Supplementary Materials

Transcriptional Bypass of DNA-protein and DNA-peptide Conjugates by T7 RNA Polymerase

Shaofei Ji[†], Jenna Thomforde[§], Colette Rogers[‡], Iwen Fu[£], Suse Broyde[£], and Natalia Y. Tretyakova[§,€,*]

[†]Department of Chemistry; [‡]Department of Biochemistry, Molecular Biology and Biophysics; [§]Department of Medicinal Chemistry; [€]Masonic Cancer Center
University of Minnesota, Minneapolis, MN, 55455, United States

[£]Department of Biology
New York University, New York, NY 10003, USA

*Corresponding author:

Masonic Cancer Center, University of Minnesota, 2231 6th Street SE, 2-147 CCRB, Minneapolis, MN 55455, USA; Tel: 612-626-3432; Fax: 612-624-3869; e-mail: trety001@umn.edu

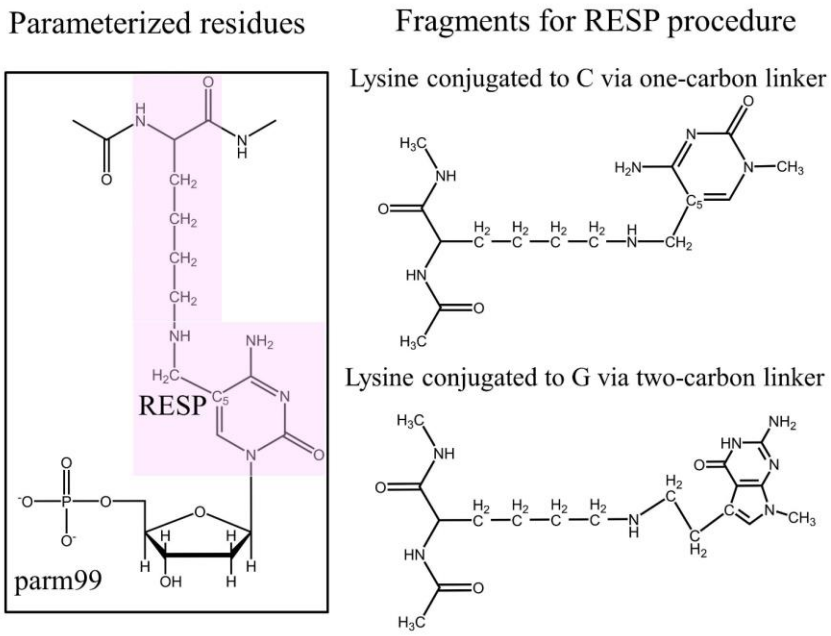
Contents

Supplementary Materials and Methods

MD parameters for non-standard residues

Parameters, except for partial charges and atom types for the DNA-protein cross-links at the lesion site and the incoming NTP parameters were assigned according to GAFF[4] and the AMBERff14SB[5, 6] force field, using the antechamber module of AMBER14.[7]

DNA-Peptide cross-links Partial charges for the lysine linked to cytosine via the one-carbon linker were computed based the fragment shown below. The N- and C- termini of lysine were capped with ACE and NME residues, respectively. The charges for these capping groups were assigned in the AMBER FF9X force fields. For the cytosine base, the N1 atom was capped with a CH3 group. The whole fragment was geometry optimized at the B3LYP/6-31G(d) level of theory using Gaussian09.[8] Subsequently, the partial atomic charges were determined using the restrained electrostatic potential[9] (RESP) fit procedure at the HF/6-31G(d) level of theory. Then the non-standard base/side chains were merged with the corresponding standard parm99 nucleotide/amino acid residues to obtain topologies for the non-standard nucleotides/residues. The capping methyl group on the N1 atom was replaced by the C1' carbon atom of deoxyribose and the residual fractional charge was added to this C1' carbon. The partial charges and atom types are given in Table S5. A similar procedure was applied to the DpC adduct, where the lysine side chain was conjugated to the C7 position of 7-deaza-dG via a two-carbon linker. The partial charges and atom types are given in Table S6.



Chemical structures of DNA-protein cross-links and the fragments for RESP charge calculations.

RNA incoming nucleotide (NTP) The GTP/ATP parameters were taken from Meagher *et al*.[10]; each NTP has a total charge of −4. To develop CTP and N1-protonated ATP parameters, we compared the parameters of the monophosphate in the AMBERff14SB force field with those of the triphosphate in GTP/ATP from Meagher *et al*.[10] The Amber99sb force field parameters for cytosine/adenine were employed for the base and sugar ring of CTP and ATP. We first modeled CTP based on PDB 1AZS[11] by replacing the guanine base with cytosine. The partial charges for the cytosine base were obtained in a similar fashion as for the original nucleic acid parameters.[12] The N1 atom of the cytosine base was capped with a methyl group and the partial atomic charges of this capped base were determined using the restrained electrostatic potential (RESP)[9] fit procedure at the HF/6-31G(d) level of theory. The capping methyl group on the N1 atom was replaced by the C1' carbon atom of ribose and the residual fractional charge was added to this C1' carbon. The obtained RESP charges for the base were merged with the corresponding ones for ATP/GTP developed by Meagher *et al*.[10] to obtain the charges for the whole CTP, resulting in a total charge of −4. The partial charges and atom types are given in Table S7.

For the case of N1-protonated ATP, we capped an N1-protonated adenine base at the N9 atom with a methyl group. Then the atomic partial charges were determined using the RESP[9] formalism from a wavefunction computed at the HF/6-31G(d) level. These RESP charges for the protonated adenine base were merged with the corresponding ones for ATP developed by Meagher *et al*.[10] to obtain the topology for the complete ATP+, resulting in a total charge of −3. The partial charges and atom types are given in Table S8.

Molecular dynamics simulation protocols

Construction of molecular topology and coordinate files for the initial models was performed using the *tleap* module of AmberTools14[7]. The polymerases were explicitly solvated with the TIP3P[13] water model with at least a 10 Å buffer. To neutralize the system, we used Na^+ ions. Then we added approximately 100 Na^+ and Cl^- ions to bring the salt concentration close to the physiological value of ~ 0.15M.

All systems were subjected to energy minimization, equilibration, and production dynamics using the PMEMD module of AMBER16[14]. All simulations were carried out according to the following simulation protocol: first, the counterions and water molecules were minimized for 2500 steps of steepest descent and 2500 steps of conjugate gradient energy minimization, with a force constant of 50 kcal/mol/Å^2 restraint on the solute atoms. Then, 30 ps initial MD at 10 K with 25 kcal/mol/Å^2 restraints on solute were performed to allow the solute to relax. Next, the system was heated from 10 K to 300 K at constant volume for 30 ps with 10 kcal/mol restraints on the solute. Restraints on the solute were then relaxed with 30 ps of 10 kcal/(molÅ^2), 40 ps of 1 kcal/(molÅ^2), 50 ps of 0.1 kcal/(molÅ^2), and 100 ps of 0.05 kcal/(molÅ^2) restraints. Subsequently, unrestrained dynamics was propagated in the NPT ensemble with a 2 fs timestep. Production MD was conducted at 1 atmosphere, 300 K. Constant pressure was maintained with a weak-coupling (Berendsen[15]) barostat with a time constant of 1 ps. The simulation temperature was regulated by a Berendsen thermostat with a coupled thermostat of 4 ps time constant. In all MD simulations, the SHAKE[16] algorithm for constraining the length of bonds to hydrogen was used. The short-range cutoff for nonbonded interactions was 9.0 Å, and long-range electrostatic interactions were treated with the particle-mesh Ewald method.[17] The simulations were run for ~ 1 μs and the

trajectories were saved every 10 ps for further analysis. The simulations were run initially for equilibration using the CPU version of the PMEMD.MPI implementation of SANDER from AMBER16,[14] followed by production runs using the GPU version of the PMEMD.CUDA implementation of SANDER in AMBER16[14] on NVIDIA Tesla K80 cards.[18]

Structural stability

In each simulation, the polymerase enzyme reached a stable state after ~ 200 ns MD simulation (Figures S4A) and the active site remained stable after ~ 250 ns (Figures S4B). Therefore, structural analyses were obtained from the MD frames with the first 300 ns discarded, resulting in a 700 ns production run for ensemble averaging. Within these ensembles, each domain has an average Ca RMSD of about 1−2 Å, indicating a stable conformation throughout the simulations (Table S1).

