# Electronic Supplementary Information (ESI)

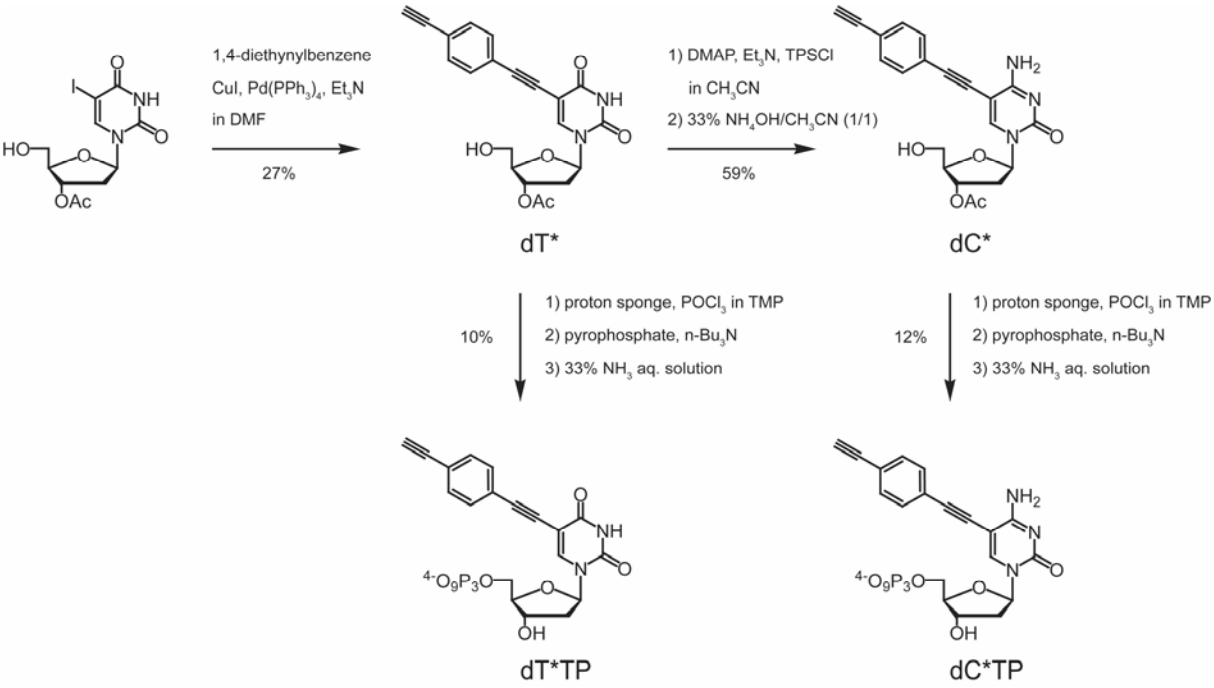
# Interactions of Non-Polar and "Click-able" Nucleotides in the Confines of a DNA Polymerase

Samra Obeid, Holger Bußkamp, Wolfram Welte, Kay Diederichs & Andreas Marx

Konstanz Research School Chemical Biology, University of Konstanz, Universitätsstrasse 10, D 78457 Konstanz, Germany

**CONTENT**

## Nucleotide Synthesis




**Scheme S1.** Synthesis strategy of dT*TP and dC*TP.

### 3'-O-Acetyl-5-(2-(4-ethynylphenyl)ethynyl)-2'-deoxyuridine (dT*)

200 mg (0.51 mmol) 5-iodo-2',3'-dideoxyuridine, 127 mg (1.0 mmol) 1,4-diethynylbenzene and 19 mg (0.1 mmol) CuI (copper(I) iodine) were dissolved in DMF. The solution was degassed before 5.8 mg (5.1 µmol) $Pd(PPh_3)_4$ (tetrakis(triphenylphosphine)palladium (0)) catalyst was added. After a second time of degassing 140 µl (1.0 mmol) previously distilled $Et_3N$ was added. The reaction was stirred at room temperature and monitored by TLC (ethyl acetate / petrol ether = 3 / 1). After 5 h the reaction mixture was extracted against saturated $NaHCO_3$ solution. The aqueous solution was washed several times with diethyl ether. The combined organic layers were dried over $MgSO_4$. Next the solvent was removed and the product was purified by flash chromatography (solvent: ethyl acetate and petrol ether; stepwise gradient 2/1 – 1/1 – 1/2 – 1/3) to yield 5-(2-(4-ethynylphenyl)ethynyl)-2'-deoxyuridine (0.13 mmol, 27% yield).

