Cooperative motion of a key positively charged residue and metal ions for DNA replication catalyzed by human DNA Polymerase-η

Vito Genna,[1] Roberto Gaspari,[1,2] Matteo Dal Peraro[3,4] and Marco De Vivo[1,5*]

[1] Laboratory of Molecular Modeling & Drug Discovery,
Istituto Italiano di Tecnologia, Via Morego 30, 16163, Genoa, Italy

[2] CONCEPT Lab.
Istituto Italiano di Tecnologia, Via Morego 30, 16163, Genoa, Italy
[3] Institute of Bioengineering, School of Life Sciences,

École Polytechnique Fédérale de Lausanne - EPFL, Lausanne, Switzerland
[4] Swiss Institute of Bioinformatics - SIB, Lausanne, Switzerland

[5]IAS-5 / INM-9 Computational Biomedicine Forschungszentrum Jülich
Wilhelm-Johnen-Straße 52428 Jülich, Germany

*Corresponding author:
Marco De Vivo
Email: marco.devivo@iit.it

Supporting Information

Table of contents

Supplementary text

Supplementary figures

Supplementary tables

Well-Tempered Metadynamics

In well-tempered metadynamics, a history-dependent repulsive potential, $V_{bias}(s,t)$ (where the quantity s is the vector of our CVs), is built up by adding a biasing potential term, in the form of a small Gaussian "hill", along the CVs after some MD steps:

$$V_{bias}(s,t) = \quad [1]$$

where δs and ω are the width and height of the Gaussian hills, which are centered at the values of the CVs that have already been visited by the system and deposited at a time interval τ_G. As the hills are build up along the CVs, the system is forced to escape local *minima* exploring therefore higher energy regions of the free energy surface, $F_G(s) = -V_{bias}(s, t \to \infty)$.

Notably, the error associated with this approach is demonstrated be related on the metadynamics parameters, $w, \delta s$ and τ_G, Its origin lies in the false assertion that the estimate of the free energy obtained from the biasing potential tends towards the real and unbiased value. Indeed, once the biasing potential fully compensates for $F(s)$, the addition of extra Gaussian hills serves to introduce error, so that $-V_{bias}(s,t)$ fluctuates around the correct value of $F(s)$, with the magnitude of the fluctuations depending on the size of the hills added. Barducci et al. have proposed an approach called "well-tempered" and "smoothly converging" form of the algorithm, herein used, by which the height of the Gaussian hills added are modified as follows:

$$w = \omega e^{-[V_{bias}(s,t)/\Delta T]_{\tau_G}} \quad [2]$$

where ω is the bias deposition rate with units of energy per unit of time, $V_{bias}(s,t)$ is the estimate of the free energy at the current CVs positions and the current time step (as also described in ref Barducci et al.), and ΔT is a tunable temperature-like parameter that regulates how rapidly w decreases.

In this scheme the final value of the biasing potential is a scaled approximation to $F(s)$:

$$F(s) = -\frac{T + \Delta \mathrm{T}}{\Delta T} V_{bias}(s, t \to \infty) \quad [3]$$

Through this approach, at the beginning of metadynamics run, the biasing potential is zero and therefore $w = \omega$, which can be quite a large value, allowing the wells to be filled quickly and leading to a fast exploration of phase space. As the wells begin to fill up, ω is scaled and progressively smaller perturbations are made to V_{bias}, allowing it to smoothly converge to a correct approximation of $F(s)$. The value of ω and ΔT are chosen to achieve the efficiency.

Confidence Interval analysis

To determine whether the *d-newbond* distributions are significantly different with respect to the different Arg61 conformations (i.e. **A-**, **B-** and **C-conf**) we performed a confidence interval analysis. The null hypothesis is expressed by $\boldsymbol{\mu_X=\mu_Y}$, where μ represents the mean of the distribution, while X and Y are two different conformations among **A**, **B** and **C**.

In order to test the null hypothesis, we computed the distributions of the *d-newbond* differences, namely d_{X-Y}, for all possible cases and evaluated the mean μ and the associated standard error σ_E. Values are reported in Fig. S3-D, together with the confidence interval, C.I., at which the null hypothesis can be rejected. The populations of **A-**, **B-** and **C-conf** were respectively 10586, 8640, 7950 and were obtained by sampling the MD trajectory every 10 ps, in the various conformations (see Fig. S3, panel D).