Structural clustering

Post-processing of all simulations was carried out using the CPPTRAJ module[19] of AMBER14[7]. The best representative structure was obtained from the last 700 ns simulation range using cluster analysis, which was performed using the average linkage hierarchical agglomerative method[20] and RMSD as the distance metric. We wished to study the structural and dynamic properties of the active site. For this purpose, we used the heavy atoms of the templating base and incoming NTP, as well as the neighboring two base pairs in the duplex region and the two Mg2+ ions for clustering.

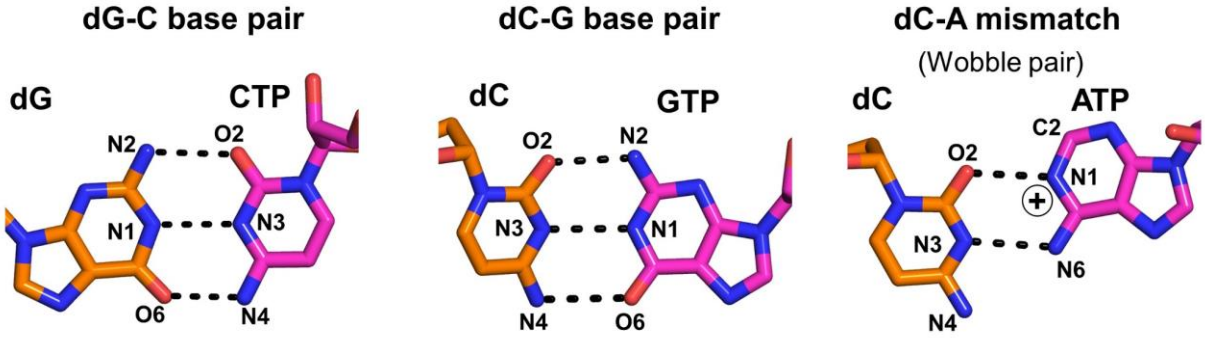
Major groove width

The shortest distance between P atoms across the major groove in the duplex is used to define major groove width. The cross-strand distance from Pi on strand I to Pi+m on strand II is measured. In the case of T7 RNA polymerase, the major groove width corresponds to a minimum when m = 6 for all the cases except dC-ATP where the minimum occurs at m = 5. Due to the short length of the duplex, only three pairs of P-P distance measurement for the major groove width are possible. In the case of DNA Pol η, the major groove width corresponds to a minimum near m = 5 for all the cases, except dC-ATP where the minimum occurs at m = 4.

Supplementary Tables

Table S1. Ensemble average values for hydrogen bonding interactions of the active site base-pair. The occupancies (%), average distances (Å) and angles (º) of hydrogen bonds are listed for occupancies > 20 %. A hydrogen bond is formed if the donor-acceptor distance is < 3.5 Å and the angle of donor-H-acceptor is > 140º. The hydrogen bonding schemes are shown in the Figure below.

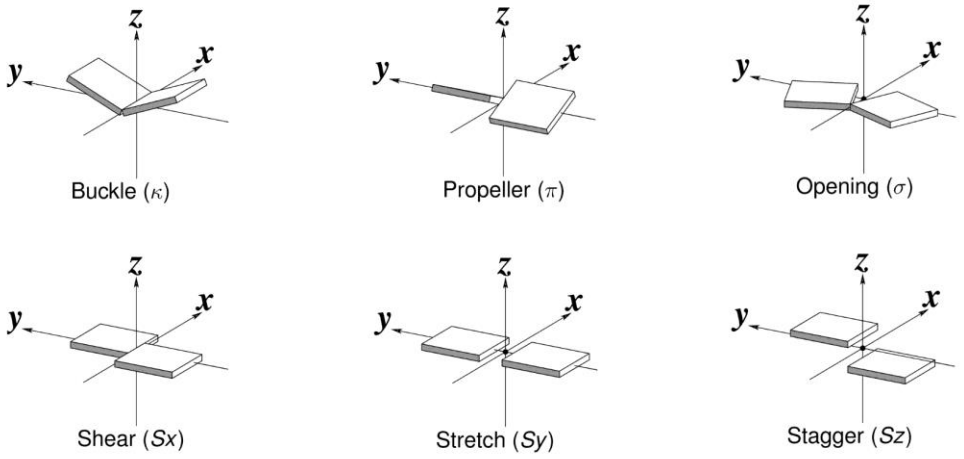
Base-Pair	Hydrogen Bond	Unmodified DNA Base	Modified DNA Base
dG-CTP	dG:N2−H···O2:CTP	98 %, 2.95 Å, 164º	96%, 2.85 Å, 163º
	dG:N1−H···N3:CTP	97 %, 2.95 Å, 162º	98%, 2.99 Å, 164º
	dG:O6···H−N4:CTP	97 %, 2.86 Å, 164º	57%, 2.97 Å, 157º
dC-GTP	dC:O2···H−N2:GTP	99%, 2.80 Å, 163º	98%, 2.92 Å, 163º
	dC:N3···H−N1:GTP	99%, 2.94 Å, 165º	97%, 3.01 Å, 161º
	dC:N4−H···O6:GTP	77%, 2.96 Å, 157º	90%, 2.87 Å, 161º
dC-ATP+	dC:O2···H−N1:ATP	98%, 2.78 Å, 163º	20%, 3.01 Å, 149º
	dC:N3···H−N1:ATP		95%, 2.97 Å, 162º
	dC:N3···H−N6:ATP	94%, 2.91 Å, 160º	
	dC:N4−H···N6:ATP		65%, 3.09 Å, 153º



Hydrogen bonding scheme of C-G matched and C-A+ mismatched base-pairs for the unmodified cases.

Table S2. Ensemble averages and standard deviations of the base-pair parameters[21] at the active site. The parameters Buckle, Propeller and Opening (º), Shear, Stretch, and Stagger (Å), are illustrated below.

Models T7RNAP	**Buckle**	**Propeller**	**Opening**	**Shear**	**Stretch**	**Stagger**
dG-CTP	-9.9 ± 7.9	0.8 ± 9.7	-2.3 ± 3.1	-0.02 ± 0.34	-0.05 ± 0.12	0.22 ± 0.46
dC-GTP	-22.5 ± 6.5	-3.9 ± 5.9	3.0 ± 2.8	0.01 ± 0.28	0.08 ± 0.11	0.36 ± 0.35
dC-ATP+	-26.9 ± 8.8	2.6 ± 7.8	6.8 ± 4.2	2.20 ± 0.43	-0.21 ± 0.18	-0.35 ± 0.44
10mer-dG-CTP	-24.8 ± 7.7	13.2 ± 10.3	1.5 ± 4.0	-0.16 ± 0.50	-0.04 ± 0.14	-0.88 ± 0.48
11mer-dC-GTP	-24.9 ± 7.3	5.8 ± 6.7	-0.4 ± 3.1	0.35 ± 0.26	0.07 ± 0.12	0.40 ± 0.36
11mer-dC-ATP+	-14.5 ± 9.2	6.5 ± 8.2	5.1 ± 4.1	0.73 ± 0.46	-0.15 ± 0.16	-0.74 ± 0.44

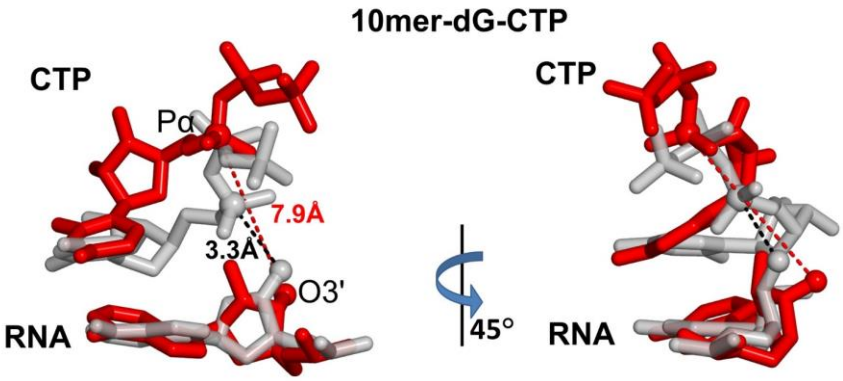


Imaged created by 3DNA[21] illustrating the positive values of designated parameters. Note that ideal B-DNA adopts values near 0 for all parameters.

