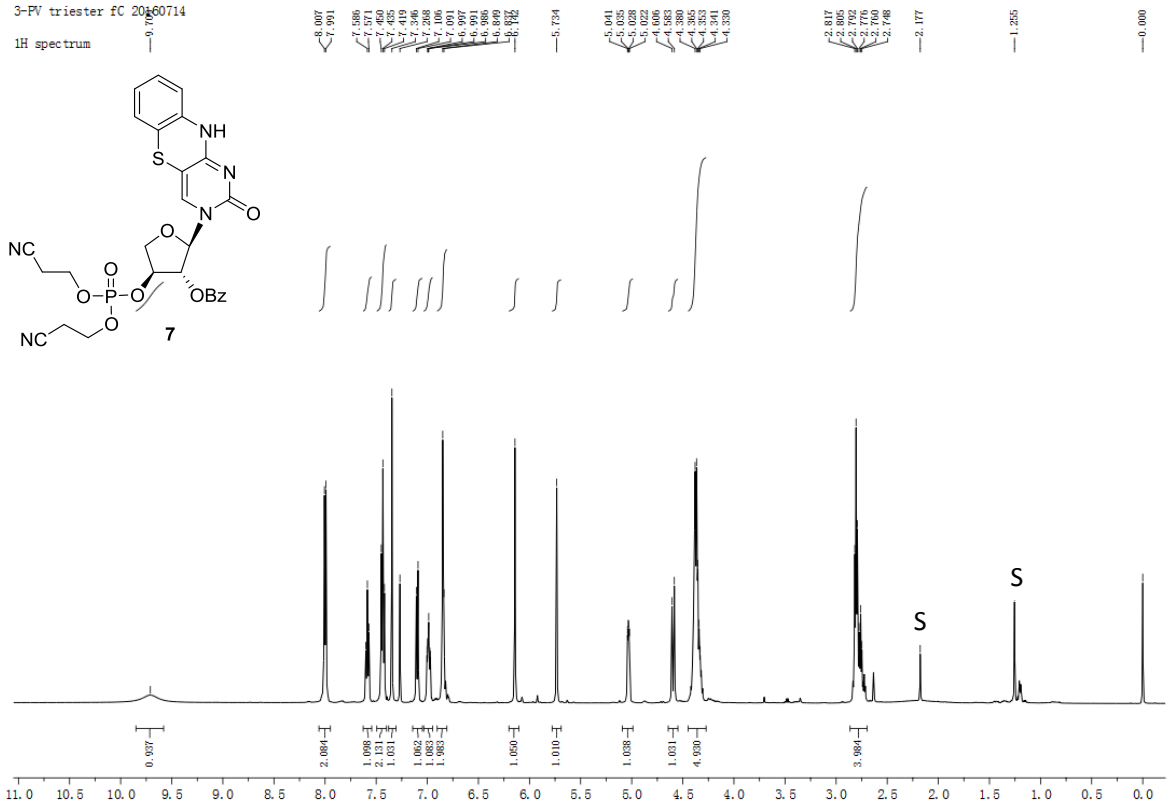


**Figure S9**. $^{1}$H NMR spectrum of compound **7** $^{1}$H NMR (500 MHz, $CDCl_3$).



Downloaded from https://academic.oup.com/nar/article/45/10/5629/3796333 by guest on 25 April 2024