Transition State Theory

Transition State Theory (TS-theory) in the Eyring-Polanyi (Laidler,K. and King,C., 1983. *J. Phys. Chem.*, 15, 2657*)* formulation reads:

$$k_{a\to b} = \frac{kT}{h} e^{\frac{\Delta G^*_{a\to b}}{kT}} \qquad (1)$$

where $k_{a\to b}$ generally represents the rate of interconversion from state a to state b, $\Delta G^*_{a\to b}$ is the associated free energy barrier (~8.00 kcal mol^{-1} in our case), kT is the Boltzmann factor (~0.60 kcal mol^{-1} at room temperature), and h is the Planck constant.

In our case state a and state b correspond to the MgB-MgC-PPi adduct bound and unbound to the active site, respectively. In eq. (1) we can set $\Delta G^*_{a\to b}$ to the value estimated by metadynamics simulations (see Result section in the manuscript). TS-theory provides a strongly simplified picture of the PPi unbinding kinetic and serves here as a tool to compare the order of magnitude of the PPi leaving rate with the whole enzymatic turnover rate.

Our estimation of $k_{a\to b}$ is $10^4\ s^{-1}$ to be compared with the experimentally measured PPi leaving constants, k_{cat} = ~25 s^{-1}, in Dpo4, a member of Y-family polymerase (Beckman et al., **2008**, *J. Bio. Chem.*, 283, 36711-36723). Since kcat depends of many enzymatic steps, besides PPi leaving, (see

enzymatic turnover from reagents to products), our data are indicate PPi unbinding process as a non-rate limiting step in DNA polymerization catalyzed by Pol-η, in agreement with the findings proposed by Beckman et al. (**2008**, *J. Biol. Chem.*, 283, 36711-36723) and Zhao and collogues (**2014**, *FEBS J.*, 281, 4394-4410) for Y-family members Dpo4 and Pol-ϰ, respectively.