Table S3. Ensemble averages and standard deviations of the distance between the O3' of the primer terminus and the Pα of the NTP. The O3'−Pα distance is illustrated below.

Models	10mer-dG-CTP	11mer-dC-GTP	11mer-dC-ATP	dG-CTP	dC-GTP	dC-ATP
O3'−Pα Distance (Å)	**7.0 ± 0.6**	4.9 ± 0.3	3.5 ± 0.1	5.1 ± 0.3	3.4 ± 0.2	3.6 ± 0.1

Note that the O3'−Pα distance in the 10mer-dG-CTP system is the largest and most dynamic, which is reflected in the largest standard deviation.



Superimposition of final (red) and initial (gray) structures in the 10mer-dG-CTP MD trajectories shows great enlargement of the O3'−Pα distance. Due to the presence of the peptide, the incoming CTP is repositioned so that it is lifted away from the RNA terminus and propeller-twisted (~ 13°; see Table S3).

Table S4. Ca RMSD (Å) relative to the average structure for each domain[a] of the T7 RNA polymerase systems over the last 700 ns MD simulation**.**

Models	Enzyme	N-terminus	Thumb	Fingers	Palm
dG-CTP	1.9±0.4	2.1±0.5	1.2±0.3	2.0±0.6	0.6±0.1
dC-GTP	1.6±0.2	1.7±0.3	1.7±0.4	1.5±0.3	0.8±0.2
dC-ATP	1.7±0.4	2.3±0.6	0.8±0.4	1.2±0.3	0.7±0.1
10mer-dG-CTP	1.7±0.3	2.0±0.5	1.2±0.3	1.3±0.2	1.0±0.2
11mer-dC-GTP	1.9±0.3	1.7±0.3	1.8±0.5	2.4±0.5	0.8±0.1
11mer-dC-ATP	1.7±0.3	1.9±0.5	1.4±0.2	1.5±0.3	0.8±0.1

[a]Domain residues in T7 RNA polymerase: N−terminus, 1−324; thumb, 325−411; fingers, 566−784; palm, 412−565, 785−883.

The standard deviations of the RMSDs are also given.

Table S5. AMBER atom name, atom type and partial charge assignments for templating dC with C5 atom conjugated to side chain of lysine via a one-carbon linker.

Atom name	Atom type	Partial charge	Atom name	Atom type	Partial charge
P	P	1.165900	N	N	-0.347900
OP1	O2	-0.776100	H	H	0.274700
OP2	O2	-0.776100	CA	CX	-0.399596
O5′	OS	-0.495400	HA	H1	0.212599
C5′	CI	-0.006900	CB	C8	0.377877
H5′	H1	0.075400	HB2	HC	-0.042061
H5′′	H1	0.075400	HB3	HC	-0.042061
C4′	CT	0.162900	CG	C8	-0.248471
H4′	H1	0.117600	HG2	HC	0.060894
O4′	OS	-0.369100	HG3	HC	0.060894
C1′	CT	-0.011600	CD	C8	0.024343
H1′	H2	0.196300	HD2	HC	0.024951
N1	N*	-0.010350	HD3	HC	0.024951
C6	CM	-0.019150	CE	C8	0.106891
H6	H4	0.239286	HE2	HP	0.075354
C5	CM	-0.368312	HE3	HP	0.075354
C4	CA	0.867372	NZ	N3	-0.639520
N4	N2	-1.089384	HZ2	H	0.427093
H41	H	0.483155	HZ3	H	0.427093
H42	H	0.483155	CM	CT	0.100380
N3	NC	-0.762594	HM1	HP	0.101702
C2	C	0.768523	HM2	HP	0.101702
O2	O	-0.556668	C	C	0.734100
C3′	CT	0.071300	O	O	-0.589400
H3′	H1	0.098500			
C2′	CT	-0.085400			
H2′	HC	0.071800			
H2′′	HC	0.071800			
O3′	OS	-0.523200			

Table S6. AMBER atom name, atom type and partial charge assignments for templating dG with C7 atom conjugated to side chain of lysine via a two-carbon linker.

Atom Name	Atom Type	Partial Charge	Atom Name	Atom Type	Partial Charge
P	P	1.165900	CN	CM	-0.165887
OP1	O2	-0.776100	HN1	HA	0.141420
OP2	O2	-0.776100	HN2	HA	0.141420
O5′	OS	-0.495400	CM	CT	-0.246083
C5′	CT	-0.006900	HM1	HP	0.188887
H5′	H1	0.075400	HM2	HP	0.188887
H5′′	H1	0.075400	NZ	N3	0.094276
C4′	CT	0.162900	HZ1	H	0.185539
H4′	H1	0.117600	HZ2	H	0.185539
C3′	CT	0.071300	CE	CT	-0.035943
H3′	H1	0.098500	HE2	HP	0.094562
C2′	CT	-0.085400	HE3	HP	0.094562
H2′	HC	0.071800	CD	CT	-0.119875
H2′′	HC	0.071800	HD2	HC	0.028534
O3′	O	-0.523200	HD3	HC	0.028534
O4′	OS	-0.369100	CG	CT	0.242795
C1′	CT	0.035800	HG2	HC	-0.027703
H1′	H2	0.174600	HG3	HC	-0.027703
N9	N2	0.050430	CB	CT	0.060989
C8	CM	-0.116159	HB2	HC	0.029341
H8	HA	0.201235	HB3	HC	0.029341
C7	CZ	-0.149970	CA	CT	-0.620813
C5	CM	-0.744223	N	DU	-0.175790
C6	C	0.987862	H	H	0.191790
O6	O	-0.669103	HA	H1	0.193453
N1	NA	-0.811978	C	CZ	0.914275
H1	H	0.413755	O	O	-0.639776
C2	CA	0.887564			
N2	N2	-0.951631			
H21	H	0.422167			
H22	H	0.422167			
N3	NC	-0.748573			
C4	CM	0.852036			

Table S7. AMBER atom name, atom type and partial charge assignment for CTP.

Atom Name	Atom Type	Partial Charge
O1G	O3	-0.952600
PG	P	1.265000
O2G	O3	-0.952600
O3G	O3	-0.952600
O3B	OS	-0.532200
PB	P	1.385200
O1B	O2	-0.889400
O2B	O2	-0.889400
O3A	OS	-0.568900
PA	P	1.253200
O1A	O2	-0.879900
O2A	O2	-0.879900
O5′	OS	-0.598700
C5′	CT	0.055800
H5′1	H1	0.067900
H5′2	H1	0.067900
C4′	CT	0.106500
H4′	H1	0.117400
C3′	CT	0.202200
O3′	OH	-0.654100
HO′3	HO	0.437600
H3′	H1	0.061500
C2′	CT	0.067000
H2′	H1	0.097200
O2′	OH	-0.613900
HO′2	HO	0.418600
O4′	OS	-0.354800
C1′	CT	0.006600
H1′	H2	0.202900
N1	N*	-0.048400
C6	C4	0.005300
H6	H4	0.195800
C5	C4	-0.521500
H5	HA	0.192800
C4	CA	0.818500
N4	N2	-0.953000
H41	H	0.423400
H42	H	0.423400
N3	NC	-0.758400
C2	C	0.753800
O2	O	-0.625200

Table S8. AMBER atom name, atom type and partial charge assignment for N1-protonated ATP.