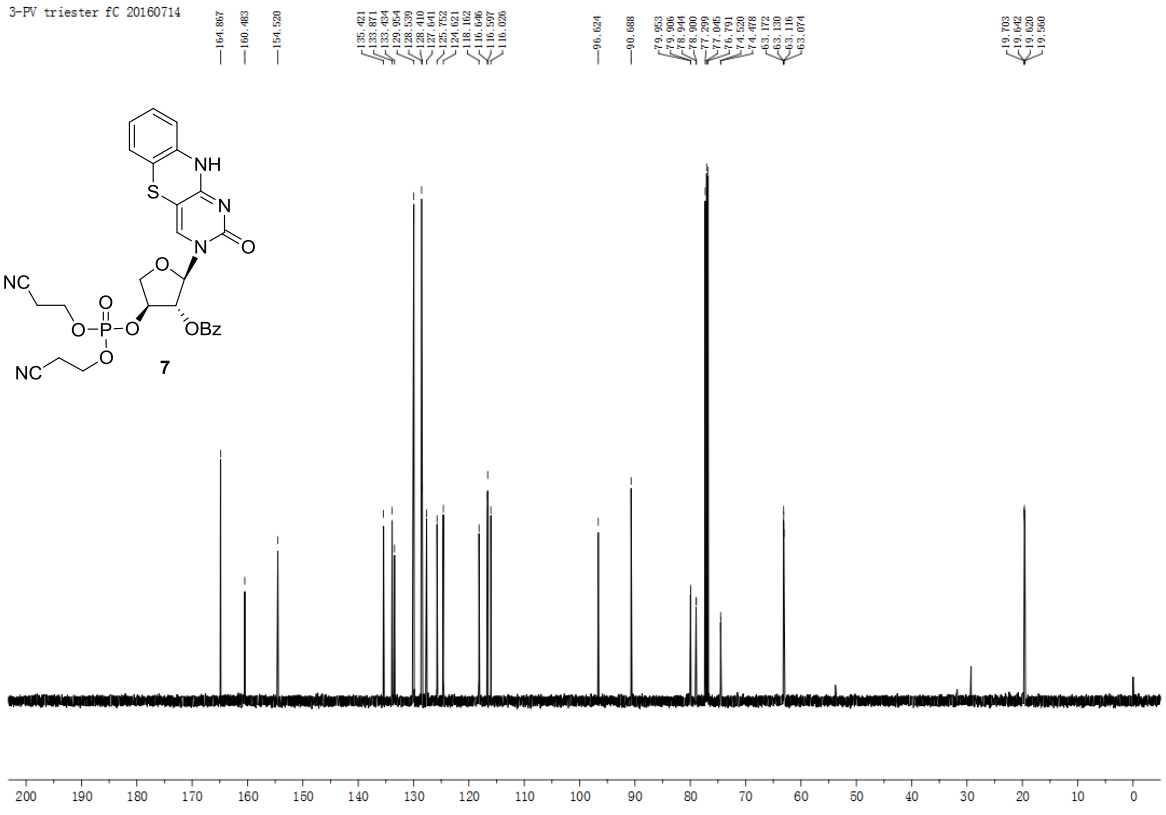


**Figure S10**. $^{13}C$ NMR spectrum of compound **7** (125.8 MHz, $CDCl_3$).

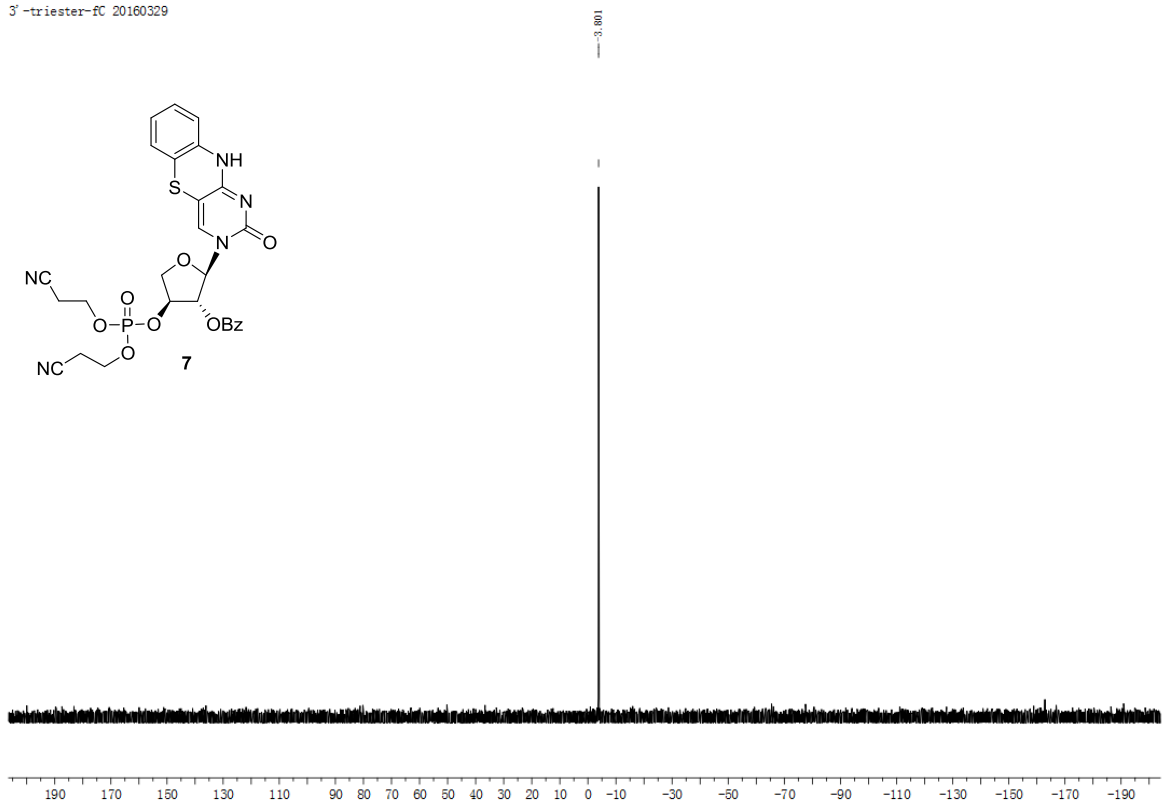


**Figure S11**. $^{31}P$ NMR spectrum of compound **7** (162 MHz, $CDCl_3$).



Downloaded from https://academic.oup.com/nar/article/45/10/5629/3796333 by guest on 25 April 2024