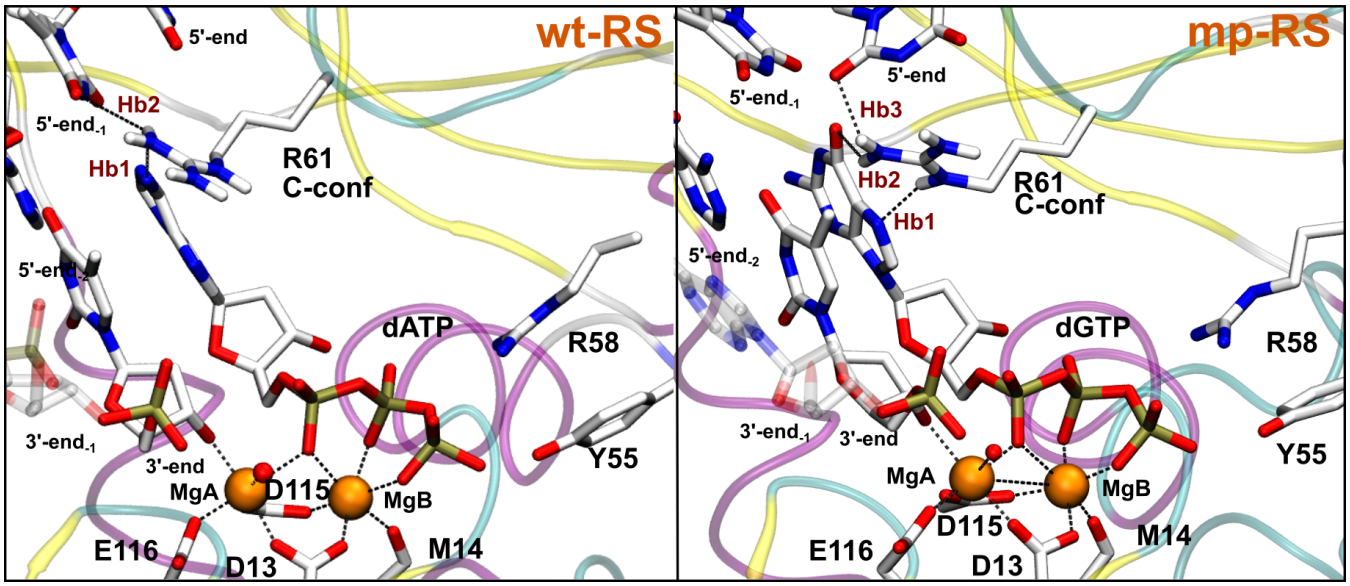


Figure S1. Different prereactive states of Poly-η. Left: Wild-type **wt-RS** system showing Arg61 in C-conf, where it establishes two hydrogen bonds (Hb1, Hb2) with the N7 atom of the incoming adenine and the apical oxygen of the templating thymine, respectively. **Right**: Model system containing a dGTP:T mispair, **mp-RS**. Here, Arg61 establishes three hydrogen bonds (Hb1, Hb2 and Hb3): Hb1 with the N7 atom of the incoming adenine; Hb2 with the O atom of the guanine; and, Hb3 with the O2 oxygen of terminal thymine. White is used for carbon, red for oxygen, blue for nitrogen, tan for phosphorus and orange for the two magnesium ions.

