Supporting Information for

DNA Mismatch Synthesis Complexes Provide Insights into Base Selectivity of a B family DNA Polymerase

Shuangluo Xia, Jimin Wang & William H. Konigsberg*

Table S1. Crystallographic Statistics for Data Collection and Structure Refinement of dNTP/dT-Containing Ternary Complexes of qm RB69pol.

	dATP/dT	dGTP/dT	dCTP/dT	dTTP/dT
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (a,b,c (Å))	75.1,120.2,130.8	75.1,119.5,130.1	75.0,120.1,130.9	75.1,119.8,130.4
Resolution (Å)	50.0 - 2.05	50.0 - 1.90	50.0 - 2.19	50.0 - 1.97
No of unique reflections	74,061	91,587	60,668	83,285
Redundancy	3.6 (3.7)	3.2 (3.0)	3.1 (2.8)	3.9 (3.7)
Completeness (%)	98.5 (99.5)	98.8 (96.7)	98.4 (96.9)	99.0 (98.6)
R_{merge} (%)	10.8 (68.0)	7.9 (82.2)	12.0 (77.7)	10.2 (94.9)
I / σ	10.6 (1.6)	13.9 (1.1)	7.6 (1.0)	11.5 (1.3)
Final model				
Amino acid residues	903	903	903	903
Water molecules	754	545	346	691
Ca^{2+} ions	5	5	5	5
Template nucleotides	18	18	18	18
Primer nucleotides	13	13	13	13
dNTP	1	1	1	1
Refinement Statistics				
Reflections	70,263	86,970	57,254	79,070
R (%)	18.4 (24.1)	18.1 (27.6)	19.5 (26.7)	19.0 (26.6)
R_{free} (%)	23.7 (27.0)	21.3 (30.1)	25.1 (32.0)	23.4 (28.6)
r.m.s.d				
Bond length (Å)	0.006	0.006	0.009	0.007
Bond angles (°)	1.056	1.111	1.167	1.156
PDB code	4FJ5	4FJ7	4FJ8	4FJ9

Footnotes:

a, Statistics for the highest resolution shell are in parenthesis.

b, $R_{merge} = \Sigma_{hkl} \Sigma_j |I_j(hkl) - \langle I(hkl) \rangle| / \Sigma_{hkl} \langle I(hkl) \rangle$, statistics for merging all observations for given reflections.

c, $R = \Sigma_{hkl} |F_{obs}(hkl) - F_{calc}(hkl)| / \Sigma_{hkl} F_{obs}(hkl)$, statistics for crystallographic agreement between the measured and model-calculated amplitudes. R_{free} is the agreement for cross-validation data set.

e, Root mean squares deviations (rmsd) to ideal values.

Table S2. Crystallographic Statistics for Data Collection and Structure Refinement of dNTP/dG-Containing Ternary Complexes of qm RB69pol.

	dATP/dG	dGTP/dG	dCTP/dG	dTTP/dG
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (a,b,c (Å))	75.0,119.9,130.5	74.8, 120.4,130.7	74.9,120.2,130.6	75.0,120.3,130.3
Resolution (Å)	50.0 - 2.12	50.0 - 1.90	50.0 - 2.18	50.0 - 2.00
No of unique reflections	67,112	90,886	61,410	80,229
Redundancy	3.2 (3.2)	3.6 (3.5)	3.5 (3.3)	3.2 (2.6)
Completeness (%)	97.4 (94.8)	95.9 (96.2)	98.6 (97.1)	98.0 (88.6)
R_{merge} (%)	6.8 (29.2)	7.3 (86.5)	8.9 (72.8)	9.0 (63.7)
I / σ	15.3 (4.1)	16.9 (1.1)	12.4 (1.8)	11.9 (1.1)
Final model				
Amino acid residues	903	903	903	903
Water molecules	593	624	364	608
Ca^{2+} ions	5	5	6	5
Template nucleotides	18	18	18	18
Primer nucleotides	13	13	13	13
dNTP	1	1	1	1
Refinement Statistics				
Reflections	63,686	86,052	58,032	75,930
R (%)	18.3 (20.9)	18.7 (28.0)	20.3 (32.7)	18.6 (28.9)
R_{free} (%)	22.5 (25.0)	22.2 (30.7)	25.6 (38.5)	22.5 (35.0)
r.m.s.d				
Bond length (Å)	0.006	0.009	0.009	0.008
Bond angles (°)	1.062	1.124	1.178	1.097
PDB code	4FJX	4FK4	4FK0	4FK2

Footnotes:

a, Statistics for the highest resolution shell are in parenthesis.

b, $R_{merge} = \Sigma_{hkl} \Sigma_j |I_j(hkl) - \langle I(hkl) \rangle| / \Sigma_{hkl} \langle I(hkl) \rangle$, statistics for merging all observations for given reflections.

c, $R = \Sigma_{hkl} |F_{obs}(hkl) - F_{calc}(hkl)| / \Sigma_{hkl} F_{obs}(hkl)$, statistics for crystallographic agreement between the measured and model-calculated amplitudes. R_{free} is the agreement for cross-validation data set.

e, Root mean squares deviations (rmsd) to ideal values.