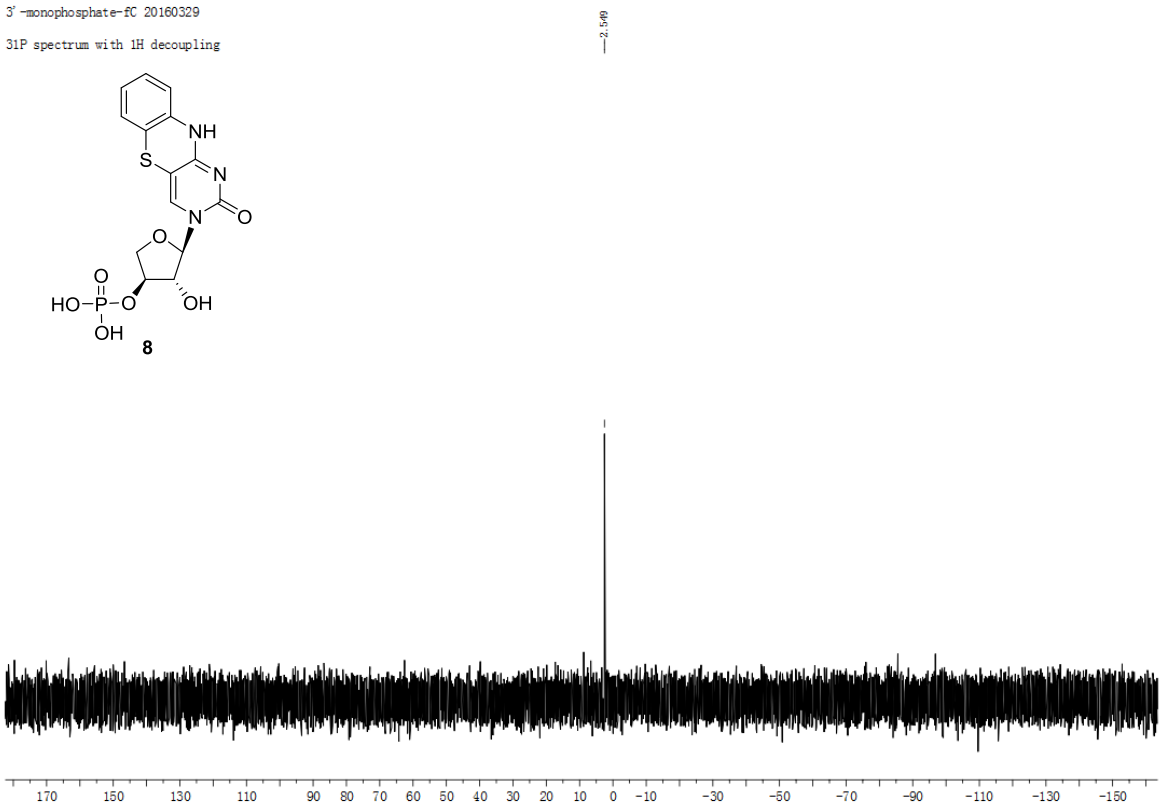


**Figure S12**. $^{31}$P NMR spectrum of compound **8** (162 MHz, $D_2O$).

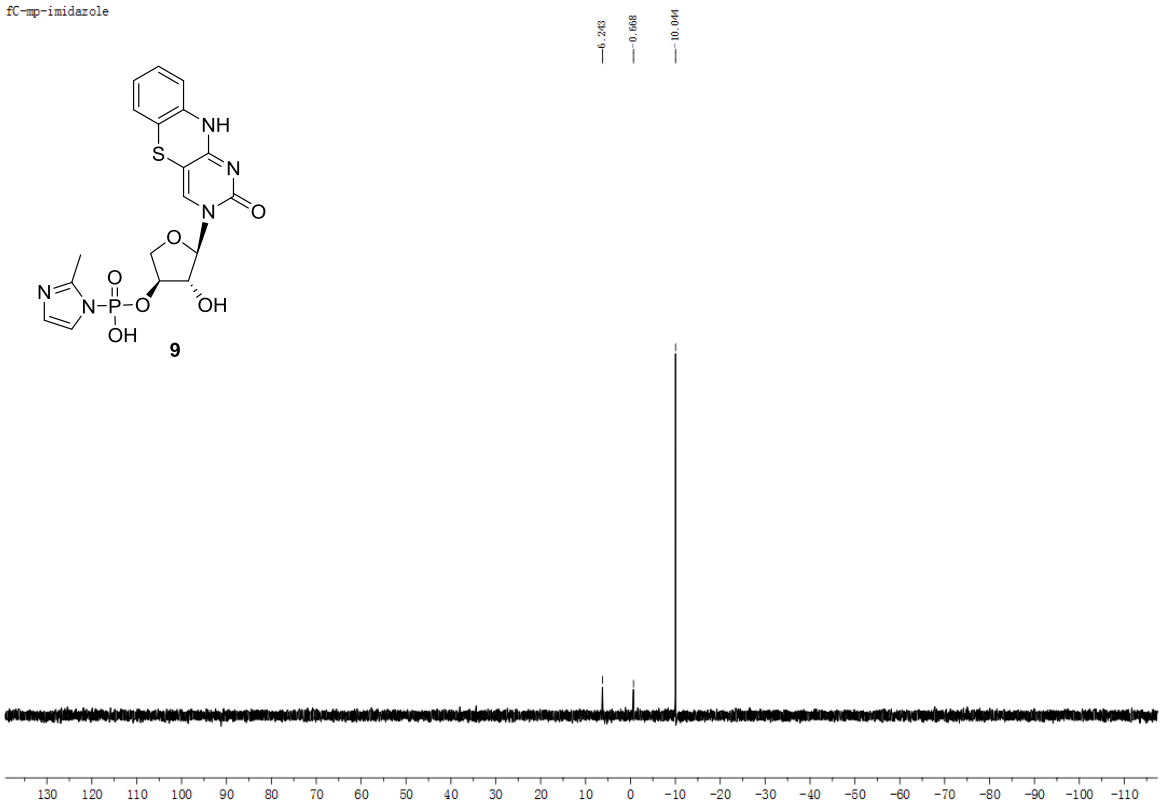


**Figure S13**. $^{31}$P NMR spectrum of compound **9** (162 MHz, DMSO-$d_6$).



