KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry

Karin Betz[1,5], Denis A. Malyshev[2,5], Thomas Lavergne[2], Wolfram Welte[1], Kay Diederichs[1], Tammy J. Dwyer[3], Phillip Ordoukhanian[4], Floyd E. Romesberg[2*], Andreas Marx[1*]

*[1]Departments of Chemistry and Biology, and Konstanz Research School Chemical Biology, Universität Konstanz, Konstanz, Germany. [2]Department of Chemistry and [4]The Center for Protein and Nucleic Acid Research, The Scripps Research Institute, La Jolla, California 92037, U.S.A. [3]Department of Chemistry and Biochemistry, University of San Diego, San Diego, CA 92110, U.S.A. [5]These authors contributed equally to this work. *e-mail: floyd@scripps.edu or andreas.marx@uni-konstanz.de*

Supplementary Information

Table of contents

Nature Chemical Biology: doi:10.1038/NChemBio.966

Supplementary Methods

Protein, Nucleotides, and Oligonucleotides. Protein expression and purification were conducted as described[1-7]. In brief, an *E. coli* codon-optimized KlenTaq gene (amino acids 293–832 of *Taq* gene; purchased from GeneArt, Germany) in the vector pET-21b was expressed in *E. coli* strain BL21 (DE3) without any purification tags. After heat denaturation of the crude cell lysate (20 min, 80°C) a PEI-precipitation was performed. The resulting supernatant was purified by anion exchange chromatography (Q Sepharose) followed by size-exclusion chromatography (Superdex 75).

Unmodified, HPLC purified oligonucleotides were purchased from Thermo Scientific with the following sequences: primer-11mer: 5'-d(GAC CAC GGC GC)-3'; template-G: 5'-d(AAA GCG CGC CGT GGT C)-3' and template-T: 5'-d(AAA TGG CGC CGT GGT C)-3'. d**NaM** phosphoramidite and d**5SICS** nucleoside were obtained from Berry & Associates Inc. (Dexter, MI) and the latter was phosphorylated using Ludwig & Eckstein conditions[8] (see Lavergne, *et al*[9] for detailed protocol of d**5SICS**TP synthesis and purification). The d**NaM** oligonucleotide (template-NaM: 5'-d(AAA **NaM** GGC GCC GTG GTC)) was prepared using standard automated DNA synthesis methodology with ultra-mild DNA synthesis phosphoramidites on CPG ultramild supports (15 μmol, Glen Research; Sterling, VA) and an ABI Expedite 8905 synthesizer. After automated synthesis, the DMT-ON oligonucleotide was first purified by Glen-Pak cartridge (Glen Research) and then by 8 M urea 20% PAGE, followed by Synergi Fusion-RP HPLC (Phenomenex, Torrence, CA) to single-band purity (>98%) using acetonitrile/100 mM triethyl ammonium bicarbonate buffer (pH 7.5) gradient (5–25% ACN over 30 min) MS (MALDI-ToF, matrix: THAP) m/z: $[M+H]^+$ calcd. 4955.3; found 4954.6.

Primer-ddC (5'-d(GAC CAC GGC GC)ddC-3) containing a 2',3'-dideoxy end (ddC) was synthesized on a Applied-Biosystems 392 DNA/RNA synthesizer using the 2',3'-ddC-CPG (5'-dimethoxytrityl-N-succinoyl-long chain alkylamino-CPG, 2',3'-deoxycytosine), which was purchased from Glen Research. The primer was purified by two rounds of reverse phase HPLC (DMT-ON followed by DMT-OFF)

Crystallization and Structure Determination. Purified KlenTaq enzyme was stored at 4°C in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA and 1 mM β-mercaptoethanol. Primers and templates were annealed prior to addition of protein and triphosphates.

All crystals were obtained using the hanging drop vapor diffusion method by equilibration against 0.5 ml of the respective reservoir solution for 4–8 days at 18°C. Protein and reservoir solutions were mixed in a 1:1 ratio. Prior to measurement crystals were flash frozen in liquid nitrogen using the respective reservoir solution and 10–20% glycerol or ethylene glycol as a cryoprotectant.

$KTQ_{dG\text{-}dCTP}$ (PDB ID: 3RTV): The closed ternary complex of $KTQ_{dG\text{-}dCTP}$ was obtained by incubating KlenTaq (10 mg/ml) with a primer-11mer/template-G complex, ddGTP and dCTP in a molar ratio of 1:2:10:10 in the presence of 20 mM $MgCl_2$. After incorporation of ddGMP the polymerase is trapped in a catalytically active complex with dCTP bound in the waiting position and two coordinated magnesium ions. Crystals grew under the following conditions: 30% PEG 4000, 0.2 M NH_4OAc, 0.01 M $Mg(OAc)_2$ and 0.05 M sodium cacodylate (pH 6.5).