Table S3. Crystallographic Statistics for Data Collection and Structure Refinement of dNTP/dC-Containing Ternary Complexes of qm RB69pol.

	dATP/dC	dGTP/dC	dCTP/dC	dTTP/dC
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (a,b,c (Å))	75.1,119.9,130.6	75.0, 120.1,130.6	78.3,119.6,130.2	75.2,119.8,130.6
Resolution (Å)	50.0 - 2.02	50.0 – 2.11	50.0 - 2.20	50.0 - 2.00
No of unique reflections	72,855	68,358	59,242	80,794
Redundancy	3.0 (2.6)	3.6 (3.5)	3.0 (2.2)	3.1 (3.1)
Completeness (%)	94.6 (94.6)	98.8 (96.4)	94.8 (80.7)	99.1 (99.4)
R_{merge} (%)	12.4 (71.8)	6.1 (27.9)	13.3 (63.9)	11.0 (68.7)
I / σ	8.3 (1.0)	19.9 (3.8)	8.0 (1.1)	11.1 (1.5)
Final model				
Amino acid residues	903	903	903	903
Water molecules	384	734	164	574
Ca^{2+} ions	5	5	5	5
Template nucleotides	18	18	18	18
Primer nucleotides	13	13	13	13
dNTP	1	1	1	1
Refinement Statistics				
Reflections	68,757	64,834	55,752	76,430
R (%)	22.0 (30.1)	17.5 (19.4)	21.0 (26.3)	18.6 (24.3)
R_{free} (%)	25.9 (34.3)	22.2 (24.3)	26.4 (32.2)	22.5 (26.6)
r.m.s.d				
Bond length (Å)	0.009	0.006	0.011	0.008
Bond angles (°)	1.155	1.067	1.345	1.094
PDB code	4FJG	4FJH	4FJI	4FJJ

Footnotes:

a, Statistics for the highest resolution shell are in parenthesis.

b, $R_{merge} = \Sigma_{hkl} \Sigma_j |I_j(hkl) - \langle I(hkl) \rangle| / \Sigma_{hkl} \langle I(hkl) \rangle$, statistics for merging all observations for given reflections.

c, $R = \Sigma_{hkl} |F_{obs}(hkl) - F_{calc}(hkl)| / \Sigma_{hkl} F_{obs}(hkl)$, statistics for crystallographic agreement between the measured and model-calculated amplitudes. R_{free} is the agreement for cross-validation data set.

e, Root mean squares deviations (rmsd) to ideal values.

Table S4. Crystallographic Statistics for Data Collection and Structure Refinement of dNTP/dA-Containing Ternary Complexes of RB69pol quadruple mutant

	dATP/dA	dGTP/dA	dCTP/dA	dTTP/dA
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (a,b,c (Å))	74.7,120.2,130.6	74.9,120.5,130.8	75.2,119.7,130.3	75.0,120.5,130.8
Resolution (Å)	50.0 - 2.00	50.0 - 1.87	50.0 - 2.02	50.0 - 2.00
No of unique reflections	79,818	97,730	76,737	81,721
Redundancy	3.2 (3.2)	3.5 (3.4)	3.1 (3.0)	3.6 (3.6)
Completeness (%)	99.5 (100.0)	99.2 (96.1)	98.5 (98.6)	99.6 (99.9)
R_{merge} (%)	8.2 (81.7)	6.2 (72.0)	9.5 (74.7)	7.6 (94.8)
I / σ	11.6 (1.1)	17.0 (1.3)	12.2 (1.1)	14.8 (1.1)
Final model				
Amino acid residues	903	903	903	903
Water molecules	616	814	464	411
Ca^{2+} ions	5	5	5	5
Template nucleotides	18	18	18	18
Primer nucleotides	13	13	13	13
dNTP	1	1	1	1
Refinement Statistics				
Reflections	75,753	92,786	72,632	77,371
R (%)	18.2 (24.8)	17.9 (28.2)	19.1 (26.6)	19.9 (28.9)
R_{free} (%)	22.5 (30.2)	22.2 (33.6)	23.4 (31.7)	24.1 (31.2)
r.m.s.d				
Bond length (Å)	0.008	0.007	0.008	0.009
Bond angles (°)	1.247	1.180	1.130	1.133
PDB code	4FJK	4FJL	4FJM	4FJN

Footnotes:

a, Statistics for the highest resolution shell are in parenthesis.

b, $R_{merge} = \Sigma_{hkl} \Sigma_j |I_j(hkl) - \langle I(hkl) \rangle| / \Sigma_{hkl} \langle I(hkl) \rangle$, statistics for merging all observations for given reflections.

c, $R = \Sigma_{hkl} |F_{obs}(hkl) - F_{calc}(hkl)| / \Sigma_{hkl} F_{obs}(hkl)$, statistics for crystallographic agreement between the measured and model-calculated amplitudes. R_{free} is the agreement for cross-validation data set.

e, Root mean squares deviations (rmsd) to ideal values.

Table S5. B factors for incoming dNTP, the nucleotides at position n and n-1 of template strand in wt and qm RB69 pol structures.

