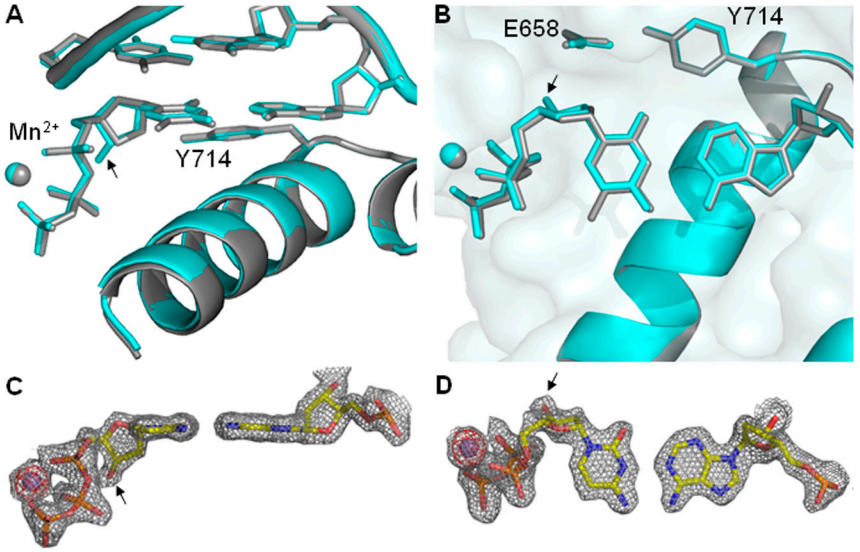
Supporting Information

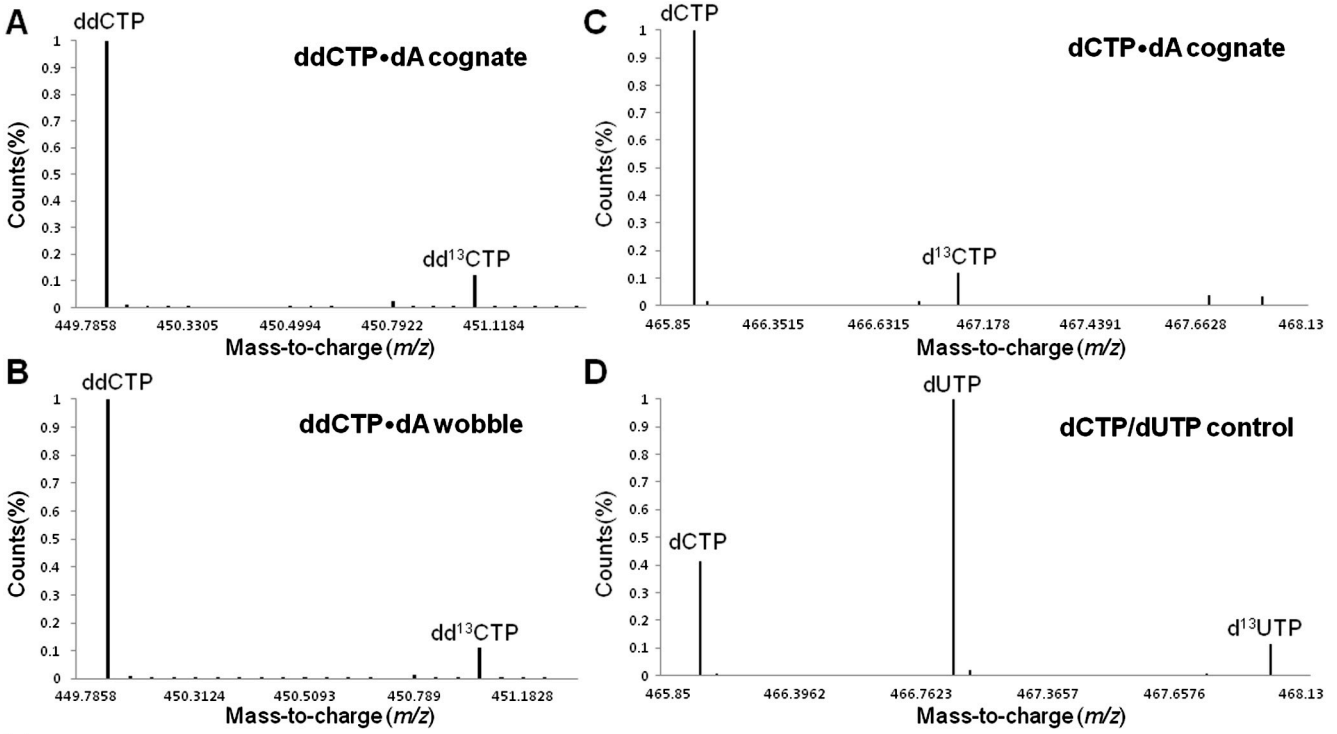
Wang et al. 10.1073/pnas.1114496108



Base pair parameters of dCTP•dA at the polymerase insertion site

Base pair	λ_{primer}(°)	$\lambda_{template}$(°)	$d_{C1'-C1'}$(Å)	Shear(Å)	Stretch(Å)	Stagger(Å)	Buckle(°)	Propeller(°)	Opening(°)
Molecule 1									
dCTP•dA	53.8	56.6	10.6	0.03	-0.01	0.02	5.84	-1.78	4.29

Fig. S1. Structure of a dCTP•dA base pair at the polymerase insertion site. A complex of dCTP•dA was captured by exchanging ddCTP in ddCTP•dA crystals with dCTP. Comparisons with a cognate dideoxy base pair captured at the insertion site show that dCTP•dA adopts a similar shape. (*A*), (*B*) In both deoxy (cyan) and dideoxy (gray) structures, the O helix is closed (*A*), the triphosphate is undistorted, and the active site fully assembled (*B*). (*C*, *D*) Two views of composite omit maps (gray) at 1.5σ contour of the dCTP•dA base pair showing the cognate base pair shape and 3′ hydroxyl group (arrow). The presence of Mn^{2+} ion (purple), which is observed in the deoxynucleotide, is confirmed by anomalous difference map (red). The structural comparisons were based on the more ordered molecule A.



E

Complex	Nucleotide	Molecular formula	Accurate mass $(M-H)^-$ (all ^{12}C)		Accurate mass $(M-H)^-$ (one ^{13}C)		$^{13}C/^{12}C$ ratio (%)	
			Theoretical	Observed	Theoretical	Observed	Theoretical	Observed
ddCTP•dA cognate	ddCTP	$C_9H_{16}N_3O_{12}P_3$	449.9874	449.9883	450.9908	450.9907	11.1	12.1
	ddUTP	$C_9H_{15}N_2O_{13}P_3$	450.9714	-	451.9748	-		
ddCTP•dA wobble	ddCTP	$C_9H_{16}N_3O_{12}P_3$	449.9874	449.9880	450.9908	450.9907	11.1	11.2
	ddUTP	$C_9H_{15}N_2O_{13}P_3$	450.9714	-	451.9748	-		
dCTP•dA cognate	dCTP	$C_9H_{16}N_3O_{13}P_3$	465.9823	465.9833	466.9857	466.9846	11.1	12.1
	dUTP	$C_9H_{15}N_2O_{14}P_3$	466.9663	-	467.9697	-		
dCTP/dUTP control	dCTP	$C_9H_{16}N_3O_{13}P_3$	465.9823	465.9831	466.9857	-		
	dUTP	$C_9H_{15}N_2O_{14}P_3$	466.9663	466.9684	467.9697	467.9722	11.1	11.3