KTQ_{dG} (PDB ID: 3SZ2): The open binary complex of KTQ_{dG} was obtained by incubating KlenTaq (7.5 mg/ml) with a primer-11mer/template-G complex and ddGTP in a molar ratio of

Nature Chemical Biology: doi:10.1038/NChemBio.966

1:1.5:5 in the presence of 20 mM $MgCl_2$. Crystals grew under the following conditions: 15% PEG 8000, 0.2 M $Mg(HCOO)_2$ and 0.1 M Tris HCl (pH 8.0).

KTQ_{dT} (PDB ID: 3SV4): The open binary complex KTQ_{dT} was obtained by incubating KlenTaq (7.0 mg/ml) with a primer-11mer/template-T complex and ddCTP in a molar ration of 1:1.2:5 in the presence of 20 mM $MgCl_2$. Crystals grew under the following conditions: 18% PEG 8000, 0.2 M $Mg(HCOO)_2$ and 0.1 M Tris HCl (pH 8.5).

KTQ_{dNaM} (PDB ID: 3SYZ): For crystallization of the binary complex of KTQ_{dNaM} KlenTaq (8.0 mg/ml) was incubated with a primer-ddC/template-**NaM** complex in a molar ratio of 1:1.5 in the presence of 20 mM $MgCl_2$. Crystals grew under the following conditions: 11% PEG 8000, 20 mM $MgCl_2$, 0.2 M $Mg(HCOO)_2$, 0.1 M Tris HCl (pH 8.0).

$KTQ_{dNaM\text{-}d5SICSTP}$ (PDB ID: 3SV3): Crystals of the ternary complex were obtained by soaking of crystals of the binary complex KTQ_{dNaM} for 20 h with 6.7 mM d**5SICS**TP. Crystals of KTQ_{dNaM} used for soaking with d**5SICS**TP grew in 13% PEG 8000, 0.1 M NH_4Cl, 0.1 M $Mg(HCOO)_2$, 0.1 M Tris HCl (pH 8.0). During soaking, two d**5SICS**TP molecules bound to the polymerase. One molecule is located in the active site as expected and a second molecule packs between R801 of one polymerase and Q633 and P816 of a symmetry related polymerase molecule. The second triphosphate seems well stabilized through a π-cation interaction with R801 as well as several hydrogen-bonds.

Data was collected at the beamline PXIII (XO6DA) and PXI (XO6SA) at the Swiss Light Source of the Paul Scherrer Institute in Villigen, Switzerland. Data reduction was performed with the XDS package[10]. Structures were solved by rigid-body refinement against a previously published KlenTaq structure (PDB: 3M8S[1]) as model. Refinement was performed using PHENIX[11] and model rebuilding was done with COOT[12]. TLS groups for refinement were generated using TLSMD server[13]. During refinement structures were evaluated using the molprobity server (http://molprobity.biochem.duke.edu/). Figures were created with PyMOL[14]. For overlays in the figures structures were aligned in PYMOL using all atoms. All structures are in the same space group $P3_121$ with similar cell dimensions (Supplementary Table 1).

*NMR analysis of free DNA duplex with d**NaM**-d**5SICS** unnatural base pair.*

Sample preparation. d**5SICS** oligonucleotide was prepared as described[15]. The d**NaM** oligonucleotide (Supplementary Fig. 1a) was prepared using standard automated DNA synthesis methodologies with ultra-mild DNA synthesis phosphoramidites on CPG ultramild supports (3*1 μmol, Glen Research; Sterling, VA) and an ABI Expedite 8905 synthesizer. After automated synthesis, the DMT-ON oligonucleotide was first purified by Glen-Pak cartridge (Glen Research) and then by Clarity Oligo-RP HPLC (Phenomenex, Torrence, CA) to single-band purity (>98%) using acetonitrile/100 mM triethyl ammonium bicarbonate buffer (pH 7.5) gradient (5–25% ACN over 30 min). MS (MALDI-ToF, matrix: THAP) m/z: $[M+H]^+$ calcd. 3760.6; found 3759.6.

After desalting (Sephadex NaP-10 column; GE Healthcare), the oligonucleotides were lyophilized. Oligomer concentrations were determined from UV absorbance at 260 nm, with the extinction coefficient for both d**5SICS** and d**NaM** estimated to be the same as that of guanine. The oligomers were mixed in a 1:1 stoichiometry, annealed at 85°C, and then dialyzed against two liters of 0.50 M NaCl followed by two liters of deionized water in a teflon spin dialyzer (5 mL, Harvard Apparatus) using a 2 kD MWCO cellulose-acetate membrane (Harvard Apparatus). The purity of the duplex DNA sample was greater than 95% based on NMR. The NMR sample was prepared by dissolving the DNA duplex in 10 mM sodium phosphate buffer (pH 7.0)

Nature Chemical Biology: doi:10.1038/NChemBio.966

containing 100 mM NaCl and 0.1 mM EDTA and then lyophilizing and re-dissolving in 99.99% D_2O (Cambridge Isotope Laboratories) to a final volume of 500 μL. The total DNA strand concentration in the resulting samples was 1 mM. The numbering scheme utilized is shown in Supplementary Fig. 1a.