Base Pairs	dNTP/dN	dNTP	dN (n)	dN (n+1)	Overall
Correct BP (qm)	dATP/dT	10.68	15.68	43.65	24.82
	dTTP/dA	30.64	14.27	54.93	28.75
	dGTP/dC	14.03	20.30	35.80	23.09
	dCTP/dG	15.73	16.18	45.95	28.29
Pyr-Pyr (qm)	dCTP/dC	17.25	33.16	66.94	23.95
	dCTP/dT	11.85	16.62	32.07	25.60
	dTTP/dC	19.17	16.30	26.47	22.27
	dTTP/dT	23.31	20.39	31.51	26.69
Pur-Pyr (qm)	dATP/dC	11.90	18.20	27.98	19.96
	dCTP/dA	15.30	18.59	31.71	25.58
	dGTP/dT	11.45	15.90	30.22	24.27
	dTTP/dG	20.00	9.64	31.23	23.21
Pur-Pur (qm)	dATP/dA	12.22	64.75	73.69	25.14
	dATP/dG	18.01	69.88	51.11	31.94
	dGTP/dA	12.10	48.05	67.47	25.68
	dGTP/dG	14.35	55.04	50.52	25.36
Correct BP (wt)	dCTP/dG	17.54	15.36	29.53	26.13

Figure S1.

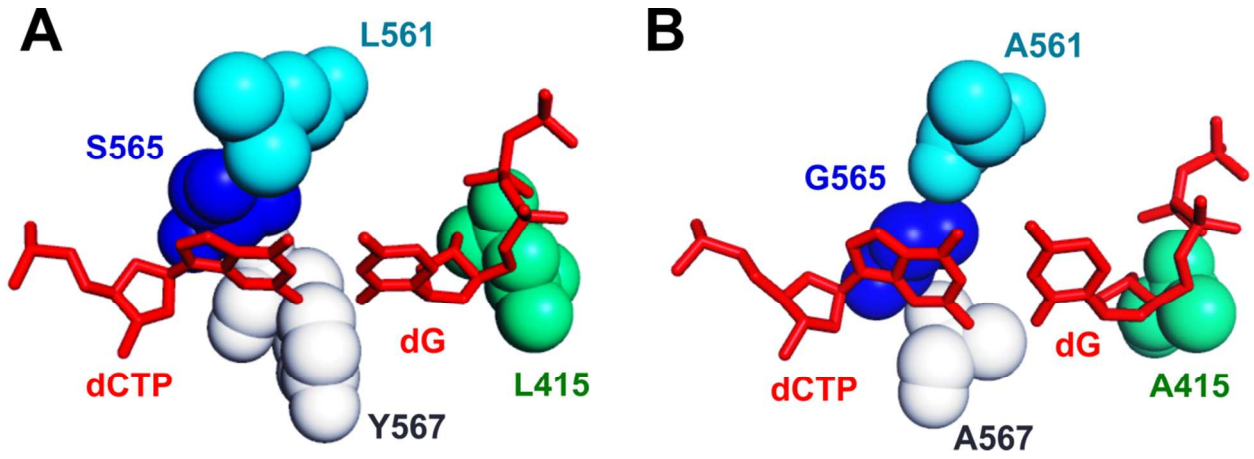

Figure S2.

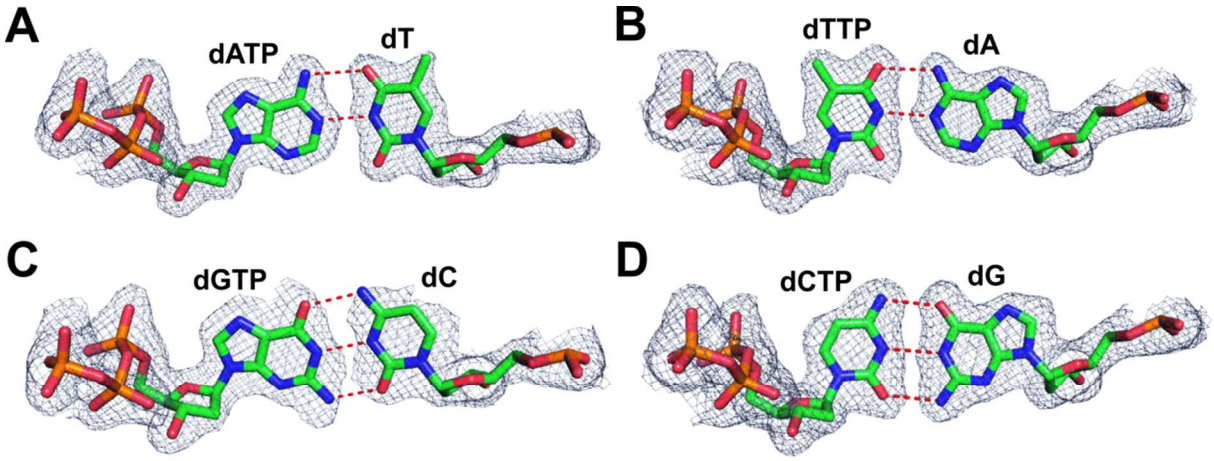

Figure S3.