$\delta_H$ (400 MHz, [$D_4$] MeOD) 2.10 (3H, s, OAc), 2.37 (1H, ddd, $^3J$ (H,H) = 6.3, 8.2 Hz, $^2J$ (H,H) = 14.1 Hz, H-2'), 2.46 (1H, ddd, $^3J$ (H,H) = 2.0, 5.8 Hz, $^2J$ (H,H) = 14.1 Hz, H-2'), 3.82 (1H, dd, $^3J$ (H,H) = 3.1 Hz, $^2J$ (H,H) = 11.8 Hz, H-5'), 3.62 (1H, s, C≡CH), 3.84 (1H, dd, $^3J$ (H,H) = 2.8 Hz, $^2J$ (H,H) = 12.0 Hz, H-5'), 4.14 (1H, q, $^3J$ (H,H) = 2.7 Hz, H-4'), 5.33 (1H, dt, $^3J$ (H,H) = 2.0, 6.1 Hz, H-3'), 6.28 (1H, dd, $^3J$ (H,H) = 5.9, 8.2 Hz, H-1'), 7.54 – 7.40 (4H, m, ArH), 8.45 ppm (1H, s, H-6).

$\delta_C$ (101 MHz, [$D_4$] MeOD) 20.9, 39.1, 62.8, 76.4, 80.6, 83.8, 83.9, 87.0, 87.1, 93.4, 100.6, 101.4, 123.8, 124.7, 132.5, 133.0, 145.1, 151.2, 172.2 ppm.

HRMS: m/z: calcd for [$C_{21}H_{17}N_2O_6^-$]: 393.1081; found: 393.1088.

**3'-O-Acetyl-5-(2-(4-ethynylphenyl)ethynyl)-2'-deoxycytidine (dC*)**

For the conversion of the modified deoxyuridine to the corresponding modified deoxycytidine 64 mg (0.16 mmol) of the modified deoxyuridine and 39.7 mg (0.33 mmol) DMAP (4-(dimethylamino)-pyridine) were dissolved in acetonitrile. Under ice bath cooling 45 µl (0.33 mmol) $Et_3N$ and subsequently 93.4 mg (0.31 mmol) TPSCl (triisopropyl benzenesulfonyl chloride) were added. The reaction mixture was stirred for 1.5 h at 0°C. Next 1 ml of a mixture of 33% $NH_4OH$ / $CH_3CN$ = 1 / 1 was added to the reaction mixture. After 1.5 h stirring at 0°C the reaction mixture was allowed to warm up to room temperature and stirring was continued for another 1.5 h. The reaction mixture was diluted with $CH_2Cl_2$ and extracted against 1M aqueous $KHSO_4$ solution. After three times extraction of the aqueous phase with $CH_2Cl_2$ the organic layers were combined, dried over $MgSO_4$, and concentrated. The desired cytidine derivative (95 µmol, 59% yield) was obtained following purification by flash chromatography (solvent: ethyl acetate and MeOH; stepwise gradient 0-4% MeOH).

$\delta_H$ (600 MHz, [$D_6$] DMSO) 2.07 (3H, s, OAc), 2.22 (1H, ddd, $^3J$ (H,H) = 6.3, 8.3 Hz, $^2J$ (H,H) = 14.3 Hz, H-2'), 2.33 (1H, ddd, $^3J$ (H,H) = 1.9, 5.7 Hz, $^2J$ (H,H) = 14.0 Hz, H-2'), 4.08 – 4.01 (1H, m, H-4'), 3.73 – 3.62 (2H, m, H-5'), 4.34 (1H, s, C≡CH), 5.21 (1H, dt, $^3J$ (H,H) = 2.0, 6.2 Hz, H-3'), 5.35 – 5.24 (1H, br.s, OH), 6.17 (1H, dd, $^3J$ (H,H) = 5.8, 8.3 Hz, H-1'), 7.16 (1H, br s, $NH_2$), 7.55 – 7.48 (2H, m, ArH), 7.68 – 7.59 (2H, m, ArH), 7.85 (1H, br s, $NH_2$), 8.32 ppm (1H, s, H-6).

$\delta_C$ (151 MHz, [$D_6$] DMSO) 20.9, 37.9, 61.2, 74.6, 82.7, 83.1, 83.7, 85.0, 85.5, 89.7, 93.4, 121.5, 123.0, 131.4, 131.8, 135.8, 145.1, 153.3, 163.7, 170.1, 177.7 ppm.