NMR spectroscopy. Proton NMR spectra in D_2O for the duplex sample were acquired on a Varian Inova 500 MHz spectrometer. NOESY and DQF-COSY spectra were acquired using the TPPI method of phase cycling[16]. Data were collected at 25 °C and 30°C to resolve cross peak overlap in the spectra. For signal assignments, NOESY spectra with a mixing time of 300 ms were collected with a spectral width of 5913 Hz, 2048 complex points in t_2 and 512 t_1 increments (zero filled to 2048 on processing); for each t_1 value 64 scans were averaged using a recycle delay of 2 s. Presaturation was applied during the recycle delay and the mixing time to suppress the residual HOD resonance. DQF-COSY spectra were collected with 2048 complex points in t_1, 512 t_1 increments (zero filled to 2048 points on processing) with a spectral width of 5913 Hz; for each t_1 value 64 scans were collected using a recycle delay of 2 s with presaturation of the HOD resonance. All spectra were transferred to a PC laptop and processed with Felix (FelixNMR).

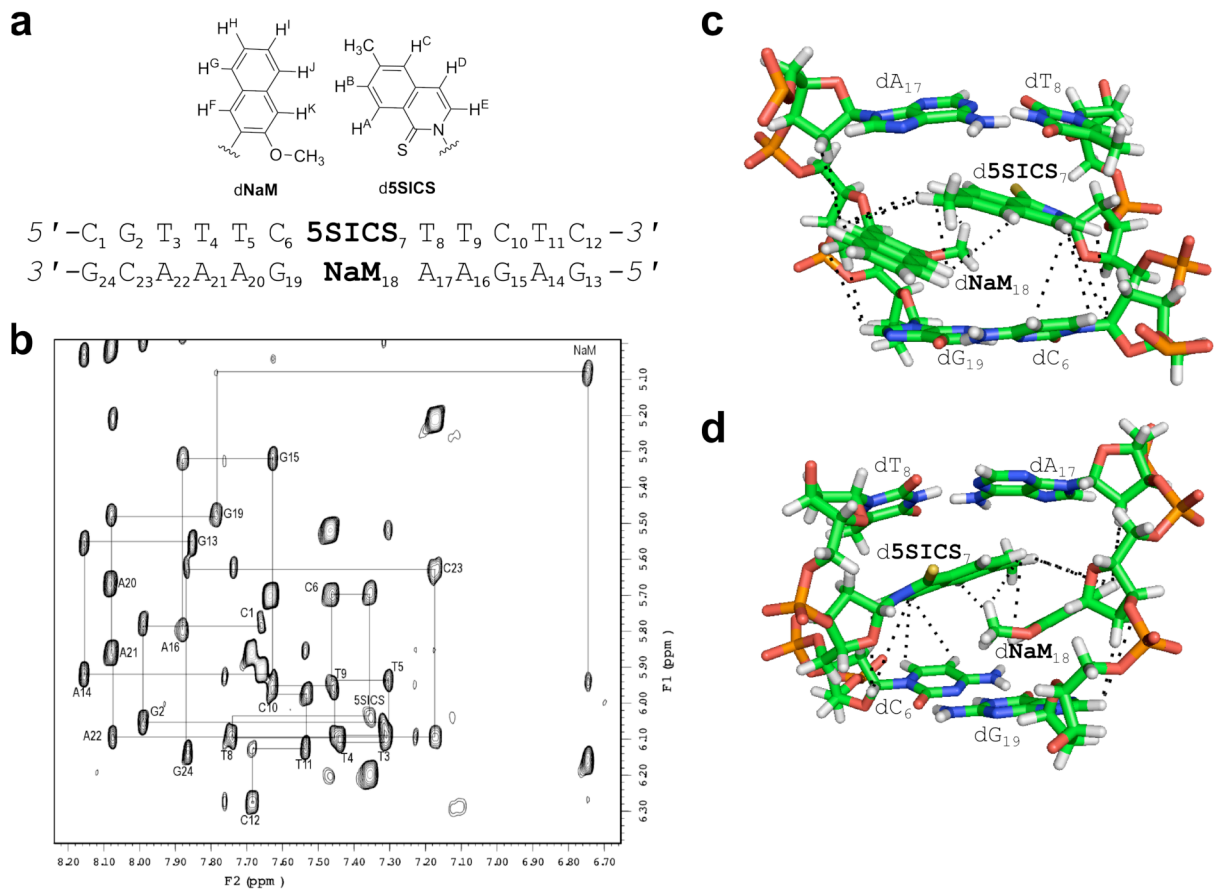
Signal assignments and NMR observations. Resonance assignments for the d**NaM**-d**5SICS** duplex followed conventional NOESY based methods[17]. The NOESY and DQF-COSY spectra support one predominant form of the d**NaM**-d**5SICS** duplex in solution. The deoxyribose C1'H of d**NaM** is shifted farther upfield (5.09 ppm) than other anomeric protons while the sequential NOE connectivities of d**5SICS** H^E to d**5SICS** C1'H and dNaM H^G to A19 C1'H are weaker than other bases. These data suggest that distortion of the duplex is minimal and localized to the region of the d**NaM**-d**5SICS** pair. Sequential connectivity of the aromatic protons with C2'H and C2"H confirmed the analysis of the C1'H-C2'H and C1'H-C2"H cross peak patterns in the DQF-COSY spectra, leading to unambiguous assignment of these protons (Supplementary Table 2). The peak patterns in the proton NOESY (Supplementary Fig. 1b) and DQF-COSY spectra for deoxyribose rings suggest largely C3'-endo conformations.