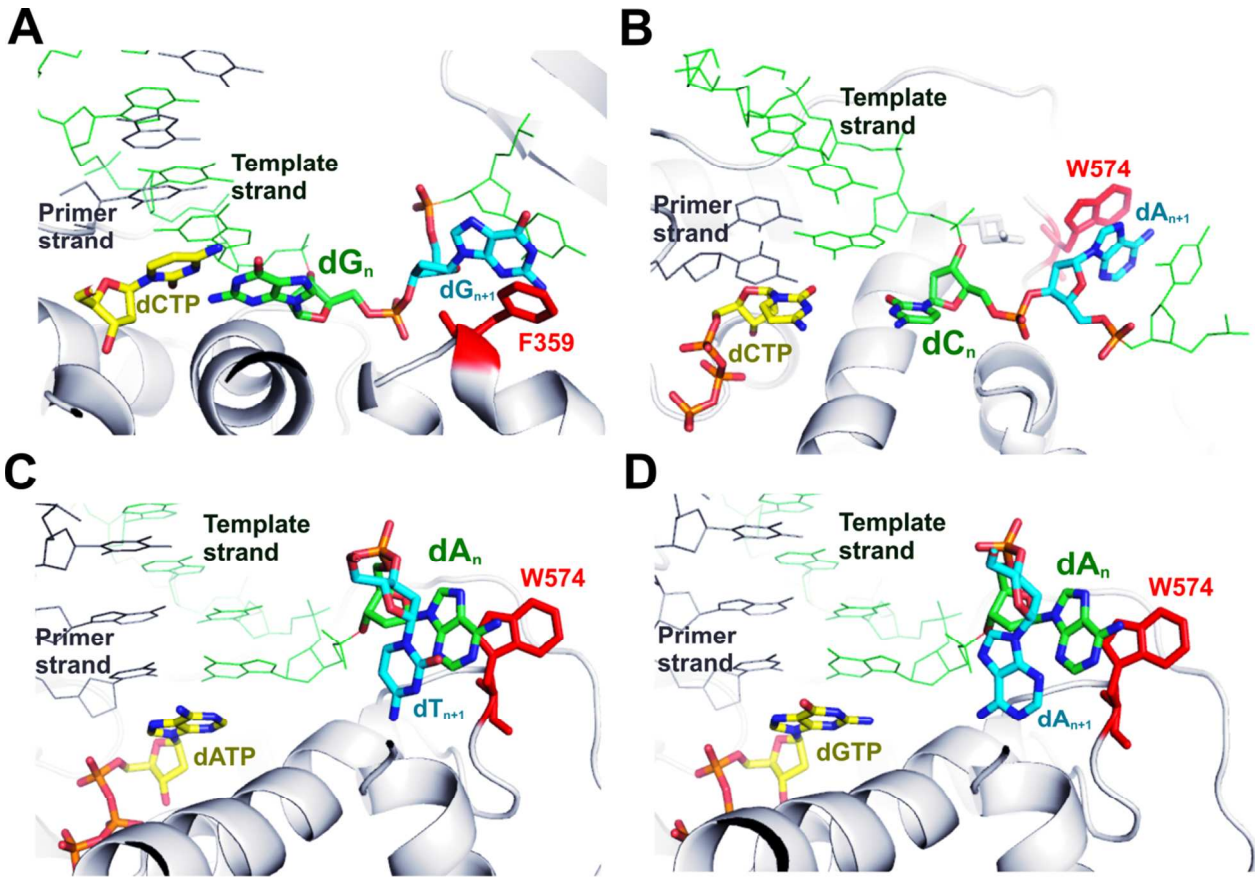
A
Template strand
Primer strand
dCTP
dGn
dGn+1
F359
B
Template strand
W574
dAn+1
Primer strand
dCTP
dCn
C
Template strand
dAn
W574
Primer strand
dATP
dTn+1
D
Template strand
dAn
W574
Primer strand
dGTP
dAn+1

Figure S4.