HRMS: m/z: calcd for [$C_{21}H_{18}N_3O_5^-$]: 392.1241; found: 392.1247.

**5-(2-(4-Ethynylphenyl)ethynyl)-2'-deoxyuridine-5'-triphosphate (dT*TP)**

14 mg (35.5 µmol) modified deoxyuridine and 11.4 mg (53.2 µmol) proton sponge were dissolved in previously distilled trimethylphosphate. The addition of 3.9 µl (42.6 µmol) $POCl_3$ was carried out under ice bad cooling. The reaction was monitored by reverse phase TLC in isopropanol / water/ ammonia = 3 / 1 / 1. After 1 h stirring 355 µl (1.78 mmol) pyrophosphate and 94 µl (355 µmol) tributylamine were added fast and simultaneously. The solution was stirred for 15 min and afterwards quenched by adding 5 ml 0.1 M TEAB buffer. The reaction mixture was extracted three times against ethyl acetate. The sample was concentrated and resolved in 10 ml water. To remove the 3' protecting group 20 ml 33% ammonia solution was added. The reaction mixture was stirred for 1 h and was afterwards

concentrated. The aqueous solution was further purified by ion exchange chromatography. The remaining sample was desalted by RP-MPLC using 0.05 M TEAA buffer and acetonitrile to afford the modified deoxyuridine-5'-triphosphate (3.6 µmol, 10% yield).

$\delta_H$ (400 MHz, $D_2O$) 1.24 (36H, t, $^3J$ (H,H) = 7.3 Hz, $NEt_3$), 2.43 – 2.37 (2H, m, H-2'), 2.53 – 2.42 (1H, m, H-2'), 3.16 (24H, q, $^3J$ (H,H) = 7.3 Hz, $NEt_3$), 3.58 (1H, s, C≡CH), 4.27 – 4.16 (3H, m, H-4'; H-5'), 4.67 – 4.62 (1H, m, H-3'), 6.28 (1H, t, $^3J$ (H,H) = 6.7 Hz, H-1'), 7.50 (2H, d, $^3J$ (H,H) = 8.0 Hz, ArH), 7.56 (2H, d, $^3J$ (H,H) = 8.0 Hz, ArH), 8.20 ppm (1H, s, H-6).

$\delta_P$ (162 MHz, $D_2O$) -7.78 – -8.25 (1P, m, $P_\gamma$), -11.23 (1P, d, $^2J$ = 20.0 Hz, $P_\alpha$), -22.04 – -22.50 ppm (1P, m, $P_\beta$).

HRMS: m/z: calcd for $[C_{19}H_{18}N_2O_{14}P_3^-]$: 590.9965; found: 590.9972.

**5-(2-(4-Ethynylphenyl)ethynyl)-2'-deoxycytidine-5'-triphosphate (dC*TP)**

15 mg (38 µmol) modified deoxycytidine and 12.3 mg (57.2 µmol) proton sponge were dissolved in previously distilled trimethylphosphate. The addition of 4.2 µl (45.8 µmol) $POCl_3$ was carried out under ice bad cooling. The reaction was monitored by reverse phase TLC in isopropanol / water/ ammonia = 3 / 1 / 1. After 1 h stirring 381 µl (191 µmol) pyrophosphate and 101 µl (380 µmol) tributylamine were added fast and simultaneously. The solution was stirred for 15 min and afterwards quenched by adding 5 ml 0.1 M TEAB buffer. The reaction mixture was extracted three times against ethyl acetate. The sample was concentrated and resolved in 10 ml water. To remove the 3' protecting group 20 ml 33% ammonia solution was added. The reaction mixture was stirred for 1 h and was afterwards concentrated. The aqueous solution was further purified by ion exchange chromatography. The remaining sample was desalted by RP-MPLC using 0.05 M TEAA buffer and acetonitrile to afford the modified deoxycytidine-5'-triphosphate (4.5 µmol, 12% yield).