d**NaM**-d**5SICS mode of pairing.** The NMR spectral data of provide evidence that the d**NaM**-d**5SICS** pair is positioned inside the double helix in a self-intercalating fashion. Key NOEs in the NOESY spectrum in support of this include d**5SICS** H^D and H^E to C6 H5 and C6 H6, plus H^D and C6 C1'H, in addition to cross-strand NOEs between d**NaM** -OCH_3 and d**5SICS** H^B and H^C and d**5SICS** H^B and d**NaM** H^F and H^K as well as d**5SICS** H^C and d**NaM** H^K (Supplementary Table 3 and Supplementary Fig. 1c). Similar to d**5SICS**:d**MMO2**[15], cross-strand NOEs between d**NaM** -OCH_3 and d**5SICS** H_B, between d**5SICS** H^B and d**NaM** H^F and H^K, between d**5SICS** -CH_3 and d**NaM** H^F and H^G, a lack of NOEs from H^C, H^D and H^E of **5SICS** to any proton on **NaM**, provide evidence that the d**5SICS**:d**NaM** pair is and oriented such that the d**NaM** -OCH_3 group and the d**5SICS** C=S are positioned with in the minor groove of the duplex.

Modeling of the d**NaM**-d**5SICS duplex** was carried out based on the duplex with d**MMO2**-d**5SICS**, whose structure was previously determined by NMR[15]. The charge distribution and geometrical parameters for the d**NaM** nucleobase were obtained from DFT (B3LYP/6-31G*) calculations in Gaussian[18]; d**MMO2** was then replaced by d**NaM** in the duplex. The system was subjected to 5000 steps of unconstrained conjugate gradient energy minimization in the Sander module of AMBER[19] resulting in the model duplex (Supplementary Fig. 1c).

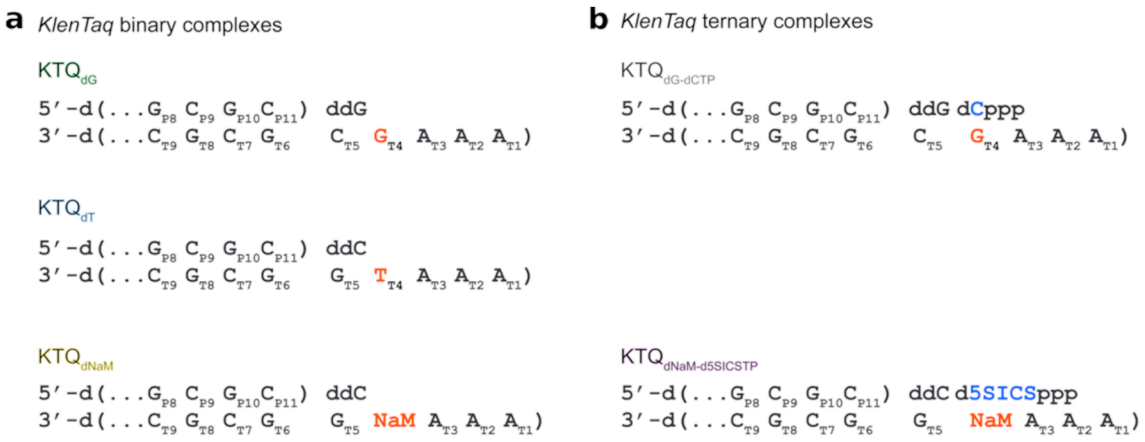
Nature Chemical Biology: doi:10.1038/NChemBio.966