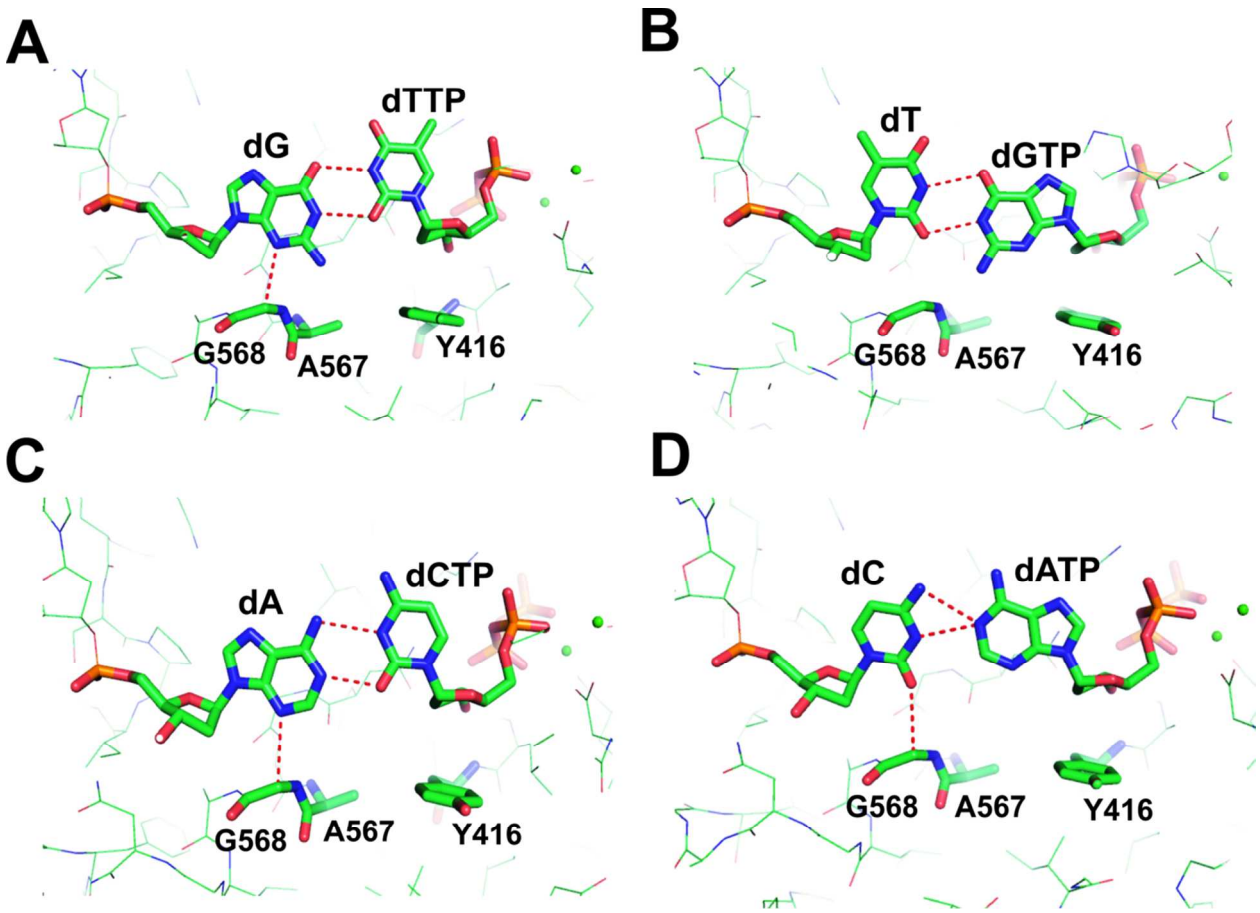

Figure S5.

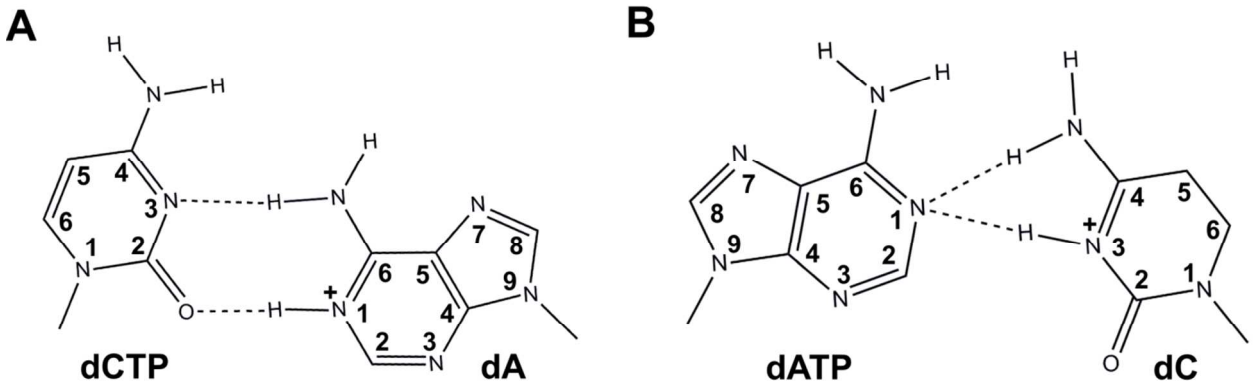

Figure S6.

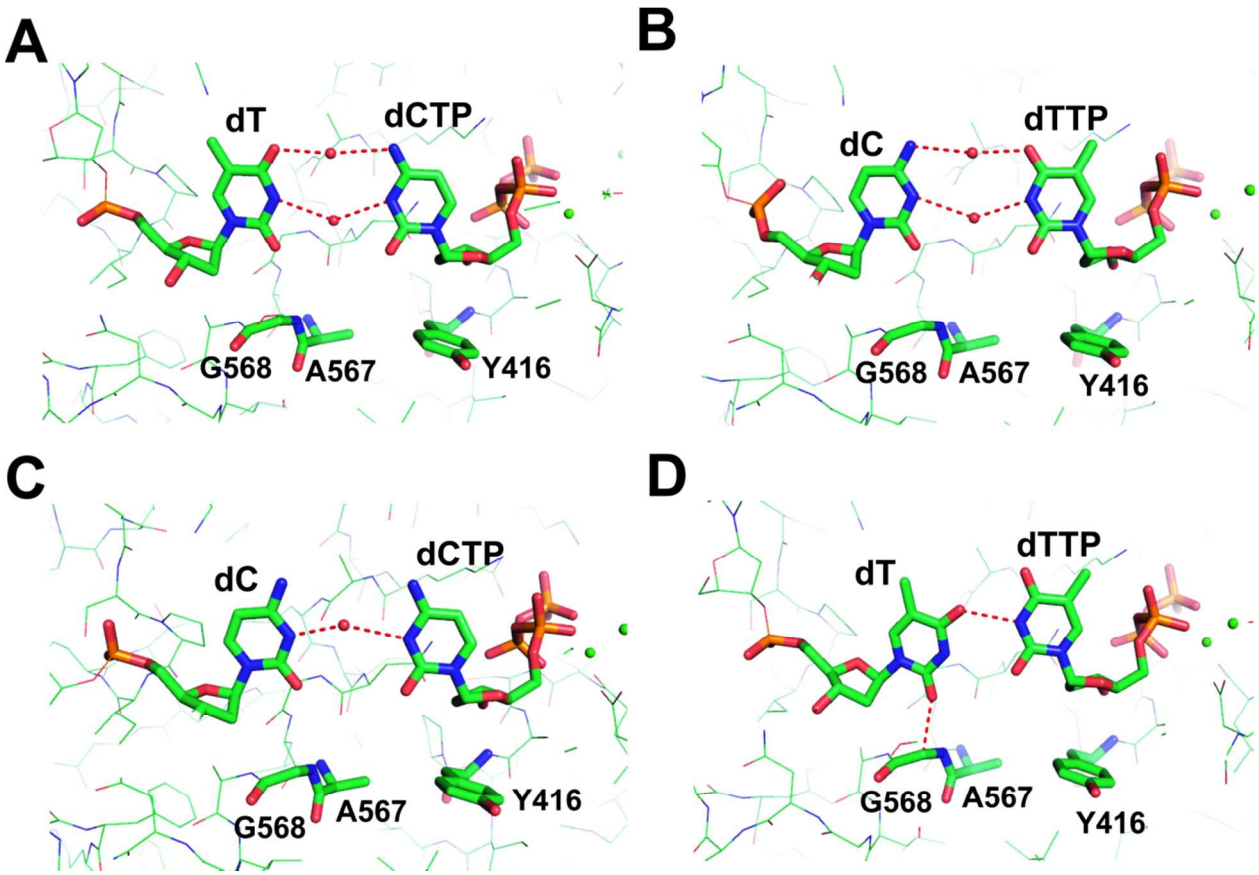
A
dT
dCTP
G568
A567
Y416
B
dC
dTTP
G568
A567
Y416
C
dC
dCTP
G568
A567
Y416
D
dT
dTTP
G568
A567
Y416