$\delta_H$ (400 MHz, $D_2O$) 1.31 (9H, t, $^3J$ (H,H) = 7.2 Hz, $NEt_3$), 2.42 – 2.30 (2H, m, H-2'), 2.57 – 2.46 (1H, m, H-2'), 3.23 (6H, q, $^3J$ (H,H) = 7.1 Hz, $NEt_3$), 3.65 (1H, s, C≡CH), 4.27 (3H, br.s, H-4'; H-5'), 4.69 – 4.60 (1H, m, H-3'), 6.31 (1H, t, $^3J$ (H,H) = 6.4 Hz, H-1'), 7.57 (2H, d, $^3J$ (H,H) = 7.0 Hz, ArH), 7.64 (2H, d, $^3J$ (H,H) = 7.0 Hz, ArH), 8.22 ppm (1H, s, H-6).

$\delta_P$ (162 MHz, $D_2O$) -8.40 – -10.63 (1P, m, $P_\gamma$), -10.62 – 12.00 (1P, m, $P_\alpha$), -20.43 – -23.90 ppm (1P, m, $P_\beta$).

HRMS: m/z: calcd for $[C_{19}H_{19}N_3O_{13}P_3^-]$: 590.0125; found: 590.0143.

**Competition experiment- single nucleotide incorporation assay**

Competition experiment was conducted as described previously [1]. In brief, incorporation opposite dA or dG: 10µl of the *KlenTaq* reactions contained 50 nM primer (5'-GTG GTG CGA AAT TTC TGA CAG ACA-3'), 75 nM template (5'-GTG CGT CTG TCA/G TGT CTG TCA GAA ATT TCG CAC CAC-3'), 200 µM dNTP mixture of either dTTP/dT*TP or dCTP/dC*TP in buffer (20 mM Tris HCl pH 7.5, 50 mM NaCl, and 2 mM $MgCl_2$) and 200 nM of *KlenTaq* polymerase. The ratio of modified nucleotide to natural nucleotide varied from 1/1 to 100/1 (1/1, 2/1, 4/1, 10/1, 20/1, 50/1 and 100/1). Reaction mixtures were incubated at 37°C. Incubation times are provided in the respective figure legends. Primer was labelled using [$\gamma$-$^{32}$P]-ATP according to standard techniques. Reactions were stopped by addition of 22.5 µl stop solution (80% [v/v] formamide, 20 mM EDTA, 0.25% [w/v] bromophenol blue, 0.25% [w/v] xylene cyanol) and analysed by 20% denaturing PAGE. Visualization was performed by phosphoimaging.