Supplementary Results



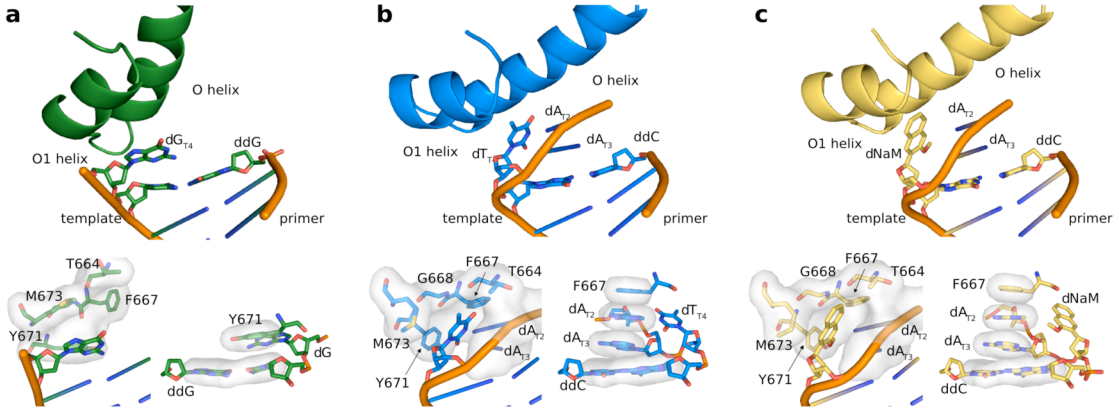
Supplementary Figure 1. NMR analysis of free DNA duplex with dNaM-d5SICS unnatural base pair. (a) Sequence of the duplex characterized by NMR and chemical structure of d**NaM**–d**5SICS** with atoms labeled. Only nucleobase moieties are shown; sugar and phosphate backbone are omitted for clarity. **(b)** Deoxyribose C1'H to aromatic portion of the NOESY spectra of the d**NaM**–d**5SICS** duplex at 25 °C with a mixing time 300 ms. Solid lines indicate the sequential connectivity for the C1'H to H^6/H^8 for both strands including H^E for d**5SICS** and H^G for **dNaM**. **(c, d)** Key NOEs (black dotted lines) from Supplementary Table 3 depicted on a modeled d**NaM**-d**5SICS** duplex. Only d**NaM**-d**5SICS** pair and flanking base pairs are shown as viewed from the major (**c**) or the minor (**d**) groove.

Nature Chemical Biology: doi:10.1038/NChemBio.966



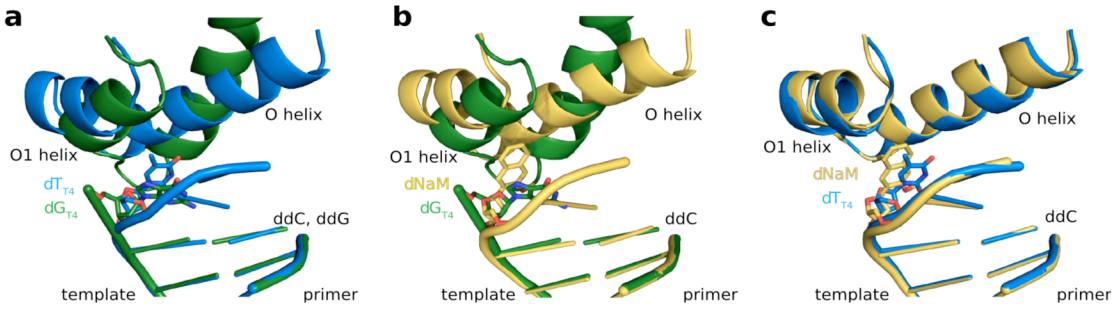
Supplementary Figure 2. Partial DNA sequences employed to obtain binary (a) and ternary (b) KlenTaq complexes solved in this study. The templating bases at the insertion position are colored red and the bound triphosphates in the ternary complexes are colored blue.