Figure S7.

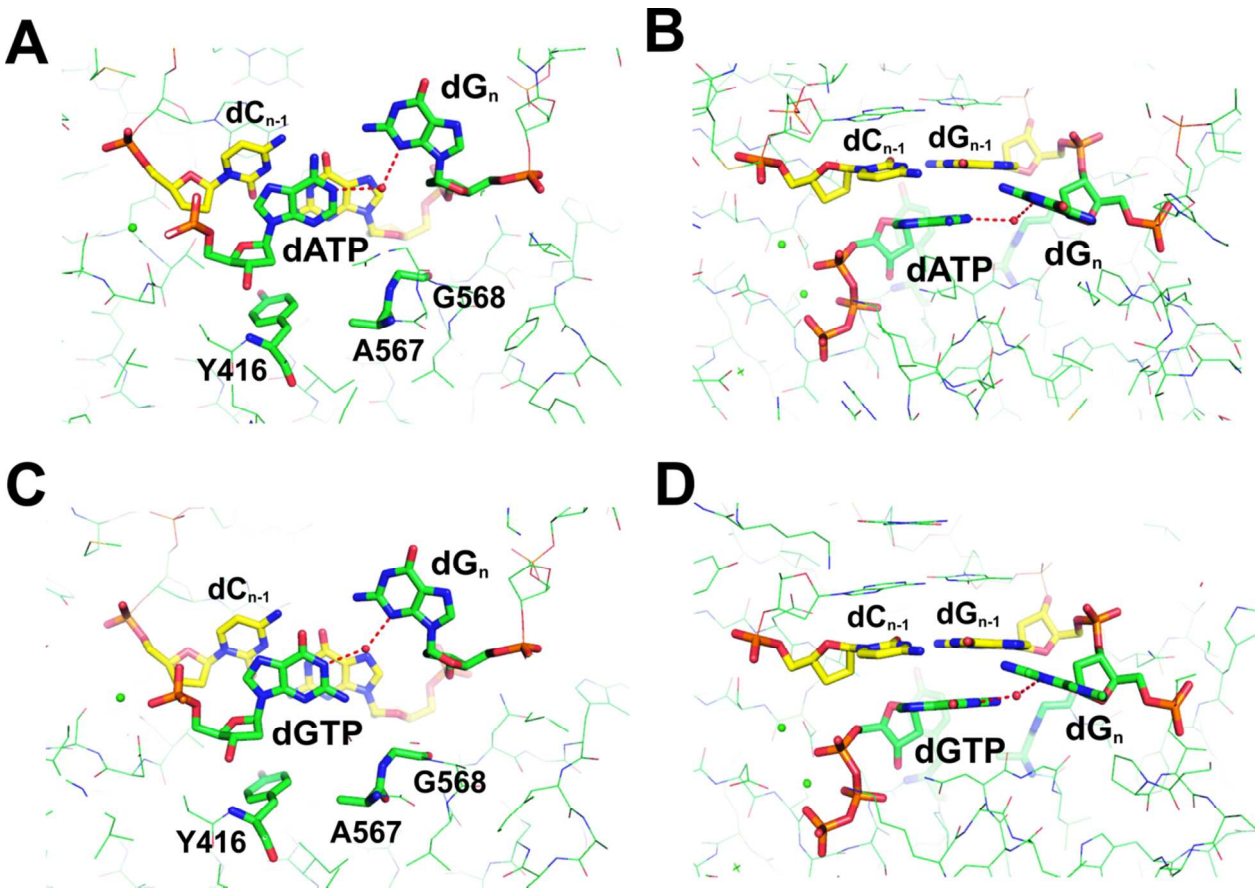
A
dC$_{n-1}$
dG$_n$
dATP
G568
A567
Y416
B
dC$_{n-1}$
dG$_{n-1}$
dATP
dG$_n$
C
dC$_{n-1}$
dG$_n$
dGTP
G568
A567
Y416
D
dC$_{n-1}$
dG$_{n-1}$
dGTP
dG$_n$

Figure S8.