Atom Name	Atom Type	Partial Charge
O1G	O3	-0.952600
PG	P	1.265000
O2G	O3	-0.952600
O3G	O3	-0.952600
O3B	OS	-0.532200
PB	P	1.385200
O1B	O2	-0.889400
O2B	O2	-0.889400
O3A	OS	-0.568900
PA	P	1.253200
O1A	O2	-0.879900
O2A	O2	-0.879900
O5′	OS	-0.598700
C5′	CT	0.055800
H5′1	H1	0.067900
H5′2	H1	0.067900
C4′	CT	0.106500
H4′	H1	0.117400
C3′	CT	0.202200
H3′	H1	0.061500
O3′	OH	-0.654100
HO3	HO	0.437600
C2′	CT	0.067000
H2′	H1	0.097200
O2′	OH	-0.613900
HO2	HO	0.418600
O4′	OS	-0.354800
C1′	CT	0.039400
H1′	H2	0.200700
N9	N*	-0.073941
C8	CK	-0.010993
H8	H5	0.168663
N7	NB	-0.514375
C5	CB	-0.074272
C6	CA	0.632310
N6	N2	-1.070016
H61	H	0.452552
H62	H	0.452552
N1	NA	-0.586809
H1	H	0.417818
C2	CQ	0.134321
H2	H5	0.123363
N3	NC	-0.682554
C4	CB	0.507381

Supplementary Figures

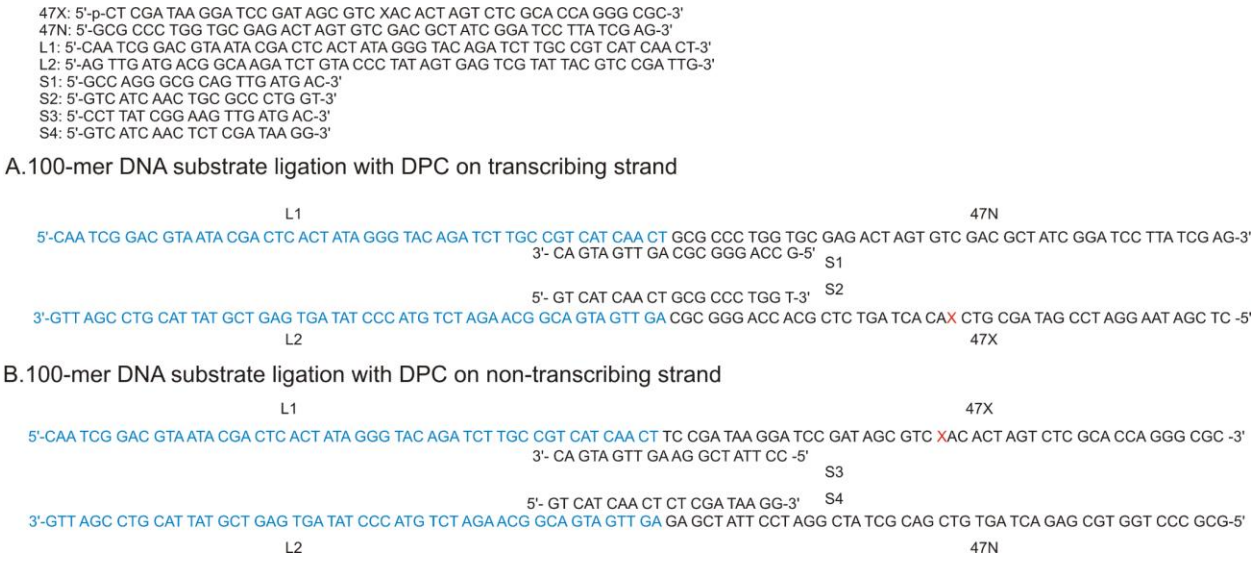


Figure S1. Ligation scheme of 47-mer oligonucleotides containing 5fC or 7-deaza-DHP-dG into double stranded 100-mer DNA duplex with lesion on transcribing (A) or non-transcribing strand (B). X = 5fC or 7-deaza-DHP-dG.

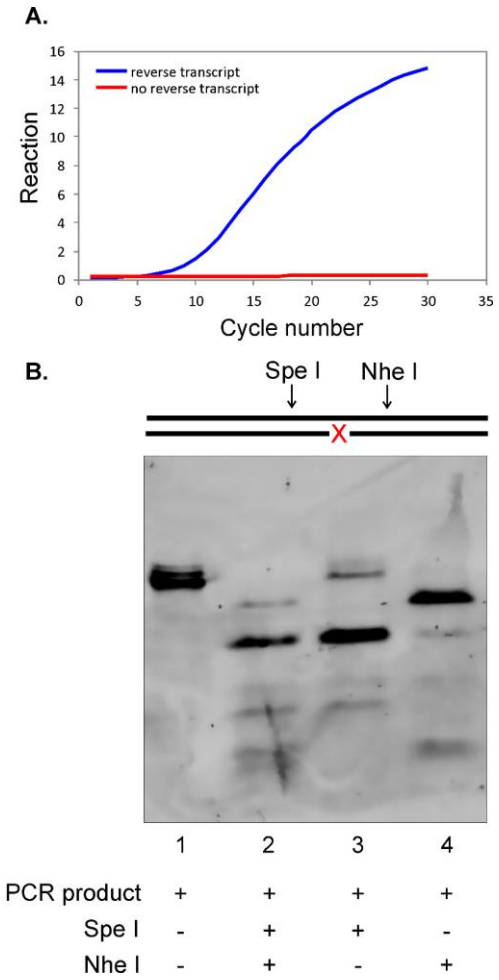


Figure S2. (A) RT-PCR amplification of transcripts and (B) the double restriction enzyme digestion of RT-PCR products for LC-MS/MS analysis.

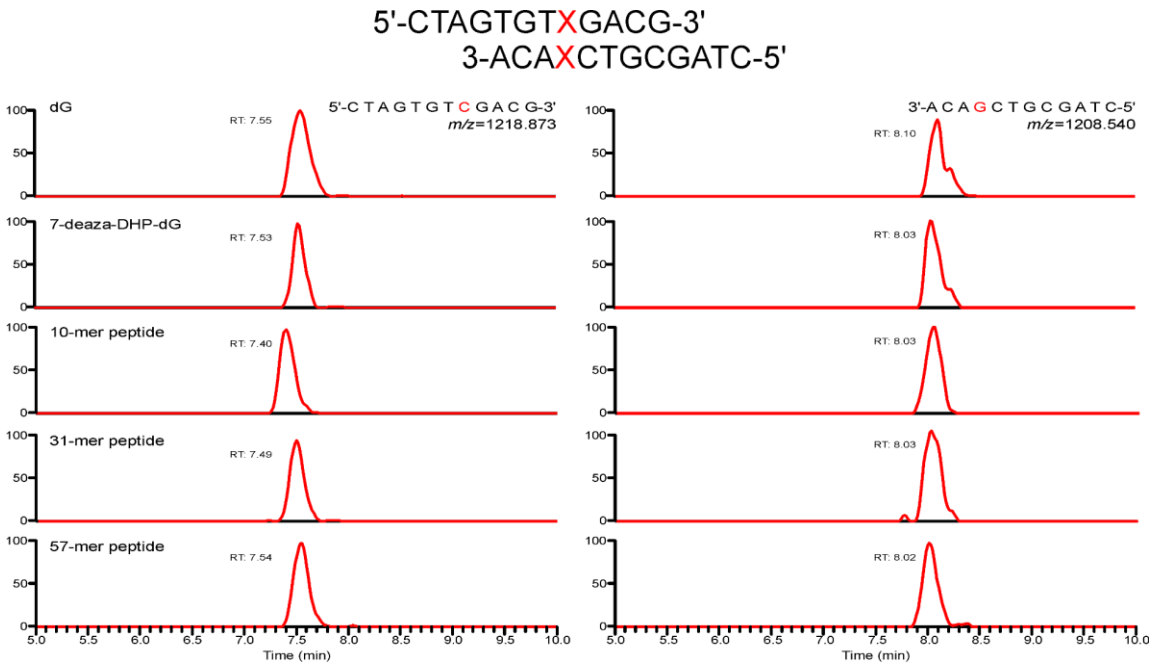


Figure S3. Extracted ion chromatograms of transcription bypass products of DNA-peptide cross-links at C7 of 7-deazaguanine. Error-free transcription products (5'-CTAGTGTCGACG-3', [M-3H]3-=1218.87) (left panel) and its complementary strands (5'-CTAGCGTCGACA, [M-3H]3-=1208.54) (right panel).