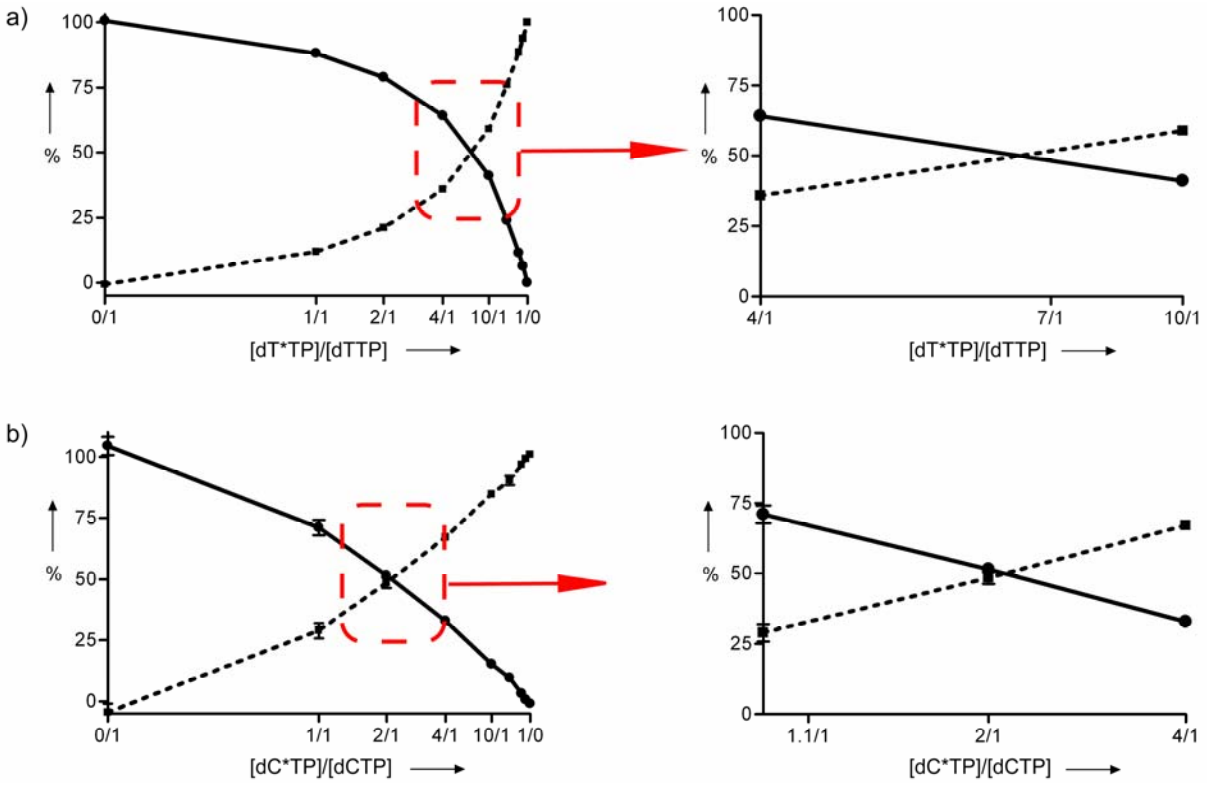
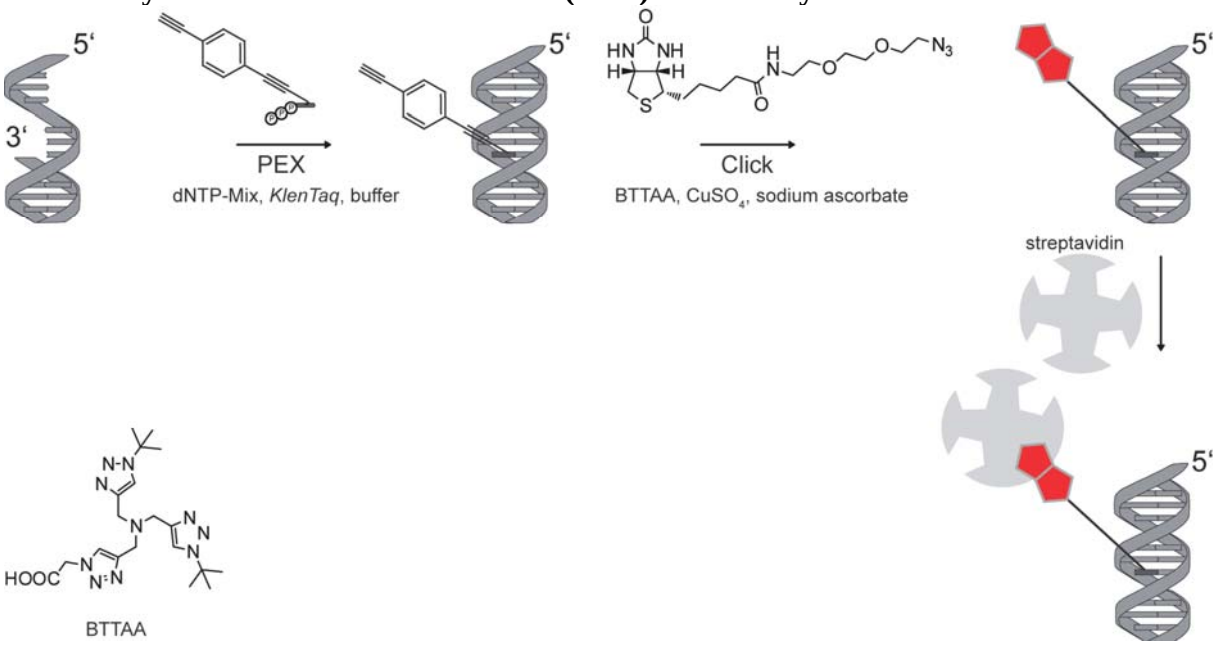



**Figure S1.** Acceptance of dT*TP and dC*TP by *KlenTaq* DNA polymerase. a) Competition experiments of dT*TP versus dTTP and dTTP. The ratio of dT*TP/dTTP was varied from 1/1 to 100/1 (1/1, 2/1, 4/1, 10/1, 20/1, 50/1, 100/1). The product bands of the PAGE analysis were quantified to evaluate the incorporation efficiencies of dT*TP (■, dashed line) and dTTP (•, solid line) catalyzed by *KlenTaq*. The conversion in % was plotted versus the concentration. The cross spot is highlighted by red squared of dashed line. A zoom into the cross spot

indicates the approximate ratio where both nucleotides are equally incorporated. b) Same as in a) using dC*TP and dCTP instead.