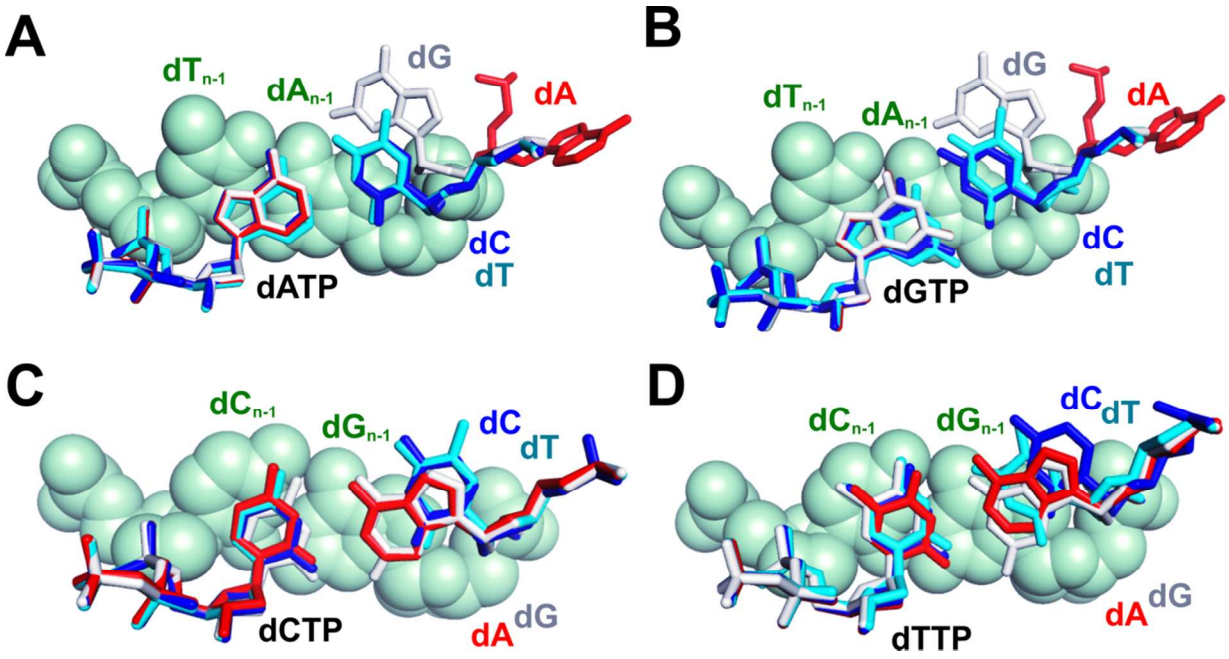
A
dT$_{n-1}$
dA$_{n-1}$
dG
dA
dC
dT
dATP
B
dT$_{n-1}$
dA$_{n-1}$
dG
dA
dC
dT
dGTP
C
dC$_{n-1}$
dG$_{n-1}$
dC
dT
dCTP
dA
dG
D
dC$_{n-1}$
dG$_{n-1}$
dC
dT
dA
dG
dTTP

Figure S9.

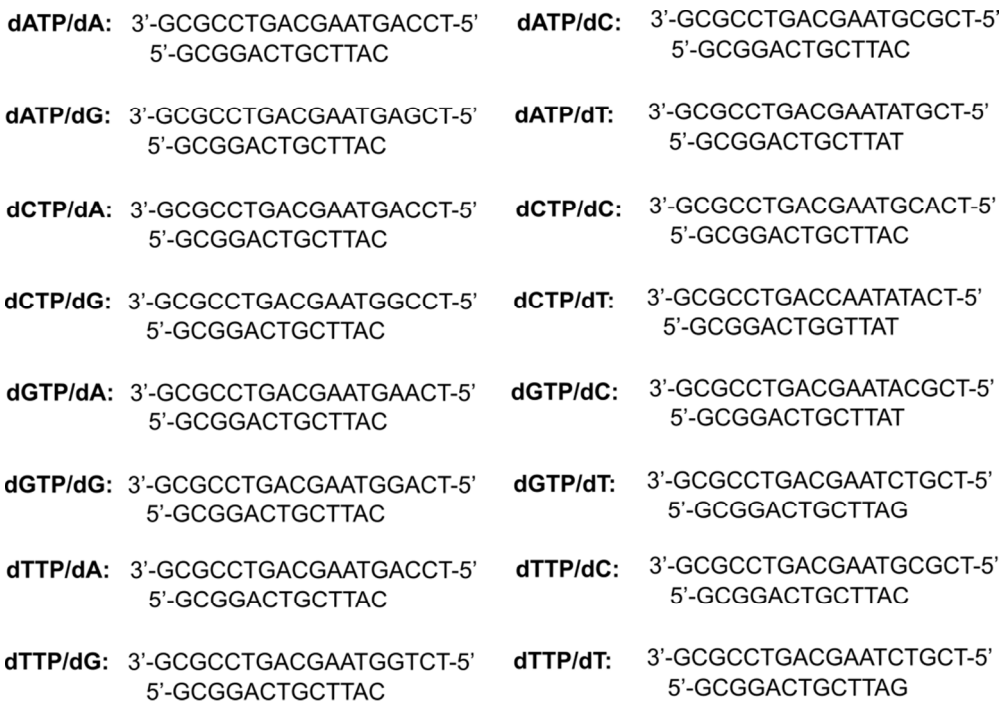
dATP/dA: 3'-GCGCCTGACGAATGACCT-5' 5'-GCGGACTGCTTAC
dATP/dC: 3'-GCGCCTGACGAATGCGCT-5' 5'-GCGGACTGCTTAC
dATP/dG: 3'-GCGCCTGACGAATGAGCT-5' 5'-GCGGACTGCTTAC
dATP/dT: 3'-GCGCCTGACGAATATGCT-5' 5'-GCGGACTGCTTAT
dCTP/dA: 3'-GCGCCTGACGAATGACCT-5' 5'-GCGGACTGCTTAC
dCTP/dC: 3'-GCGCCTGACGAATGCACT-5' 5'-GCGGACTGCTTAC
dCTP/dG: 3'-GCGCCTGACGAATGGCCT-5' 5'-GCGGACTGCTTAC
dCTP/dT: 3'-GCGCCTGACCAATATACT-5' 5'-GCGGACTGGTTAT
dGTP/dA: 3'-GCGCCTGACGAATGAACT-5' 5'-GCGGACTGCTTAC
dGTP/dC: 3'-GCGCCTGACGAATACGCT-5' 5'-GCGGACTGCTTAT
dGTP/dG: 3'-GCGCCTGACGAATGGACT-5' 5'-GCGGACTGCTTAC
dGTP/dT: 3'-GCGCCTGACGAATCTGCT-5' 5'-GCGGACTGCTTAG
dTTP/dA: 3'-GCGCCTGACGAATGACCT-5' 5'-GCGGACTGCTTAC
dTTP/dC: 3'-GCGCCTGACGAATGCGCT-5' 5'-GCGGACTGCTTAC
dTTP/dG: 3'-GCGCCTGACGAATGGTCT-5' 5'-GCGGACTGCTTAC
dTTP/dT: 3'-GCGCCTGACGAATCTGCT-5' 5'-GCGGACTGCTTAG