Fig. S2. Investigation of cytosine deamination by Mass Spectrometry. Mass spectrographs of the crystallization drops of ddCTP•dA cognate (*A*), ddCTP•dA wobble (*B*) crystals, soaking solution of dCTP•dA cognate crystals (*C*), and a solution with dUTP added to the dCTP•dA cognate soaking solution (*D*) are shown. Theoretical and observed masses of ddCTP or dCTP and correspondent deaminated ddUTP or dUTP are listed (*E*). The observed peak with the smaller m/z ratio in (*A*), (*B*), and (*C*) is consistent with ddCTP, ddCTP, and dCTP respectively. The peaks with the larger m/z in samples (*A*–*C*) are consistent with the correspondent ^{13}C isotope of the nucleotides respectively. The ratio of ^{13}C isotope to ^{12}C is within the error of experimental measurements (*E*). In the positive control containing both dCTP and dUTP (*D*), the peak with smaller m/z is consistent with dCTP, the intermediate peak with dUTP ($d^{13}CTP$ is probably masked by dUTP), and the larger peak with $d^{13}UTP$. No detectable ddUTP or dUTP peaks were observed at correspondent m/z in the mass spectrographs of solutions where the C•A crystals were harvested from. Taken together, these results show that no detectable amount of ddUTP or dUTP is formed in the C•A crystals.

Table S1. Crystallographic data collection and refinement statistics

	ddCTP•dA cognate (*n*)	ddCTP•dA wobble (*n*)	dCTP•dA (*n*)	ddTTP•dA (*n*)	ddATP•dT (*n*)	ddGTP•dC (*n*)	C•A (*n*-1)	C•A (*n*-3)	C•A (*n*-4)	C•A (*n*-6)
Data collection										
Resolution (Å)	100-1.59	100-1.58	100-1.73	100-1.52	100-1.61	100-1.62	50-1.53	100-1.65	50-1.65	50-1.60
Outer shell (Å)	1.68-1.59	1.68-1.58	1.83-1.73	1.61-1.52	1.71-1.61	1.71-1.62	1.62-1.53	1.76-1.65	1.75-1.65	1.70-1.60
R_{sym}	7.4(56.4)*	4.4(60.1)	5.8(58.3)	6.9(49.2)	6.6(49.3)	7.3(49.3)	3.3(48.0)	4.5(48.9)	4.6(47.0)	4.3(49.7)
$I/\sigma I$	17.6(3.2)	21.9(2.4)	18.0(2.6)	14.9(4.2)	14.2(3.6)	12.2(3.5)	26.2(4.1)	19.9(3.4)	21.2(3.4)	22.2(3.8)
Completeness (%)	99.0(99.9)	95.4(89.5)	98.3(99.6)	97.5(95.4)	97.2(91.4)	98.3(99.6)	97.9(99.6)	98.1(91.6)	97.4(99.5)	99.8(99.2)
Redundancy	8.3	5.4	5.8	7.2	5.6	5.5	4.8	4.9	5.0	5.7
Refinement										
Resolution (Å)	87.7-1.59	71.1-1.58	34.2-1.73	88.4-1.52	88.4-1.61	43.0-1.62	41.1-1.53	40.7-1.66	40.8-1.65	45.3-1.60
No. reflections	192,561	188,106	149,632	222,603	185,284	180,731	120,984	96,815	95,402	111,078
R_{work}/R_{free}†	18.5/20.7	21.1/23.9	19.5/22.4	18.5/21.1	19.3/22.4	18.7/21.5	17.0/18.5	17.6/20.2	16.7/19.0	17.8/19.2
No. nonhydrogen atoms										
Total	11,542	11,408	11,476	12,090	11,962	11,936	6,036	5,896	5,988	6,041
Solvent	1,202	1,468	1,368	1,823	1,696	1,687	873	697	775	670
B-factor	27.8	27.3	29.0	23.6	25.7	25.9	26.6	28.0	25.4	29.6
rms deviations										
Bond lengths (Å)	0.013	0.008	0.009	0.009	0.010	0.008	0.011	0.009	0.010	0.010
Bond angles (°)	1.489	1.169	1.193	1.245	1.266	1.221	1.358	1.223	1.243	1.290
Expected maximal error (Å)‡	0.051	0.056	0.060	0.044	0.050	0.053	0.077	0.088	0.081	0.074
Expected minimal error (Å)§	0.009	0.010	0.013	0.008	0.011	0.012	0.011	0.019	0.021	0.014
Crystal form¶	II	II	II	II	II	II	I	I	I	I
PDB code	3PX6	3PX4	3PX0	3PV8	3THV	3TI0	3TAN	3TAP	3TAQ	3TAR