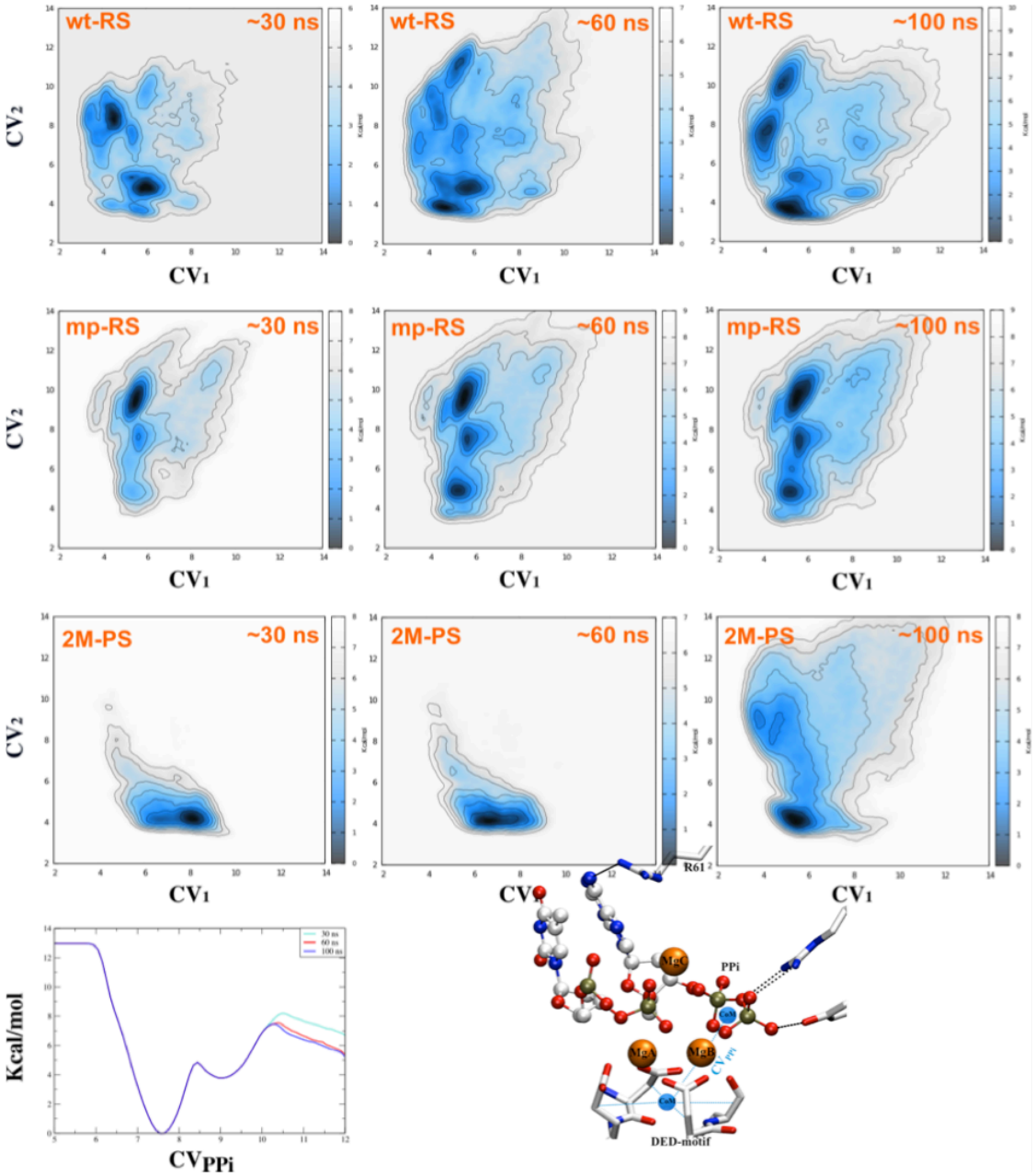


Figure S2. Convergence of the free energy simulations. Convergence was checked considering the location of the minima as a function of time. From ~60 ns to ~100 ns, no significant changes were detected. On this basis, we considered converged our well-tempered metadynamics simulations. The ensemble of conformational states of Arg61 found on the reconstructed free energy surfaces reproduced well the overall Arg61 rotamers detected throughout the unbiased MD simulations, improving the overall conformation sampling associated with the motion of this key residue. The postreactive site representation (bottom), indicates the collective variable adopted (CV_{PPi}) during PPi-leaving metadynamics, where the blue spheres indicate the two centers of mass used to define CV_{PPi}. Protein residues are represented in licorice while nucleic acid residues in CPK style. White is used for carbon, red for oxygen, blue for nitrogen, tan for phosphorus and orange for magnesium ions.