Figure S10.


dATP/dA: 3'-GCGCCTGACGAATGACCT-5' 5'-GCGGACTGCTTAC
dATP/dC: 3'-GCGCCTGACGAATGCGCT-5' 5'-GCGGACTGCTTAC
dATP/dG: 3'-GCGCCTGACGAATGGACT-5' 5'-GCGGACTGCTTAC
dATP/dT: 3'-GCGCCTGACGAATATGCT-5' 5'-GCGGACTGCTTAT
dCTP/dA: 3'-GCGCCTGACGAATGAGCT-5' 5'-GCGGACTGCTTAC
dCTP/dC: 3'-GCGCCTGACGAATGCACT-5' 5'-GCGGACTGCTTAC
dCTP/dG: 3'-GCGCCTGACGAATGGGCT-5' 5'-GCGGACTGCTTAC
dCTP/dT: 3'-GCGCCTGACTAATGTGCT-5' 5'-GCGGACTGGTTAC
dGTP/dA: 3'-GCGCCTGACGAATGAACT-5' 5'-GCGGACTGCTTAC
dGTP/dC: 3'-GCGCCTGACGAATACGCT-5' 5'-GCGGACTGCTTAT
dGTP/dG: 3'-GCGCCTGACGAATGGACT-5' 5'-GCGGACTGCTTAC
dGTP/dT: 3'-GCGCCTGACGAATCTGCT-5' 5'-GCGGACTGCTTAG
dTTP/dA: 3'-GCGCCTGACGAATGAACT-5' 5'-GCGGACTGCTTAC
dTTP/dC: 3'-GCGCCTGACGAATGCGCT-5' 5'-GCGGACTGCTTAC
dTTP/dG: 3'-GCGCCTGACGAATGGACT-5' 5'-GCGGACTGCTTAC
dTTP/dT: 3'-GCGCCTGACGAATCTGCT-5' 5'-GCGGACTGCTTAG

Figure legends

Figure S1. NBP of (A) wt and (B) qm RB69pol. dCTP/dG is shown in red stick and protein side chains are shown in space-filling mode. L415/A415 is shown in green, L561/A561 is shown in cyan, S565/G565 is shown in blue and Y567/A567 is shown in white.

Figure S2. Final 2Fo-Fc electron density map for four correct base-pairs at the NBP of qm RB69pol contoured at 1.2 σ: (A) dATP/dT; (B) dTTP/dA; (C) dGTP/dC and (D) dC/dGTP.

Figure S3. Alternative conformations of the 5'-overhang of template strand DNA in the ternary complex of (A) dCTP/dG-containing qm RB69pol; (B) dCTP/dC-containing qm RB69pol; (C) dATP/dA-containing qm RB69pol and (D) dGTP/dA-containing qm RB69pol.

Figure S4. NBP of qm RB69pol with Pu-Py mismatches. (A) dTTP/dG; (B) dGTP/dT; (C) dCTP/dA and (D) dATP/dC.

Figure S5. Protonation state of (A) dCTP/dA and (B) dATP/dC observed in the structures.

Figure S6. NBP of qm RB69pol with Py-Py mismatches. (A) dCTP/dT; (B) dTTP/dC; (C) dCTP/dC and (D) dTTP/dT.

Figure S7. NBP of qm RB69pol with Pu-Pu mismatches. (A) dATP/dG; (B) orthogonal view of A; (C) dGTP/dG and (D) orthogonal view of C.

Figure S8. Base stacking of the incoming dNTP and templating base with the penultimate base-pair. (A) dATP/dN; (B) dGTP/dN; (C) dCTP/dN and (D) dTTP/dN.

Figure S9. P/T sequences used for chemical quench experiments.

Figure S10. P/T sequences used for crystallization (the primer is dideoxy-terminated).