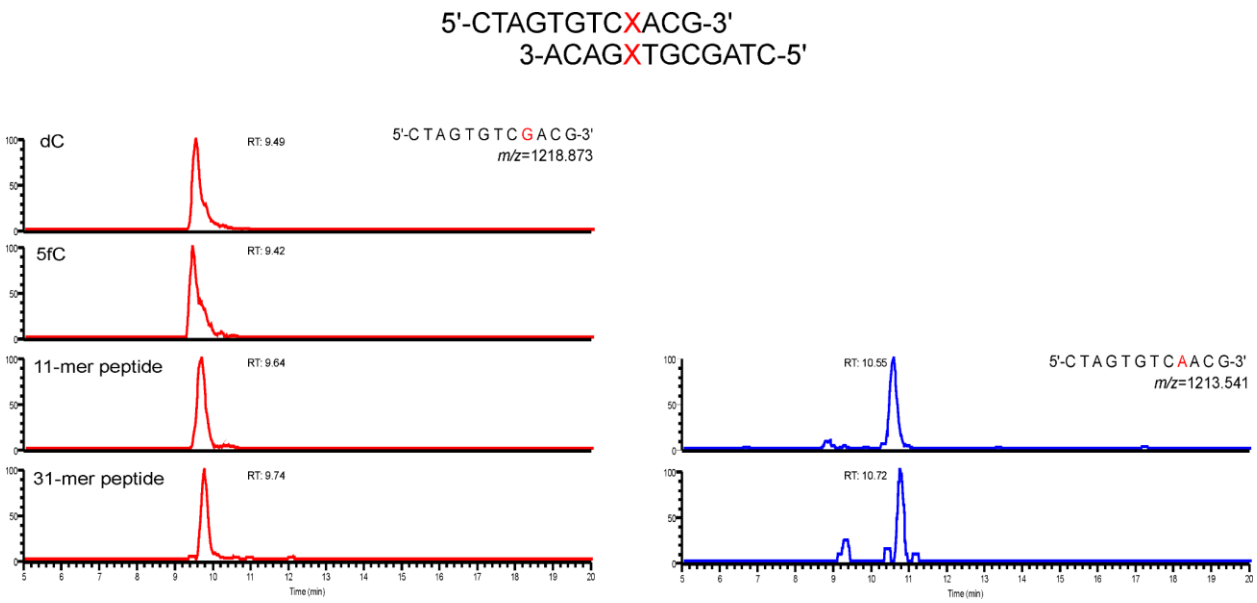


Figure S4. Extracted ion chromatograms of transcription bypass products of DNA-peptide cross-links at C5 of cytosine. Error-free transcription products (5'-CTAGTGTCGACG-3', [M-3H]3-=1218.87) (left panel) and C to T mutations (5'-CTAGTGTCAACG-3', [M-3H]3-=1213.54) (right panel).

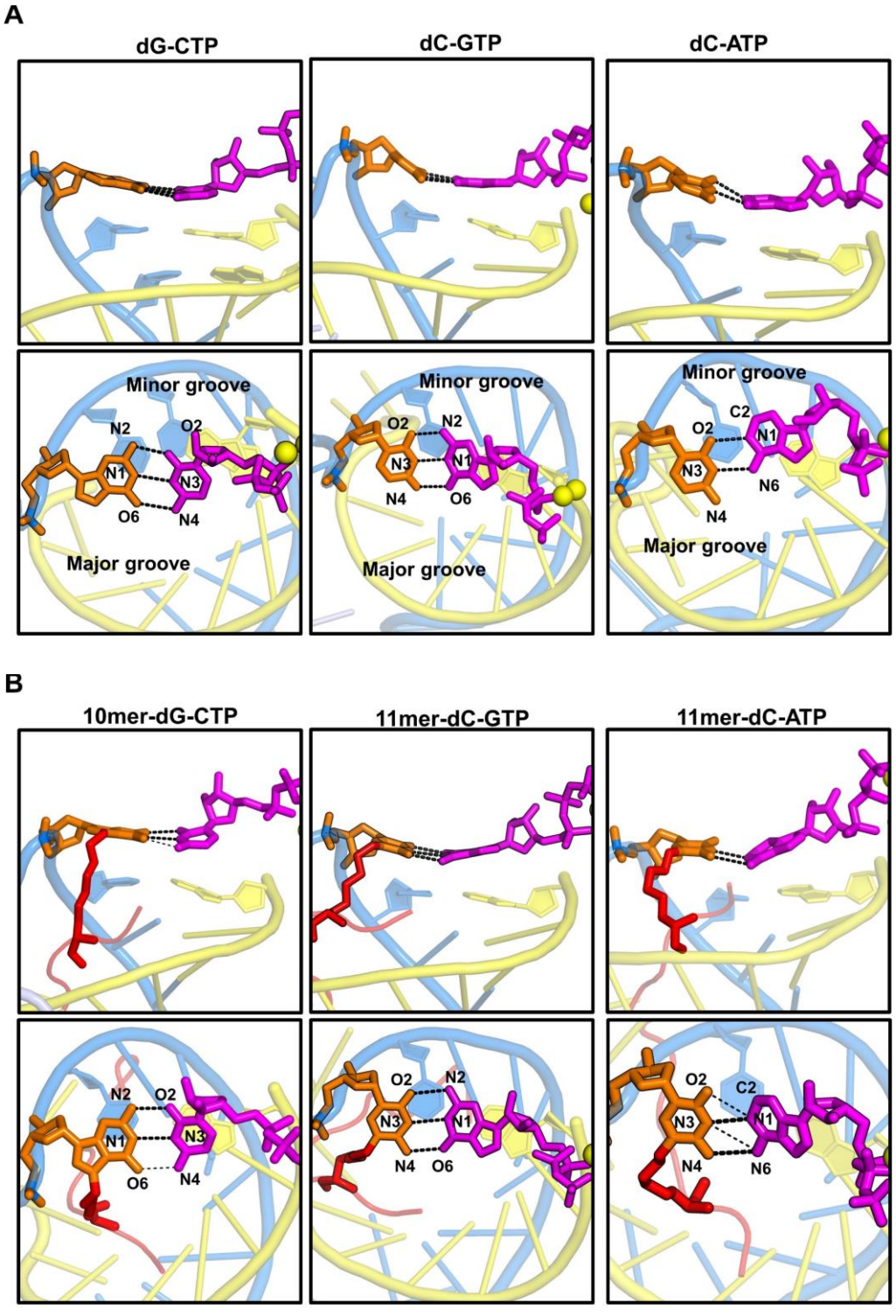

A
dG-CTP
dC-GTP
dC-ATP
Minor groove
N2
O2
N1
N3
O6
N4
Major groove
Minor groove
O2
N2
N3
N1
N4
O6
Major groove
Minor groove
C2
O2
N1
N3
N6
N4
Major groove
B
10mer-dG-CTP
11mer-dC-GTP
11mer-dC-ATP
N2
O2
N1
N3
O6
N4
O2
N2
N3
N1
N4
O6
O2
C2
N3
N1
N4
N6

Figure S5. The geometry of the template base-NTP base-pairs and their hydrogen bonds.