**Click assay – Primer extension reactions (PEX) followed by Click reactions**




**Scheme S2.** Concept – PEX reactions followed by Click reactions.

PEX reaction prior to click chemistry:

Generation of modified DNA by incorporation of dT*MP or dC*MP : 20 µl of the reaction contained 900 nM Primer (5'-GTG GTG CGA AAT TTC TGA CAG ACA-3'), 1.2 µM template (for incorporation of dT*MP: 5'-GTG CGT CTG TCA TGT CTG TCA GAA ATT TCG CAC CAC-3'; for incorporation of dC*MP: 5'-ATA CAT CTA TCG TGT CTG TCA GAA ATT TCG CAC CAC-3') 100 µM dNTPs in 1x buffer (20 mM Tris HCl pH 7.5, 50 mM NaCl, and 2 mM $MgCl_2$) and 200 nM of *KlenTaq* DNA polymerase. The reaction mixtures were incubated for 30 min at room temperature. The primer extension reactions were stopped by centrifuging the reaction mixture over a microspin G25-column.

Click reaction of modified DNA with azide-modified biotin:

BTTAA was synthesized by literature known procedures [2]. The azido-modified biotin was also synthesized by literature known procedures [3].

Click reaction was carried out under a nitrogen atmosphere. A mixture of 50 mM BTTAA, 80 mM sodium ascorbate and 10 mM $CuSO_4$ in 40 µl 250 mM sodium-phosphate buffer pH = 7 was prepared. 20 µl of this mixture was transferred to a mixture of 20 µl of the G25-purified primer extension reaction mixtures and 10 µl of a 10 mM biotin-N3 solution resulting in the following final concentrations: 2 mM biotin-N3, 20 mM BTTAA, 32 mM sodium ascorbate, 4 mM $CuSO_4$ in 50 µl 100 mM sodium-phosphate buffer pH = 7. The reaction mixture was incubated at room temperature for 4 h. The click reaction was stopped by centrifugation of the reaction mixture over a microspin G25-column.

1µl of this solution was incubated with 19 µg streptavidin (New England Biolabs) in a final volume of 20 µl for 30 minutes at room temperature. As a control 1 µl of the click-solution was diluted with water to 20 µl.

All samples were diluted with 20 µl loading dye (80% [v/v] formamide, 20 mM EDTA, 0.25% [w/v] bromophenol blue, 0.25% [w/v] xylene cyanol) and analysed by 20% denaturing PAGE. Visualization was performed by phosphoimaging.

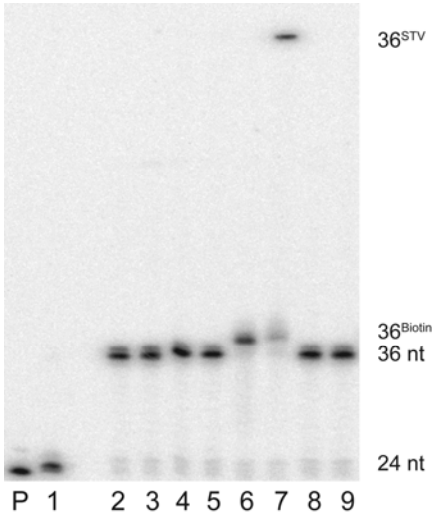



**Figure S2.** PAGE analysis of PEX using *KlenTaq* followed by Click reaction. Lane P: primer only; lane 1: in presence of dATP, dGTP and dCTP; lane 2: all four natural dNTPs; lane 3: same as lane 2 followed by incubation with streptavidin (STV); lane 4: same as in lane 1, but in the presence of dT*TP; lane 5: same as in lane 4 followed by incubation with STV; lane 6: same as in lane 4 followed by click reaction with azido-modified biotin; lane 7: same as line 6 followed by incubation with STV; lane 8: same as lane 4 followed by click reaction without BTTAA and $CuSO_4$; lane 9: same as in lane 8 followed by incubation with STV.

**Protein expression and purification**

Protein expression and purification was conducted as described[4]. In brief, an *E. coli* codon-optimized *KlenTaq* gene (amino acids 293-832 of *Taq* gene; purchased from Geneart, Germany) was cloned into a pET-21b vector without any purification tags and expressed in *E. coli* strain BL21 (DE3). It is noteworthy that codon optimization resulted in changes within the gene sequence without affecting the amino acid sequence. After heat denaturation and ultra centrifugation a PEI-precipitation was performed. The resulting material was purified by anion exchange (Q Sepharose) chromatography followed by size-exclusion chromatography (Superdex 75) in 20 mM Tris HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM β-mercaptoethanol.