Downloaded from https://academic.oup.com/nar/article/45/10/5629/3796333 by guest on 25 April 2024

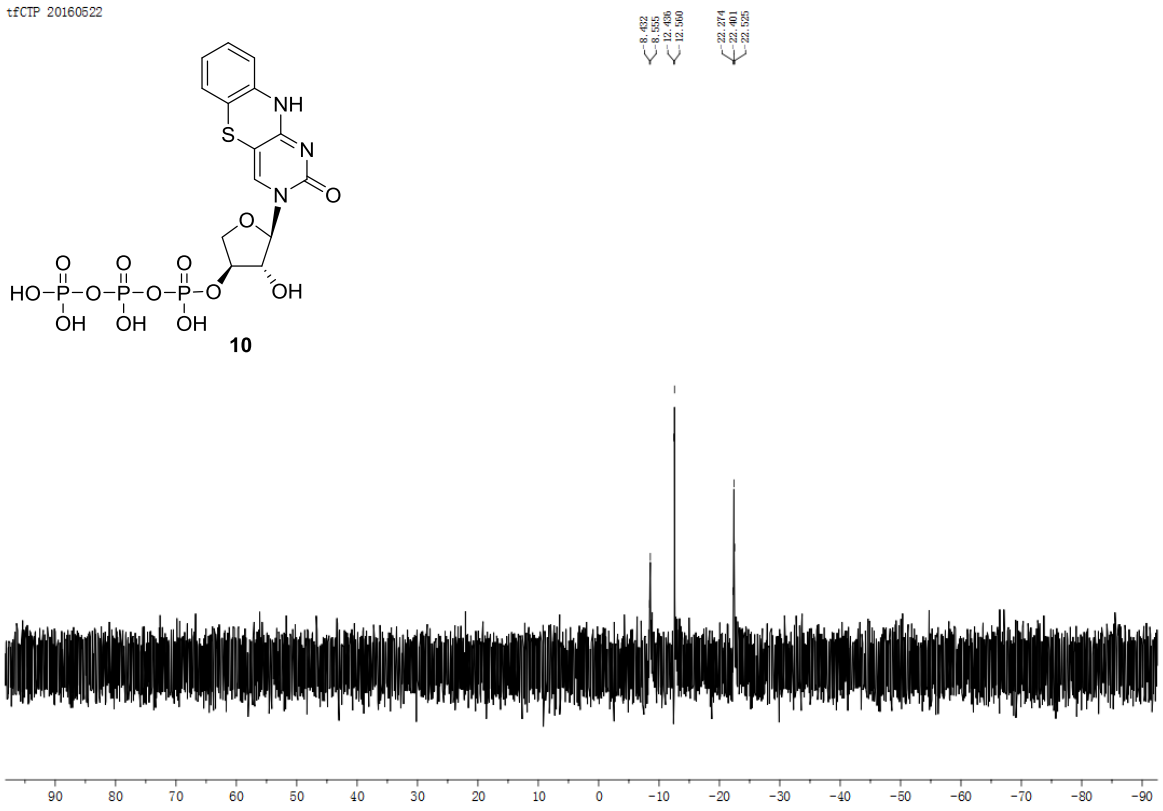


**Figure S14**. $^{31}$P NMR spectrum of compound **10** (162 MHz, $D_2O$).

Downloaded from https://academic.oup.com/nar/article/45/10/5629/3796333 by guest on 25 April 2024