The best representative structures of the MD simulations are shown for (A) unmodified models and (B) DpC-containing models. Hydrogen bonds with occupancies > 60% are displayed in thick dashed line, and lower values are shown in thin dashed lines. Hydrogen bond pairing partners, occupancy, average distances and angles are given in Table S2. The base-pair parameters, including Buckle, Propeller, Opening, Shear, Stretch, and Stagger, are given in Table S3. Views are (top) from the major groove of the hybrid duplex and (bottom) looking down the duplex helical axis. The templating DNA base and NTP are in orange and magenta sticks, respectively; the peptide is in red, the Mg2+ ion is in yellow sphere, the DNA is in cyan cartoon and the RNA is in yellow cartoon.

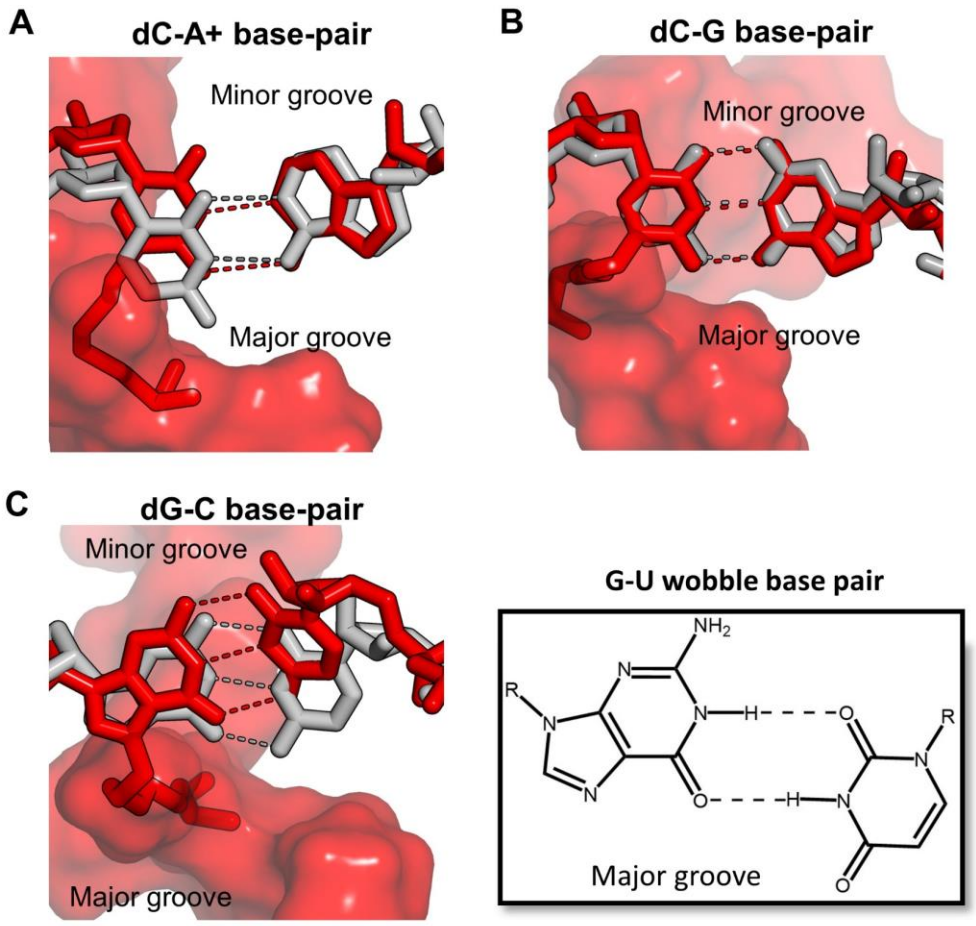


Figure S6. Template base-NTP base-pair alignment with (red) and without (gray) bulky DpCs.

(A) In the case of the unmodified dC-ATP mismatched base-pair, dC-A+ forms a stable wobble pair (Table S2). However, in the presence of the 11mer peptide, the dC-A+ forms a shifted two-hydrogen bonding scheme, in which the peptide causes the modified dC to move toward the minor groove. (B) For the dC-G matched base-pair, the peptide does not affect the position of the template base dC which is in a well-aligned Watson-Crick pair with GTP, although the GTP is modestly lifted from the RNA primer terminus in the presence of the peptide (Table S4). (C) The DpC-conjugated to dG greatly lifts the incoming CTP from the RNA terminus (Table S4) so that the Watson-Crick hydrogen bonds are distorted (Tables S2-3).

The best representative structures of each T7 polymerase system are shown. The unmodified template base cases are colored in gray and the DpC-containing cases are colored in red with the bulky DpCs in red surface rendering.

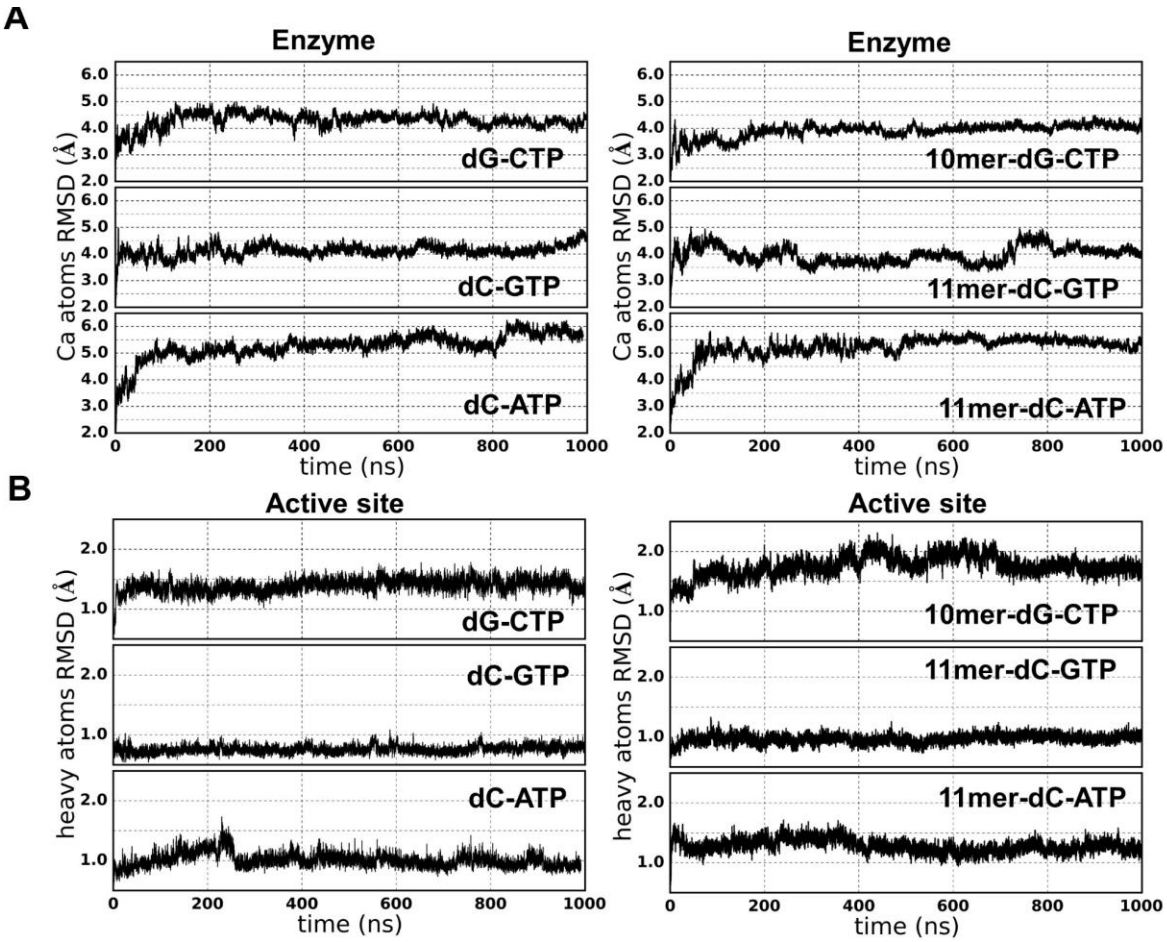


Figure S7. T7 RNA polymerase structural stabilities over the course of the 1 μs simulation for each system.