Nature Chemical Biology: doi:10.1038/NChemBio.966



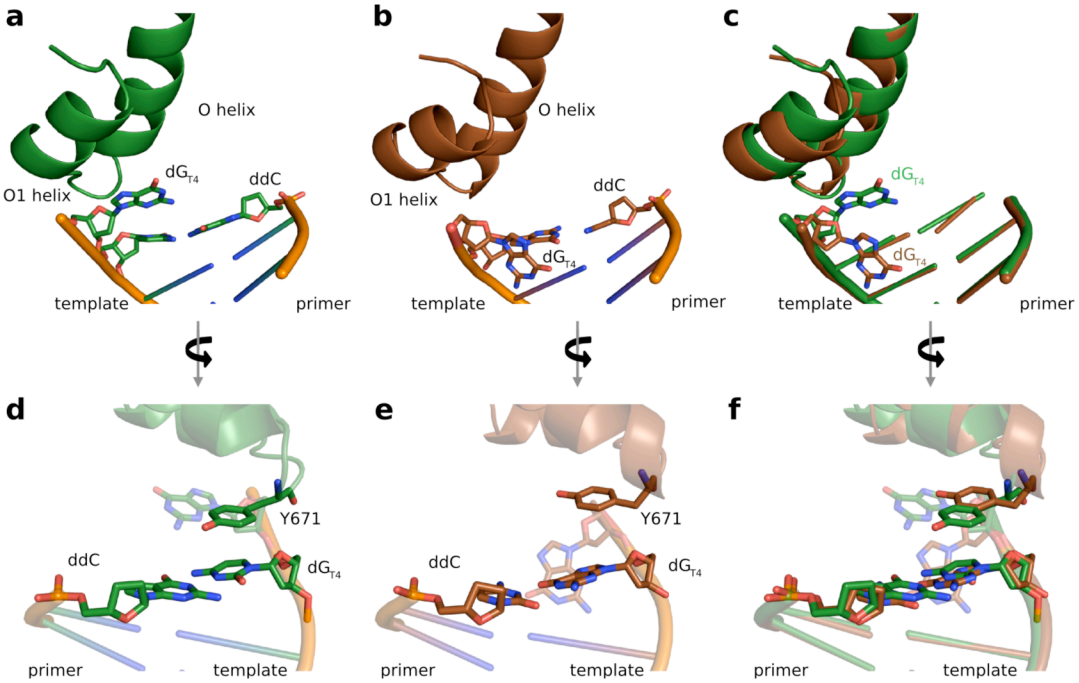
Supplementary Figure 3. Close-up view of KlenTaq binary structures. (**a**) KTQ_{dG} (green), (**b**) KTQ_{dT} (blue), and (**c**) $KTQ_{\mathbf{dNaM}}$ (yellow). At the top of each panel is shown the O and O1 helices and the primer-template end as cartoon and the nascent base pair and templating nucleotide in stick representation. Residues and their surfaces that accommodate the templating nucleobase are shown at the bottom left of each panel and the stacking arrangement of the primer-template with either Y671 (KTQ_{dG}) or downstream nucleobases of dA_{T3}, dA_{T2} and F667 (KTQ_{dT} and $KTQ_{\mathbf{dNaM}}$) is shown at the bottom right of each panel.

Nature Chemical Biology: doi:10.1038/NChemBio.966



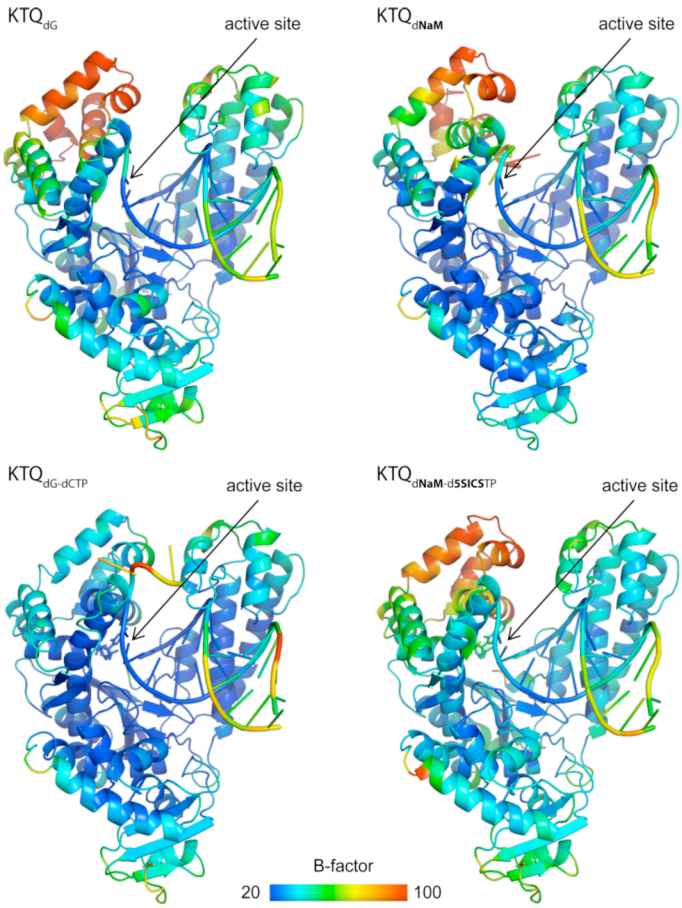
Supplementary Figure 4. Comparison of binary structures. (a) Overlay of KTQ_{dT} (blue) and KTQ_{dG} (green). **(b)** Overlay of KTQ_{dG} (green) and $KTQ_{\mathbf{dNaM}}$ (yellow). **(c)** Overlay of KTQ_{dT} (blue) and $KTQ_{\mathbf{dNaM}}$ (yellow).

Nature Chemical Biology: doi:10.1038/NChemBio.966



Supplementary Figure 5. Comparison of the binary complex KTQ$_{dG}$ (green) and the binary structure published by Waksman *et al.*[3] 4KTQ (brown). (**a-c**), Same view as in Supplementary Fig. 3 (upper panel) for KTQ$_{dG}$, 4KTQ, and an overlay of both structures. (**d-f**), Stacking of the protein side chain Y671 on top of the nascent base pair.