Data for this study were collected at SIBYLS and SER-CAT beamlines. Use of SIBYLS beamline at the Advanced Light Source, Lawrence Berkeley National Laboratory, was supported in part by the DOE program Integrated Diffraction Analysis Technologies (IDAT) and the DOE program Molecular Assemblies Genes and Genomics Integrated Efficiently (MAGGIE) under Contract Number DE-AC02-05CH11231 with the Department of Energy. Use of the Advanced Photon Source was supported by the Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-3.

*Numbers in parentheses correspond to parameter values in the outer resolution shell.

†5% free reflections were generated in XDSCONV (1) by combining inherited free reflections from the starting model (1L3T, 1L5U, 1L5U, and 1L3V for Crystal Form I C•A (*n*-1), C•A (*n*-3), C•A (*n*-4), and C•A (*n*-6) respectively and 2HVI for all Crystal Form II) and randomly selected reflections beyond resolution of the starting model.

‡Expected maximal and minimal error were calculated in SFcheck (2).

§Expected maximal and minimal error were calculated in SFcheck (2).

¶Crystal Form I and II both belong to space group $P2_12_12_1$. In Crystal Form I, cell dimensions are: $a = 87$ Å, $b = 93$ Å, $c = 105$ Å. There is one molecule in the asymmetric unit and the molecule adopts an open protein conformation. In Crystal Form II, cell dimensions are: $a = 93$ Å, $b = 108$ Å, $c = 149$ Å. There are two molecules in the asymmetric unit.

1 Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell dimensions. *J Appl Cryst* 26:795–800.

2 Vaguine AA, Richelle J, Wodak SJ (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr D Biol Crystallogr* 55:191–205.

Table S2 Substrates for preparing BF primer-template complexes

Nucleotides placed at the insertion site

Name	Primer (bottom) and template (top) sequences	Nucleotides and metal ions used in cocrystallization
ddCTP•dA wobble	5′-CATAGGAGTCAGG-3′ 3′-CTCAGTCC-5′	ddCTP, Mg^{2+}
ddCTP•dA cognate	5′-CATAGGAGTCAGG-3′ 3′-CTCAGTCC-5′	ddCTP, Mn^{2+}
ddTTP•dA	5′-CATAAGAGTCAGG-3′ 3′-CTCAGTCC-5′	ddTTP, Mg^{2+}
ddATP•dT	5′-CATTTGAGTCAGG-3′ 3′-CTCAGTCC-5′	ddATP, Mg^{2+}
ddGTP•dC	5′-CATCCGAGTCAGG-3′ 3′-CTCAGTCC-5′	ddGTP, Mg^{2+}

Mismatches incorporated into the DNA duplex

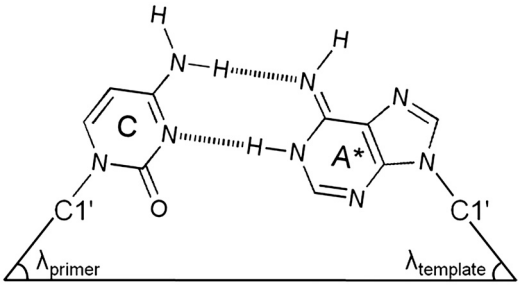
Position	Primer (bottom) and template (top) sequences	Nucleotides used in catalysis in the crystal
C•A (*n*-1)	5′-GACGTACGTGATCGCA-3′ 3′-GCACTAGCG-5′	dCTP
C•A (*n*-1) presynthesized	5′-GACGTACGTGATCGCA-3′ 3′-CGCACTAGCG-5′	None
C•A (*n*-3)	5′-GACGTACGTGATCGCA-3′ 3′-CGCACTAGCG-5′	dATP, dCTP
C•A (*n*-4)	5′-GACGTACGTGATCGCA-3′ 3′-CGCACTAGCG-5′	dATP, dCTP, dGTP
C•A (*n*-6)	5′-GACGTACGTGATCGCA-3′ 3′-CGCACTAGCG-5′	dATP, dCTP, dGTP, dTTP