(A) Ca RMSDs of the polymerase protein. (B) Heavy atom RMSDs of the templating base, the incoming NTP, and the end base-pair of the DNA-RNA hybrid duplex.

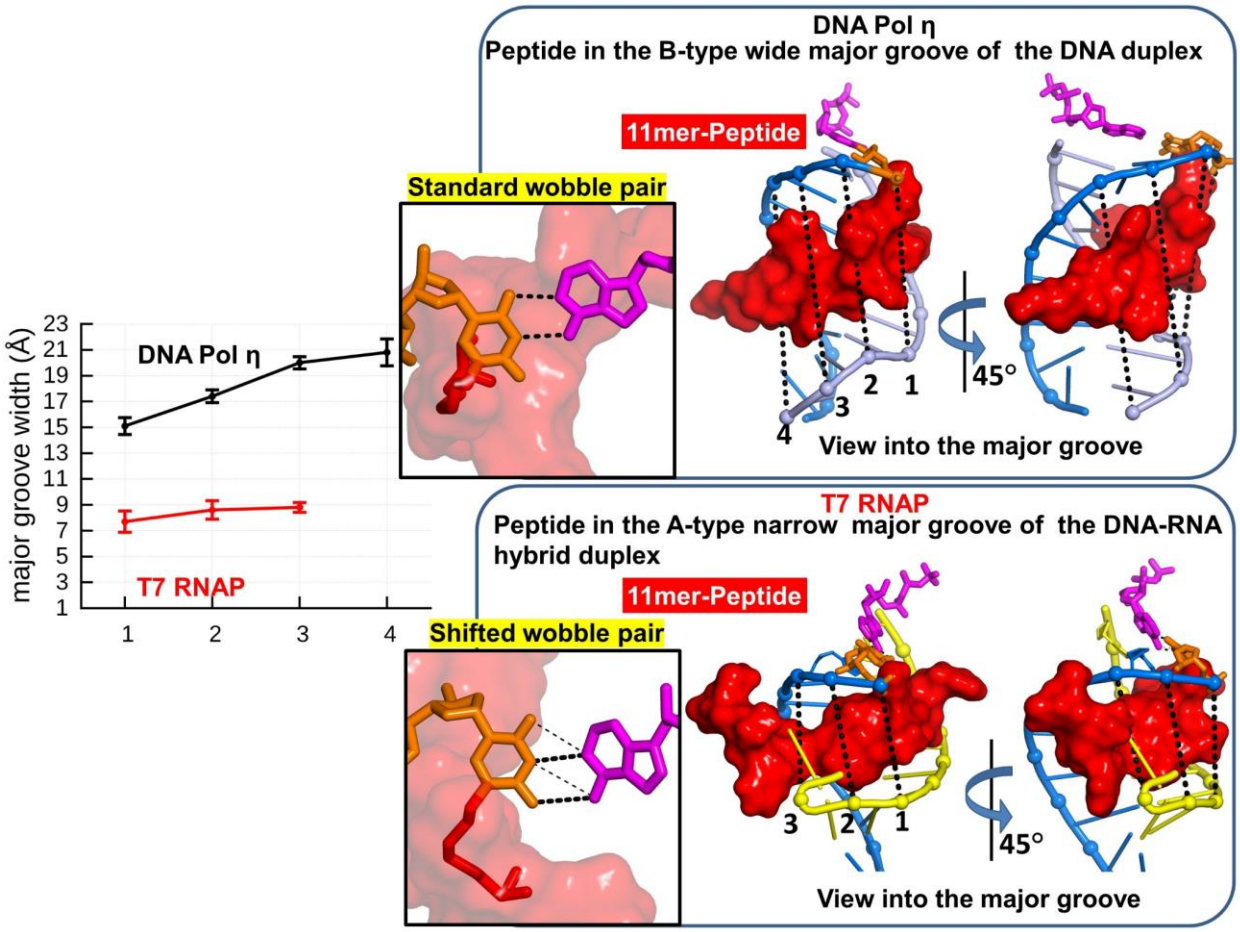


Figure S8. DpC-containing dC-dATP and dC-ATP base-pair in DNA pol η[22] and T7 RNA polymerase, respectively. The major groove width of the DNA-RNA hybrid duplex in T7 RNA polymerase is much narrower than in the DNA duplex in DNA pol η; this impacts the orientation of the conjugated-DpC in the major groove and consequently the hydrogen bonding in the dC-A+ mismatched base-pair.

DpC-containing dC-dATP forms a stable C-A+ wobble pair in the DNA polymerase. In the RNA polymerase, two hydrogen bonds between DpC-containing dC-ATP are still retained. However, this hydrogen bonding scheme is shifted from the standard C-A+ wobble pair, due to the bulky peptide in the confined major groove, which shifts the template dC toward the minor groove. The major groove width is between 15 and 21 Å in the B-type DNA duplex in DNA polymerase η and is around 8−9 Å in the A-type DNA-RNA hybrid in T7 RNA polymerase. The major groove width (black dotted line) is defined in the Supplementary Methods Section. The peptide is shown in red surface rendering, template dC is in orange sticks, incoming ATP/dATP is in magenta sticks, DNA is in blue and light blue cartoon rendering and RNA is in yellow cartoon rendering.