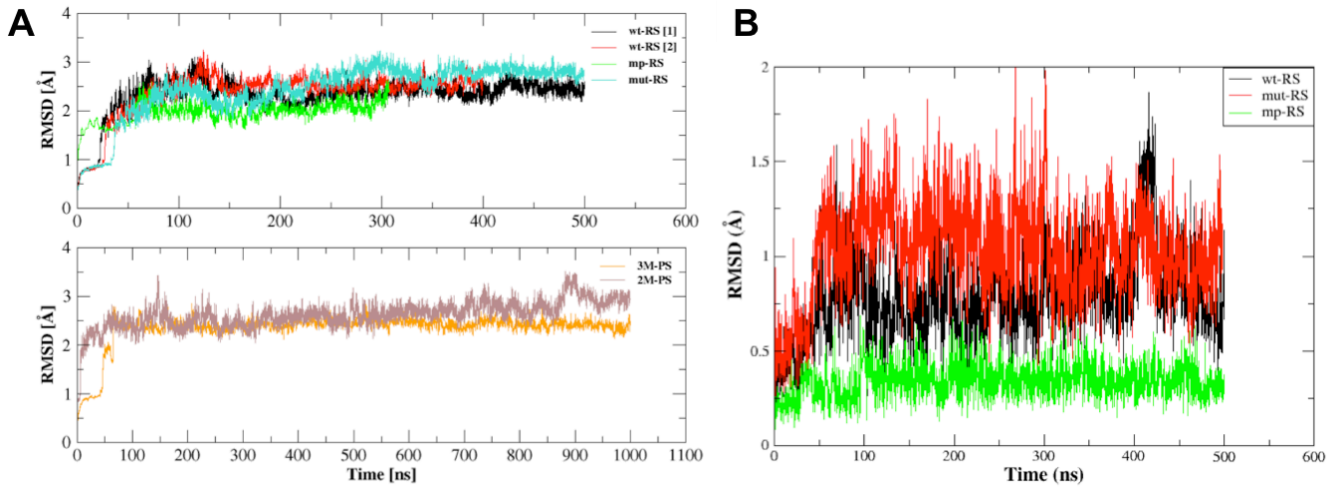
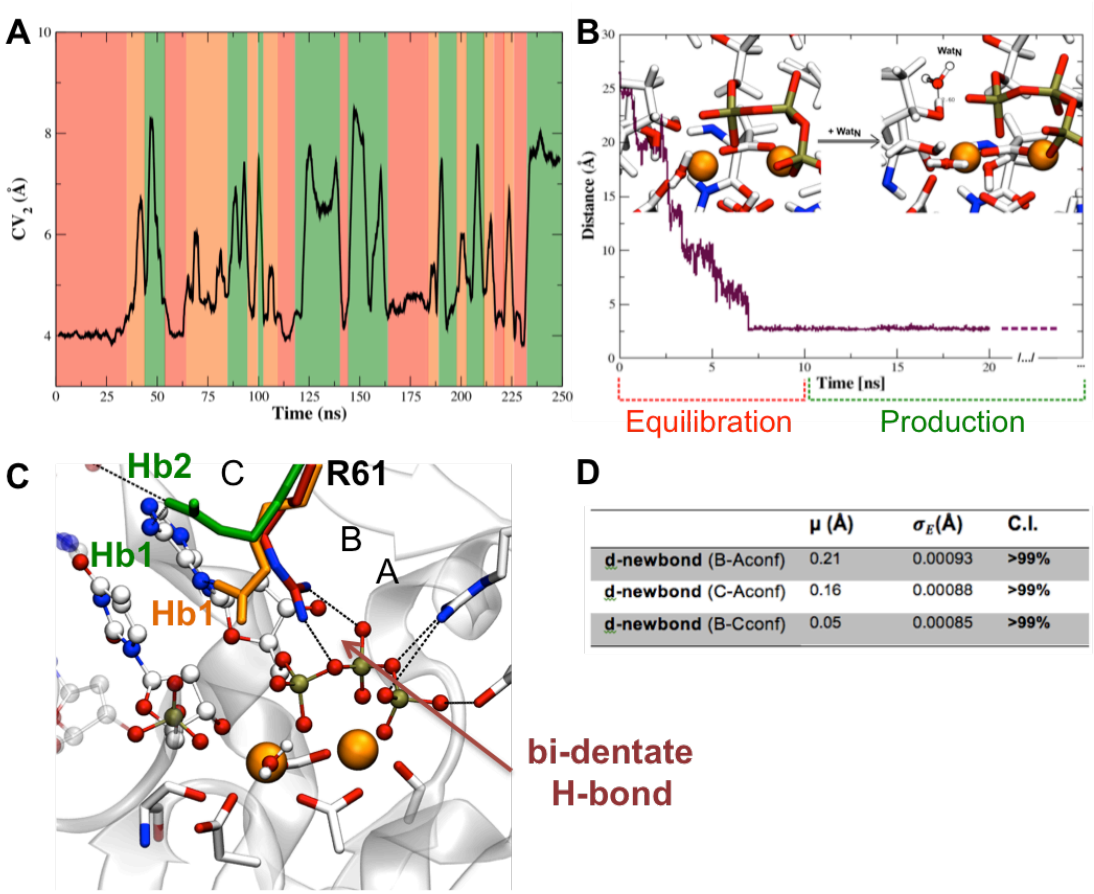


Figure S3. Root Mean Square Deviation (RMSD) of Pol-η and incoming nucleotide, backbone heavy atoms. (A, upper), Prereactive state: wt-RS system is in black, a second replica of wt-RS system is in red, the mp-RS system is in green, and mut-RS system (Arg61Ala mutant) is in cyan. The ternary Pol-η/dsDNA/dATP(dGTP) complex showed high stability in each prereactive state. The RMSD value for protein backbone is: 2.53 ± 0.27 Å for wt-RS, 2.73 ± 0.25 Å for mut-RS and 2.12 ± 0.24 Å for mp-RS. **(A, bottom)**, Postreactive state: in orange is indicated the 3M-PS system while in brown the 2M-PS system. The ternary Pol-η/dsDNA/PPi complex showed high stability in both systems. The backbone RMSD of the whole ternary complex with respect to the X-ray structure is 2.72 ± 0.33 Å and 2.53 ± 0.52 Å for the 3M-PS and 2M-PS, respectively. The DED-motif, which coordinated both MgA and MgB in the catalytic pocket, also exhibited high stability with an RMSD value of only 0.27 ± 0.07 Å, preserving the octahedral coordination for MgA and MgB, as well as MgC in 3M-PS, which conserved the polyhedral coordination geometry well. **(B)** Root Mean Square Deviation (RMSD) of the dATP heavy atoms in prereactive states. Black color indicates the wt-RS system; red indicates the mut-RS system, while green indicates dGTP RMSD value in the mp-RS system.