**Crystallization and Structure Determination**

The closed ternary complexes of *KlenTaq* were obtained by incubating *KlenTaq* in presence of DNA primer (5′-d(GAC CAC GGC GC)-3′), a suitable template (5′-d(AAA RSG CGC CGT GGT C)-3′), and suitable nucleotides. S is replacement character for G or C coding either for the incorporation of ddCMP or ddGMP, respectively. R is replacement character for A or G coding for the insertion of either a dT*TP or dC*TP. Thereby, one ddC/GMP is incorporated, while the corresponding 5-modified dT/C*TP is captured in the position waiting for insertion.

*KlenTaq*(dT*): The crystallization was set up using purified *KlenTaq* (20 mg⁄ml; buffer: 20 mM Tris HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM β-mercaptoethanol), DNA template/primer duplex, ddCTP and dT*TP in a molar ratio of 1:3:10:20 and in presence of 20 mM $MgCl_2$. The crystallization solution was mixed in 1:1 ratio with the reservoir solution containing 0.05 M sodium cacodylate (pH 6.5), 0.2 M $NH_4OAc$, 0.01 M $Mg(OAc)_2$, and 28% PEG 8000.

*KlenTaq*(dC*): The crystallization was carried out in the similar fashion as in *KlenTaq*(dT*) except using *KlenTaq*:DNA template/primer duplex:ddGTP:dC*TP in a molar ratio of 1•3•10•20 in presence of 20 mM $MgCl_2$. The reservoir solution containied 0.05 M sodium cacodylate (pH 6.5), 0.2 M $NH_4OAc$, 0.01 M $Mg(OAc)_2$, and 28% PEG 8000.

Crystals were produced by the hanging drop vapor diffusion method, equilibrating against 1 ml of the reservoir solution for 5 d at 18 °C. They were frozen in liquid nitrogen and kept at 100 K during data collection. Data of *KlenTaq*(dT*) was measured at the beamline PXIII (X06DA), whereas data of *KlenTaq*(dC*) was measured at the beamline PXI (X06SA) at the Swiss Light Source of the Paul Scherrer Institute (PSI) in Villigen, Switzerland, at a wavelength 1.000 Å and using a Mar225 CCD detector or PILATUS 6M detector, respectively. Data reduction was performed with the XDS package [5]. The structures were solved by difference Fourier techniques using *KlenTaq* wild-type (PDB 1QTM) as model. Refinement was performed with PHENIX [6] and model rebuilding was done with COOT [7]. Figures were made with PyMOL [8].

## Table S1: Data collection and refinement statistics

| | ***KlenTaq*(dT*)** | ***KlenTaq*(dC*)** |
|---|---|---|
| **Data collection** | | |
| Space group | P $3_1$ 2 1 | P $3_1$ 2 1 |
| Cell dimensions | | |
| *a* = *b*, *c* (Å) | 108.9, 91.0 | 108.2, 90.4 |
| Resolution (Å) | 47.14 – 2.19 (2.32 – 2.19) | 46.86 – 1.79 (1.90 – 1.79) |
| $R_{meas}$ | 11.5 (99.1) | 6.5 (101.7) |
| $I / \sigma_I$ | 16.1 (2.1) | 18.4 (2.0) |
| Completeness (%) | 98.5 (90.6) | 99.5 (97.3) |
| Redundancy | 10.4 (7.2) | 8.3 (7.8) |
| **Refinement** | | |
| Resolution (Å) | 47.14 – 2.20 (2.27 – 2.20) | 46.86 – 1.80 (1.83 – 1.80) |
| No. Reflections | 31776 | 56899 |
| $R_{work}$ / $R_{free}$ | 16.0 (24.1) / 21.0 (30.8) | 15.5 (23.3) / 18.3 (28.9) |
| No. of residues | | |
| Protein | 539 | 539 |
| Primer/template | 12/16 | 12/16 |
| Ligand/ion | dT*TP/2 $Mg^{2+}$ | dC*TP/2 $Mg^{2+}$ |
| Water | 242 | 382 |
| *B*-factors (Å$^2$) | | |
| Protein | 35.4 | 31.0 |
| Primer/template | 32.4/37.1 | 29.9/31.4 |
| Ligand/ion | 30.9/33.0 | 20.4/20.4 |
| Water | 36.5 | 39.4 |
| R.m.s. deviations | | |
| Bond lengths (Å) | 0.008 | 0.006 |
| Bond angles (°) | 1.242 | 1.188 |
| Ramachandran statistics [9] | | |
| Most favored | 93.6 | 93.6 |
| Additionally allowed | 6.2 | 6.2 |
| Generously allowed | 0 | 0 |
| Disallowed | 0.2 | 0.2 |
| PDB | 4ELT | 4ELU |