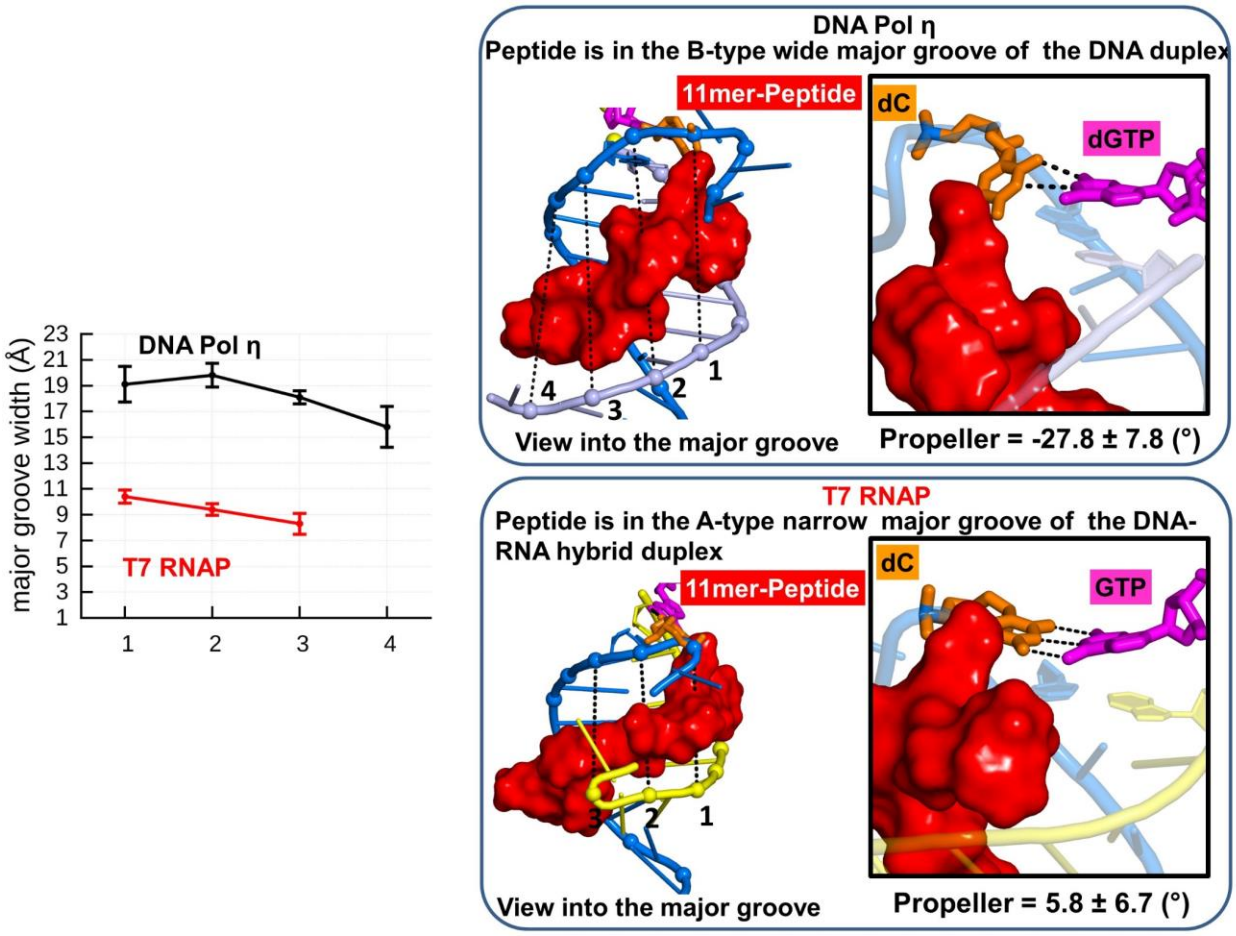


Figure S9. DpC-containing dC-dGTP and dC-GTP base-pair in DNA pol η[22] and T7 RNA polymerase, respectively. Hydrogen bonds of the DpC-containing dC-GTP in T7 RNA polymerase are better-aligned than the dC-dGTP base-pair in DNA pol η[22].

The DpC in the wide and spacious major groove of DNA Pol **η** draws the template base dC toward the peptide; it causes the template base dC to be greatly propeller-twisted (~ 28°) so that only one full and two ~ half hydrogen bonds are formed. By contrast, in T7 RNA polymerase, the much narrower major groove confines the DpC so that the Watson-Crick hydrogen bonds between dC and GTP are well-aligned.

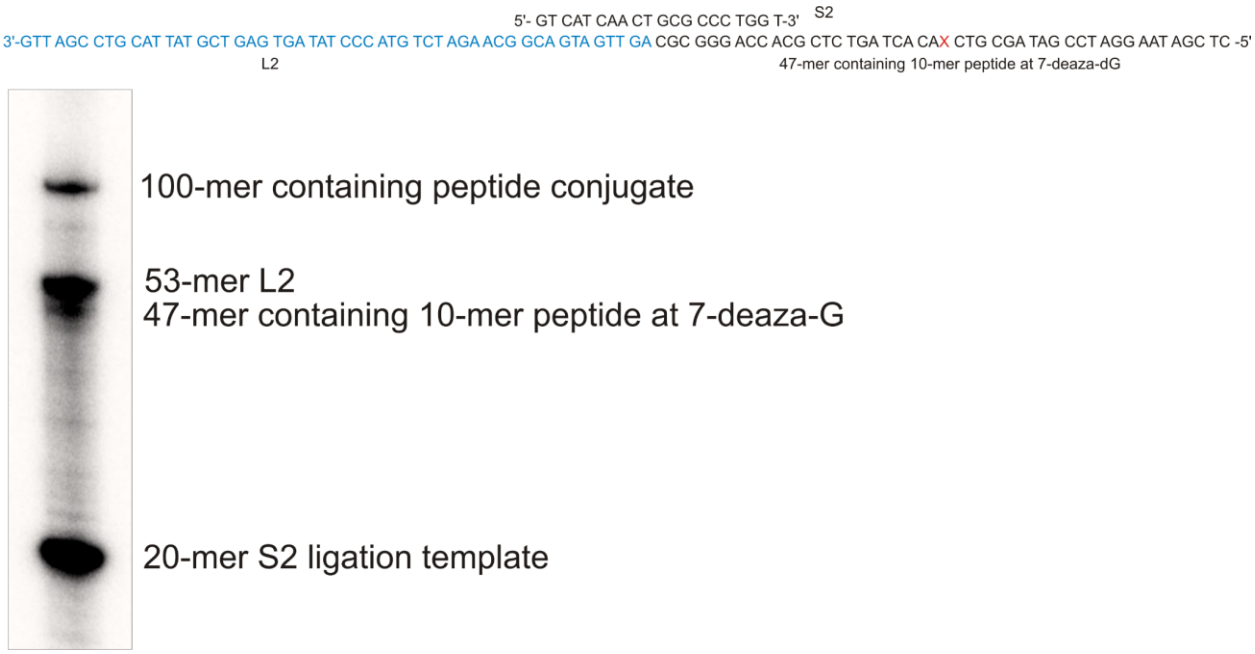


Figure S10. Representative gel to construct 100-mer oligo containing site-specific DNA-peptide cross-links by ligation.

Supplementary Movie

Movie S1: T7 RNA polymerase ternary complex, containing the 11mer peptide cross-linked to the templating dC opposite to the N1 protonated incoming ATP.

References

(1) Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The protein data bank. *Nucleic Acids Res. 28*, 235-242.

(2) Yin, Y. W., and Steitz, T. A. (2004) The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell 116*, 393-404.

(3) Fiser, A., and Sali, A. (2003) Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol 374*, 461-491.

(4) Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field. *J. Comput. Chem. 25*, 1157-1174.

(5) Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: Improving the cccuracy of protein side chain and backbone parameters from ff99SB. *J. Chem.Theory Comput. 11*, 3696-3713.

(6) Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins 65*, 712-725.

(7) Case, D. A., Darden, T. A., Cheatham, T. E., 3rd, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M. J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2014) AMBER 14, University of California, San Francisco.

(8) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.;

Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. (2013) Gaussian 09, Gaussian, Inc.

(9) Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P. A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the resp Model. *J. Phys. Chem. 97*, 10269-10280.

(10) Meagher, K. L., Redman, L. T., and Carlson, H. A. (2003) Development of polyphosphate parameters for use with the AMBER force field. *J. Comput. Chem. 24*, 1016-1025.

(11) Tesmer, J. J., Sunahara, R. K., Gilman, A. G., and Sprang, S. R. (1997) Crystal structure of the catalytic domains of adenylyl cyclase in a complex with Gsalpha.GTPgammaS. *Science 278*, 1907-1916.

(12) Cieplak, P., Cornell, W. D., Bayly, C., and Kollman, P. A. (1995) Application of the multimolecule and multiconformational resp methodology to biopolymers - charge derivation for DNA, RNA, and proteins. *J. Comput. Chem. 16*, 1357-1377.

(13) Jorgensen, W. L., Chandreskhar, J., Madura, J. D., Imprey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926-935.

(14) D.A. Case, R.M. Betz., D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke,, A.W. Goetz, N. H., S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C., Lin, T. L., R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I., Omelyan, A. O., D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails,, and R.C. Walker, J. W., R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman (2016) AMBER 2016, University of California, San Francisco.

(15) Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys. 81*, 3684-3690.

(16) Ryckaert, J. P., Ciccotti, G., and C., B. H. J. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys. 23*, 327-341.

(17) Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: an *N* log(*N*) method for Ewald sums in large systems. *J. Chem. Phys. 98*, 10089-10092.

(18) Le Grand, S., Gotz, A. W., and Walker, R. C. (2013) SPFP: speed without compromise-a mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Physics Commun. 184*, 374-380.

(19) Roe, D. R., and Cheatham, T. E. (2013) PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory Data. *J. Chem. Theory Comput. 9*, 3084-3095.

(20) Shao, J. Y., Tanner, S. W., Thompson, N., and Cheatham, T. E. (2007) Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *J. Chem. Theory Comput. 3*, 2312-2334.

(21) Lu, X. J., and Olson, W. K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res. 31*, 5108-5121.

(22) Ji, S., Fu, I., Naldiga, S., Shao, H., Basu, A. K., Broyde, S., and Tretyakova, N. Y. (2018) 5-Formylcytosine mediated DNA-protein cross-links block DNA replication and induce mutations in human cells. *Nucleic Acids Res 46*, 6455-6469.