	μ (Å)	σ_E(Å)	C.I.
d-newbond (B-Aconf)	0.21	0.00093	>99%
d-newbond (C-Aconf)	0.16	0.00088	>99%
d-newbond (B-Cconf)	0.05	0.00085	>99%

Figure S4 Arg61 swinging in wt-RS. (A) Black line shows the different values of CV_2 along the simulated time scale. Red background marks CV_2 values corresponding to **A-conf**, orange to **B-conf** while green indicates **C-conf**. Here, the A↔B interconversion occurred seven times (t = ~30, ~65, ~110, ~180, ~210, ~220, ~230 ns) while nine A↔B interconversions were measured for the first replica system (at t = ~60, ~90, ~100, ~140, ~150, ~160, ~170, ~180, ~240 ns, see Fig. 2A in the manuscript). A↔C conformational switch, where **C-conf** was just transiently formed, happened six times in this second replica (t = ~55, ~115, ~140, ~145, ~165, ~230 ns) and four times during the simulated timescale of the first replica (t= ~190, ~210, ~215, ~225 ns, see Fig. 2A in the manuscript). **(B)** Notably, during the first few ns of simulation time, one water molecule (Wat_N) reached the catalytic site from the bulk, H-bonding the nucleophile 3'-OH of the incoming nucleotide. Once formed, the Wat_N-3'-OH H-bond exhibited great stability, maintaining an average length of 2.10 ± 0.08 Å throughout the collected trajectories. In this position, Wat_N is properly placed to act as a general base for the activation of the nucleophilic 3'-OH of the primer end and to initiate the catalytic phosphoryl-transfer reaction, as recently proposed by Nakamura et al. (ref.1 in manuscript) **(C)** Different Arg61 conformations detected in our MD simulations of wt-RS system**.** Dashed lines indicate H-bond interactions. Hydrogen atoms were removed for clarity. Arg61 is shown assuming A-conf (red), B-conf (orange), and C-conf (green). Typical H-bond are reported for each Arg61 conformation. **(D)** Confidence interval data of *d-newbond* in wt-RS, considering the 3 conformations (A, B and C, see text in SI for details on confidence interval calculations).