Oligonucleotides were synthesized at GF grade from Midland Certified Reagent Co. (Midland, TX) and annealed to form duplexes as described (1). Ultrapure ddNTPs were purchased from USB Co. (Cleveland, OH), and dNTPs from Promega Co. (Madison, WI)

1 Kiefer JR, Mao C, Braman JC, Beese LS (1998) Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Nature* 391:304–307.

Table S3 DNA base pair parameters at the insertion site and the duplex region

Base pair	λ_{primer}(°)*	$\lambda_{template}$(°)†	$d_{C1'-C1'}$(Å)	Shear(Å)	Stretch(Å)	Stagger(Å)	Buckle(°)	Propeller(°)	Opening(°)
Insertion site									
Molecule A‡									
C•A cognate	58.7	54.4	10.5	0.29	−0.03	0.04	5.92	−3.23	5.69
C•A wobble	70.7	51.9	10.1	1.49	−0.07	−0.01	3.45	−5.20	15.01
T•A	56.0	57.4	10.4	−0.16	−0.13	0.09	3.21	−6.16	2.97
A•T	57.7	56.7	10.4	0.08	−0.14	−0.10	2.29	−7.01	1.50
G•C	58.0	54.5	10.6	−0.06	−0.14	−0.20	−2.94	−14.66	1.60
C•G§	55.1	55.4	10.6	0.19	−0.15	0.15	4.66	−3.89	2.52
Watson-Crick¶	56.7±1.4	56.0±1.3	10.5±0.11	0.01±0.15	−0.14±0.01	−0.02±0.16	1.81±3.31	−7.93±4.68	2.14±0.71
Molecule B‡									
C•A cognate	59.6	52.4	10.4	0.62	−0.12	−0.15	6.69	−3.31	5.80
C•A wobble	NA								
T•A	55.8	55.7	10.5	−0.02	−0.14	−0.02	7.56	−5.36	1.60
A•T	59.7	55.1	10.5	−0.03	−0.04	−0.04	2.19	−5.56	1.23
G•C	58.7	54.4	10.6	−0.18	−0.14	−0.13	−1.03	−11.96	1.56
C•G§	56.9	53.5	10.6	0.46	−0.09	−0.18	8.51	−4.62	3.03
Watson-Crick¶	57.8±1.8	54.7±0.9	10.6±0.1	0.06±0.28	−0.10±0.05	−0.09±0.08	4.31±4.51	−6.88±3.41	1.86±0.80
Duplex region									
C•A(*n*-1)	64.2	44.4	10.3	1.95	−0.42	−0.27	18.44	−18.84	2.33
Watson-Crick(*n*-1)‖	58.1±2.1	58.0±0.8	10.2±0.2	0.01±0.22	−0.11±0.09	−0.31±0.25	22.6±5.9	−6.89±2.83	4.19±1.76
C•A(*n*-3)	57.8	53.0	10.3	0.77	−0.32	−0.09	14.00	−16.83	1.00
Watson-Crick(*n*-3)‖	57.2±0.8	56.6±0.9	10.4±0.1	0.04±0.16	−0.16±0.04	−0.01±0.17	9.87±5.01	−13.4±2.7	3.11±2.16
C•A(*n*-4)	65.8	43.7	10.6	2.13	−0.37	−0.58	12.54	6.43	1.10
Watson-Crick(*n*-4)‖	56.4±1.1	53.3±1.4	10.7±0.1	0.13±0.23	−0.11±0.04	−0.20±0.21	5.82±2.85	7.46±2.56	−0.07±2.52
C•A(*n*-6)	64.5	45.2	10.4	2.06	−0.40	−0.17	1.08	−17.09	5.66
Watson-Crick(*n*-6)‖	54.5±0.5	53.8±0.8	10.8±0.1	0.09±0.07	0.02±0.16	−0.15±0.11	2.06±4.71	−11.0±2.8	−0.52±0.94
C•A1**	64.0	45.7	10.4	−1.71	−0.45	0.22	11.06	−8.83	−2.06
C•A2**	72.3	49.3	10.2	2.17	−0.27	0.27	−10.28	−12.27	9.05