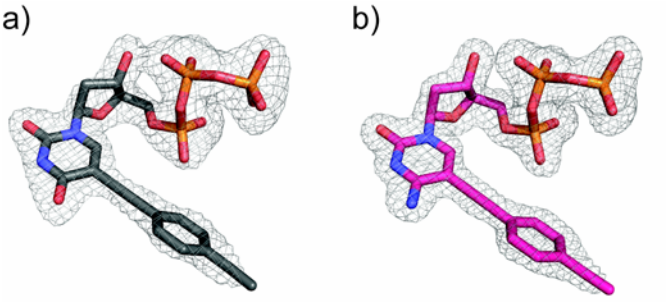



**Figure S3.** Simulated annealing omit maps of the incoming modified pyrimidine analogs. a) The final refined simulated annealing omit map mFo-DFc at 3σ is shown for dT*TP. b) The final refined simulated annealing omit map mFo-DFc at 3σ is shown for dC*TP.

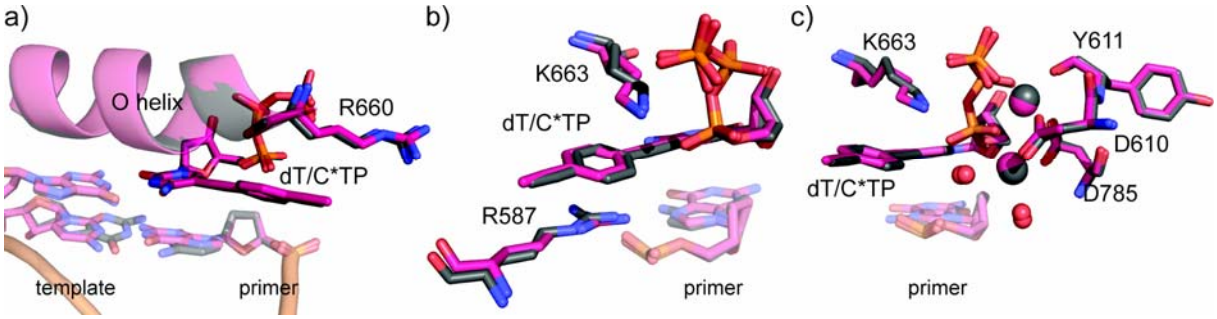



**Figure S4.** Comparison of *KlenTaq*(dC*) structure (pink) and *KlenTaq*(dT*) structure (gray). a) Zoom into the active site of *KlenTaq*(dC*). R660 show the same orientation in both structures. b) Interaction pattern of the C5 modified dC*TP and dT*TP is very much alike. c) The superimposition shows that the complexation of $Mg^{2+}$ ions by the incoming alkyne modified pyrimidine analogs and the catalytically responsible amino acids are nearly identical.

## References

1 S. Obeid, A. Baccaro, W. Welte, K. Diederichs, A. Marx, *Proc. Natl. Acad. Sci. U. S. A.* 2010, **107**, 21327.
2 (a) C. Besanceney-Webler, et al., *Angew Chem Int Ed Engl* 2011; (b) D. Soriano Del Amo, W. Wang, H. Jiang, C. Besanceney, A. C. Yan, M. Levy, Y. Liu, F. L. Marlow, P. Wu, *J. Am. Chem. Soc.* 2010, **132**, 16893.
3 M. Verdoes, et al., *ChemBioChem* 2008, **9**, 1735.
4 Y. Li, S. Korolev, G. Waksman, *EMBO J.* 1998, **17**, 7514.
5 (a) W. Kabsch, *Acta Crystallogr. D* 2010, **66**, 133; (b) W. Kabsch, *Acta Crystallogr. D* 2010, **66**, 125.
6 P. D. Adams, et al., *Acta Crystallogr. D* 2002, **58**, 1948.
7 P. Emsley, K. Cowtan, *Acta Crystallogr. D* 2004, **60**, 2126.
8 W. DeLano, *The PyMOL Molecular Graphics System*, DeLano Scientific, Palo Alto, CA, **2002**.
9 R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* 1993, **26**, 283.