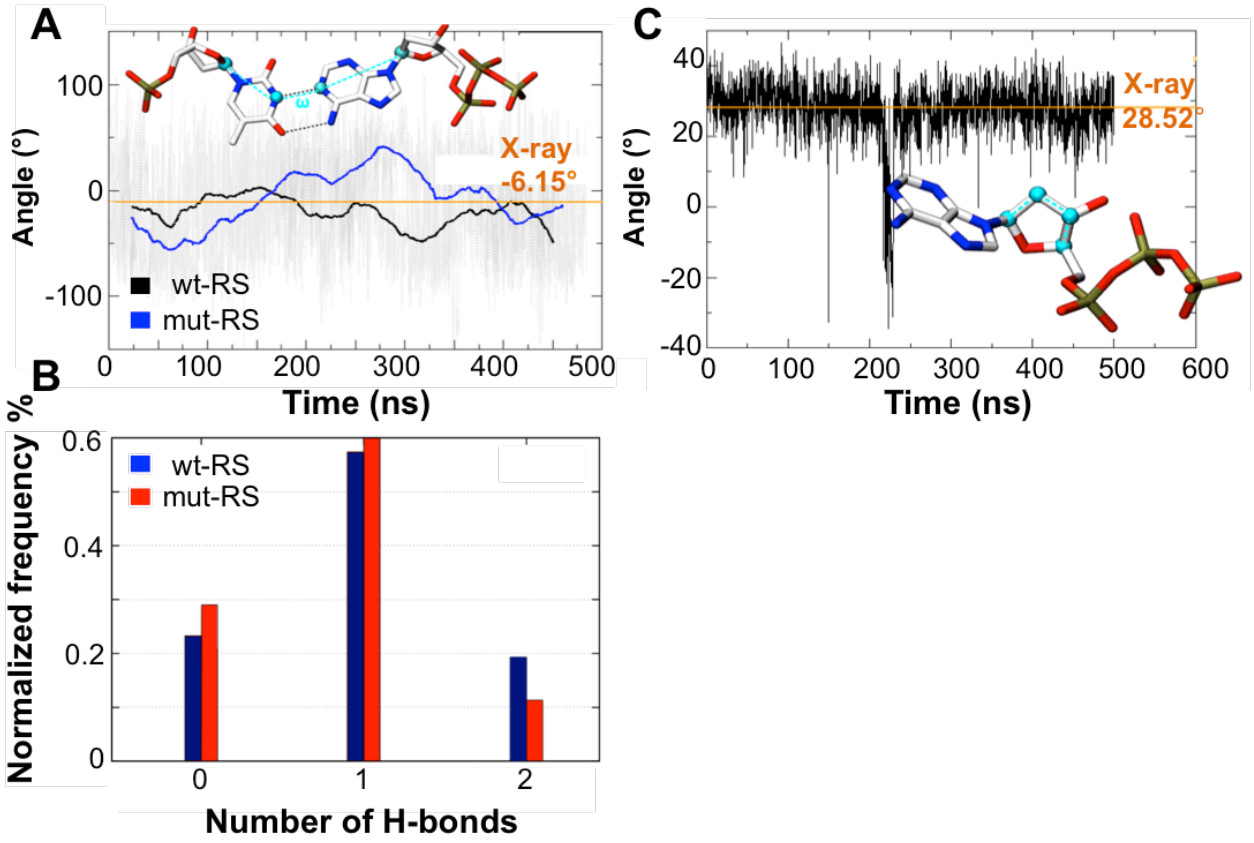


Figure S5. Further characterization of wt-RS vs. mut-RS. (A) value of the ω pseudo dihedral measured along the MD simulations of wt-RS and mut-RS systems. Angle's value quantifies the buckling of the terminal base-pair (dATP:T). Atoms forming ω pseudo dihedral (C1-N1-N3-C1) are represented in CPK and colored in cyan. Black line indicates fluctuations of the ω in wt-RS system while blue line describes those in mut-RS system. Orange line describes ω value detected in X-ray structure (PDBid 4ECS). As showed in the graph, in mut-RS system the terminal base pair get buckled, with a ω rotation of ~28.00° (with respect to the -6.15° found in the X-ray structure). Moreover, in mut-RS, the angles N-H-O and N-H-N and the distance N-O, N-N (which are characteristic of Watson and Crick's structure) increase to an average value of ~40.00° for N-H-O and N-H-N angles, and ~4.20 Å for N-O and N-N distances. This is in comparison to the X-ray, where they are ~20.00° and ~3.00 Å, respectively. **(B)** Frequency distribution of the Watson and Crick H-bond of the terminal base pair (dATP:T) in prereactive states. H-bond network between the incoming nucleotide and its templating base in wt-RS (Blue) and mut-RS (Red) systems. The graph shows a pronounced instability of the W&C interaction framework in mut-RS system. In fact, only the ~11% of the collected frames report the presence of two H-bond interactions between dATP:T. (**C**) Sugar pucker conformation adopted by dATP ribose in wt-RS. Atoms forming the analyzed pseudo dihedral are colored in cyan. Orange line shows the value of the angle detected in the crystal structure (PBDid 4ECS).