*λ_{primer} and $\lambda_{template}$ are defined as the angle between the glycosidic bond of primer or template nucleotide and the line draw between the C1′ atoms of the base pair (see inserted panel above). $d_{C1'-C1'}$ is the distance between the C1′ atoms of the base pair. All other base pair parameters are defined (1). All values were calculated in 3DNA (2).

†λ_{primer} and $\lambda_{template}$ are defined as the angle between the glycosidic bond of primer or template nucleotide and the line draw between the C1′ atoms of the base pair (see inserted panel above). $d_{C1'-C1'}$ is the distance between the C1′ atoms of the base pair. All other base pair parameters are defined (1). All values were calculated in 3DNA (2).

‡There are two molecules in the asymmetric unit in Crystal Form II. For C•A cognate and C•A wobble structures, molecule 1 (chains D, E, and F) is more ordered than molecule 2 (chains A, B, and C). Chain naming follows previously published structures (3, 4). Each molecule contains the BF polymerase, DNA primer, and template. In the structure of the ddCTP•dA cognate, the O helix of molecule 2 is in the closed conformation and ddCTP adopts a near-cognate shape. In the structure of the ddCTP•dA wobble, the O helix of molecule 2 is in the open conformation, the loop between O helix and N helix is partially disordered, and there is no base pairing at the insertion site. In the structures of four Watson-Crick base pairs, the structure of molecule 2 is similar to molecule 1 with the exception of some amino acid side chain conformations.

§This structure was determined previously (2HVI) (3).

¶Average values and standard deviations were calculated over all four cognate base pairs observed at the insertion site.

‖Averaged values of cognate base pairs in previously observed complexes (5). At the postinsertion site and *n*-3 position in the DNA duplex-binding region, average values of all four cognate base pairs (1L3S, 1L3T, 1L3U, and 1L5U) are shown. At the *n*-4 position, average values of C•G, A•T, and G•C pairs (1L3S, 1L3T, and 1L5U) are shown. At the *n*-6 position, average values of T•A, A•T, and C•G pairs (1L3S, 1L3T, and 1L3U) are shown.

**Values of the two C•A mismatches observed in a DNA dodecamer structure (1D99) (6).

1 Dickerson RE, et al. (1989) Definitions and nomenclature of nucleic acid structure parameters. *J Mol Biol* 205:787–791.

2 Lu XJ, Olson WK (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 31:5108–5121.

3 Warren JJ, Forsberg LJ, Beese LS (2006) The structural basis for the mutagenicity of O(6)-methyl-guanine lesions. *Proc Natl Acad Sci USA* 103:19701–19706.

4 Wu EY, Beese LS (2011) The structure of a high fidelity DNA polymerase bound to a mismatched nucleotide reveals an "ajar" intermediate conformation in the nucleotide selection mechanism. *J Biol Chem* 286:19758–19767.

5 Johnson SJ, Taylor JS, Beese LS (2003) Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc Natl Acad Sci USA* 100:3895–3900.

6 Hunter WN, Brown T, Anand NN, Kennard O (1986) Structure of an adenine-cytosine base pair in DNA and its implications for mismatch repair. *Nature* 320:552–555.