Nature Chemical Biology: doi:10.1038/NChemBio.966



Supplementary Figure 6. B-factors of KlenTaq binary and ternary complexes with natural and unnatural substrates. Overall structures of KTQ_{dG}, KTQ_{dNaM}, $KTQ_{dG\text{-}dCTP}$ and $KTQ_{dNaM\text{-}d5SICSTP}$ colored by B-factors using a minimum value of 20 (blue) and maximum value of 100 (red) (generated using PyMOL[14]). The spectrum bar indicates the color gradient from low to high B-factors.

Nature Chemical Biology: doi:10.1038/NChemBio.966

Supplementary Table 1. Data collection and refinement statistics

	KTQ_{dG}	$KTQ_{dG\text{-}dCTP}$	$KTQ_{\mathbf{dNaM}}$	$KTQ_{\mathbf{dNaM}\text{-}d\mathbf{5SICS}TP}$	KTQ_{dT}
Data collection					
Space group	$P3_121$	$P3_121$	$P3_121$	$P3_121$	$P3_121$
Cell dimensions					
a, *b*, *c* (Å)	108.4, 108.4, 89.8	108.3, 108.3, 90.4	108.5, 108.5, 90.5	108.5, 108.5, 91.0	108.7, 108.7, 90.2
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120
Resolution (Å)	46.9-2.15 (2.28-2.15)	46.9-1.89 (2.01-1.89)	47.0-1.95 (2.07-1.95)	47.0-2.10 (2.23-2.10)	47.1-2.00 (2.12-2.00)
R_{meas}	16.6 (142.4)	8.9 (95.0)	13.6 (102.4)	15.7 (136.3)	11.7 (102.0)
$I / \sigma I$	14.77 (1.90)	17.27 (2.33)	11.23 (2.21)	12.41 (2.33)	13.79 (2.06)
Completeness (%)	100.0 (99.9)	99.7 (98.5)	99.8 (98.6)	99.9 (99.5)	99.9 (99.6)
Redundancy	11.1 (11.0)	9.9 (10.0)	10.9 (10.2)	10.0 (9.8)	10.0 (9.6)
Refinement					
Resolution (Å)	46.9-2.15 (2.18-2.15)	46.9-1.90 (1.92-1.90)	46.6-1.95 (1.97-1.95)	47.0-2.10 (2.13-2.10)	46.9-1.99 (2.02-1.99)
No. reflections	64185	93428	86484	69989	80646
R_{work} / R_{free}	19.0 / 23.7 (30.3 / 32.6)	16.2 / 20.2 (26.9 / 31.1)	16.0 / 20.0 (26.2 / 28.0)	16.7 / 21.5 (26.6 / 30.6)	17.9 / 21.6 (31.9 38.0)
No. atoms					
Protein	8474	4374	8711	8692	8626
DNA / Triphosphate	510 / -	570 / 28	583 / -	511 / 64	562 / -
Water	290	331	314	232	303
B-factors					
Protein	44.9	35.1	42.0	48.8	44.0
DNA (primer / template)	39.5 / 40.8	37.2 / 41.2	38.6 / 45.6	41.4 / 41.2	40.5 / 43.3
Triphosphate, active site	-	22.8	-	48.7	-
Water	39.2	40.1	42.2	45.3	43.5
R.m.s. deviations					
Bond lengths (Å)	0.003	0.007	0.006	0.008	0.003
Bond angles (°)	0.758	1.102	1.077	1.144	0.770

For each structure a single crystal was used for data collection. Highest-resolution shell is shown in parentheses.

Nature Chemical Biology: doi:10.1038/NChemBio.966

Supplementary Table 2. Proton chemical shift assignments (ppm) for the d**NaM**-d**5SICS** duplex at 25 °C.[a]

Nucleotide	H6/H8	H5	C1'H	C2'H	C2''H	C3'H	-CH_3
C1	7.66	5.91	5.78	2.05	2.46	-[c]	na[b]
G2	7.99	na	6.05	2.70	2.84	5.00	na
T3	7.31	na	6.11	2.07	2.29	4.89	1.41
T4	7.44	na	6.09	2.17	2.60	-	1.62
T5	7.31	na	5.94	2.08	2.56	-	1.93
C6	7.46	5.53	5.69	2.31	2.51	4.84	na
5SICS	7.36	na	6.04	2.72	2.72	-	1.57
T8	7.75	na	6.10	2.30	2.56	-	1.73
T9	7.45	na	5.96	2.80	2.50	-	1.62
C10	7.63	5.70	5.97	2.20	2.51	-	na
T11	7.54	na	6.12	2.21	2.52	4.89	1.72
C12	7.69	5.87	6.27	2.27	2.27	4.57	na
G13	7.85	na	5.56	2.44	2.63	4.81	na
A14	8.15	na	5.92	2.70	2.82	5.04	na
G15	7.63	na	5.32	2.43	2.53	4.95	na
A16	7.88	na	5.80	2.30	2.59	4.98	na
A17	-	na	5.94	2.31	1.93	-	na
NaM	6.75	na	5.09	2.17	2.17	-	na
G19	7.79	na	5.48	2.59	2.69	4.96	na
A20	8.08	na	5.67	2.64	2.76	5.03	na
A21	8.08	na	5.86	2.63	2.81	5.02	na
A22	8.07	na	6.10	2.54	2.80	5.22	na
C23	7.17	5.22	5.63	1.80	2.26	-	na
G24	7.87	na	6.14	2.33	2.33	4.97	na

5SICS	H^A	H^B	H^C	H^D	H^E	-CH_3	
	5.26	7.11	6.04	6.20	7.36	1.57	
NaM	H^F	H^G	H^H	H^I	H^J	H^K	-OCH_3
	6.17	6.75	7.11	6.23	6.05	6.29	3.55

[a]Chemical shifts are referenced to residual HOD resonance.
[b]Not applicable.
[c]Not assigned due to overlap.

Nature Chemical Biology: doi:10.1038/NChemBio.966

Supplementary Table 3. Key NOEs defining the d**NaM**-d**5SICS** base pair

d5SICS proton	**dNaM proton**	**DNA proton**
H^B	H^F	
H^B	H^K	
H^B	$-OCH_3$	
H^C	H^K	
H^D		C6 C1'H
H^D		C6 H5
H^D		C6 H6
H^E		C6 C1'H
H^E		C6 H5
	H^F	A17 C2'H/C2"H
	C2'H/C2"H	G19 H8
$-CH_3$	H^F	
$-CH_3$	H^G	

Nature Chemical Biology: doi:10.1038/NChemBio.966

Supplementary References

1. Betz, K. et al. Structures of DNA polymerases caught processing size-augmented nucleotide probes. *Angew. Chem. Int. Ed.* 49, 5181-5184 (2010).
2. Li, Y., Kong, Y., Korolev, S. & Waksman, G. Crystal structures of the Klenow fragment of *Thermus aquaticus* DNA polymerase I complexed with deoxyribonucleoside triphosphates. *Protein Sci.* 7, 1116-1123 (1998).
3. Li, Y., Korolev, S. & Waksman, G. Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO Journal* 17, 7514-7525 (1998).
4. Li, Y., Mitaxov, V. & Waksman, G. Structure-based design of Taq DNA polymerases with improved properties of dideoxynucleotide incorporation. *Proc. Natl. Acad. Sci. USA* 96, 9491-9496 (1999).
5. Li, Y. & Waksman, G. Crystal structures of a ddATP-, ddTTP-, ddCTP-, and ddGTP-trapped ternary complex of Klentaq1: Insights into nucleotide incorporation and selectivity. *Prot. Sci.* 10, 1225-1233 (2001).
6. Obeid, S., Baccaro, A., Welte, W., Diederichs, K. & Marx, A. Structural basis for the synthesis of nucleobase modified DNA by *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. USA* 107, 21327-21331 (2010).
7. Obeid, S. et al. Replication through an abasic DNA lesion: structural basis for adenine selectivity. *EMBO J.* 29, 1738-1747 (2010).
8. Ludwig, J. & Eckstein, F. Rapid and efficient synthesis of nucleoside 5'-0-(1-thiotriphosphates), 5'-triphosphates and 2',3'-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.* 54, 631 (1989).
9. Lavergne, T., Malyshev, D.A. & Romesberg, F.E. Major groove substituents and polymerase recognition of a class of predominantly hydrophobic unnatural base pairs. *Chem. Eur. J.* 18, 1231-1239 (2012).
10. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* 26, 795 (1993).
11. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr.* 66, 213-221 (2010).
12. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr.* 60, 2126-2132 (2004).
13. Painter, J. & Merritt, E.A. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr.* 62, 439-450 (2006).
14. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.
15. Malyshev, D.A. et al. Solution structure, mechanism of replication, and optimization of an unnatural base pair. *Chem. Eur. J.* 16, 12650-12659 (2010).
16. Drobny, G., Pines, A., Sinton, S., Weitekamp, D. & Wemmer, D.E. Fourier Transform Multiple Quantum NMR. *Faraday Div. Chem. Soc. Symp.* 13, 49-55 (1979).
17. Hare, D.R., Wemmer, D.E., Chou, S.H., Drobny, G. & Reid, B.R. Assignment of the non-exchangeable proton resonances of d(C-G-C-G-A-A-T-T-C-G-C-G) using two-dimensional nuclear magnetic resonance methods. *J Mol Biol* 171, 319-336 (1983).
18. Frisch, M.J. et al. Gaussian 98. Revision A.7 edn (Gaussian, Inc., Pittsburgh, PA, 1998).
19. Case, D.A. et al. *AMBER 9*, (University of California, San Francisco, 2006).

Nature Chemical Biology: doi:10.1038/NChemBio.966