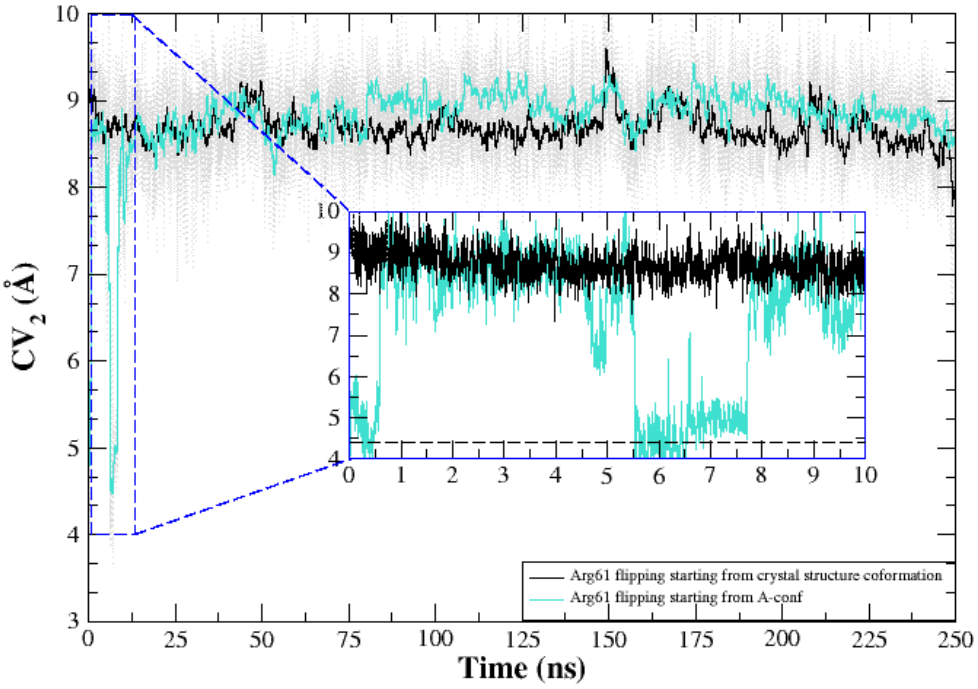


Figure S6. Arg61 swinging in mp-RS system. Black indicates CV_2 values in the mp-RS MD simulations in which Arg61 assumes its crystallographic pose as a staring conformation (C-conf, forming Hb1, Hb2 and Hb3). Cyan line represents CV_2 values in mp-RS in which Arg61 adopts A-conf as a starting pose. The graph shows the fast conformational A↔C interconversion even when A-conf is the starting pose.

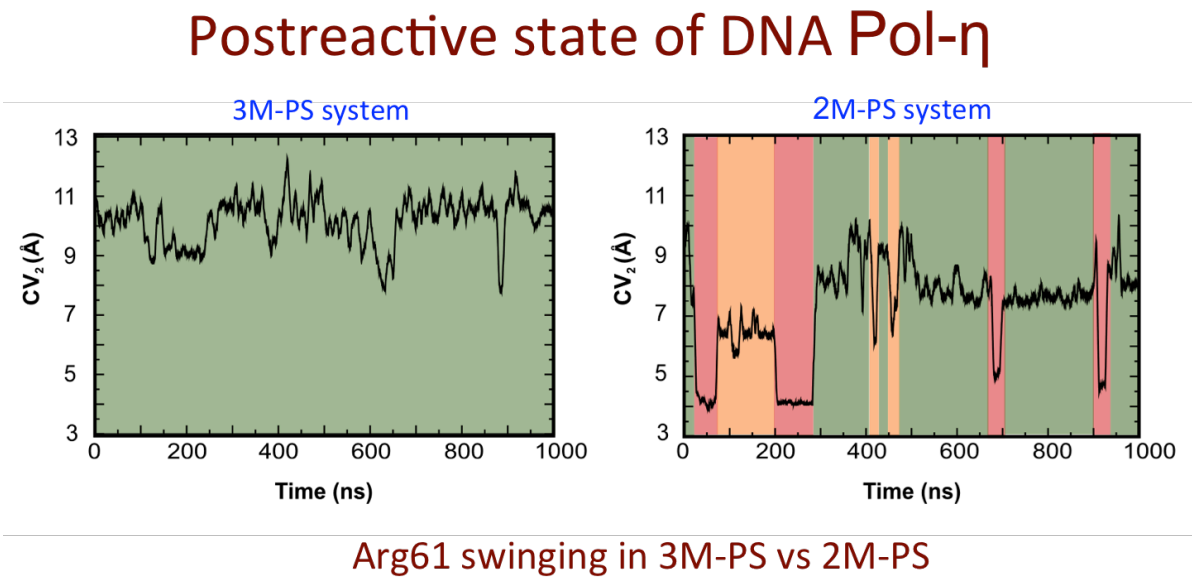


Figure S7. Arg61 swinging in 3M-PS vs. 2M-PS. 3M-PS: Pol-η in complex with dsDNA, PPi and three Mg^{2+} ions. **Right**, 2M-PS: Pol-η in complex with dsDNA, PPi and two Mg^{2+} ions. Black line indicates value of CV_2 (see manuscript) along the simulated timescale. Red, orange and green represent those CV_2 values which defining Arg61 **A- B-** or **C-conf**, respectively.

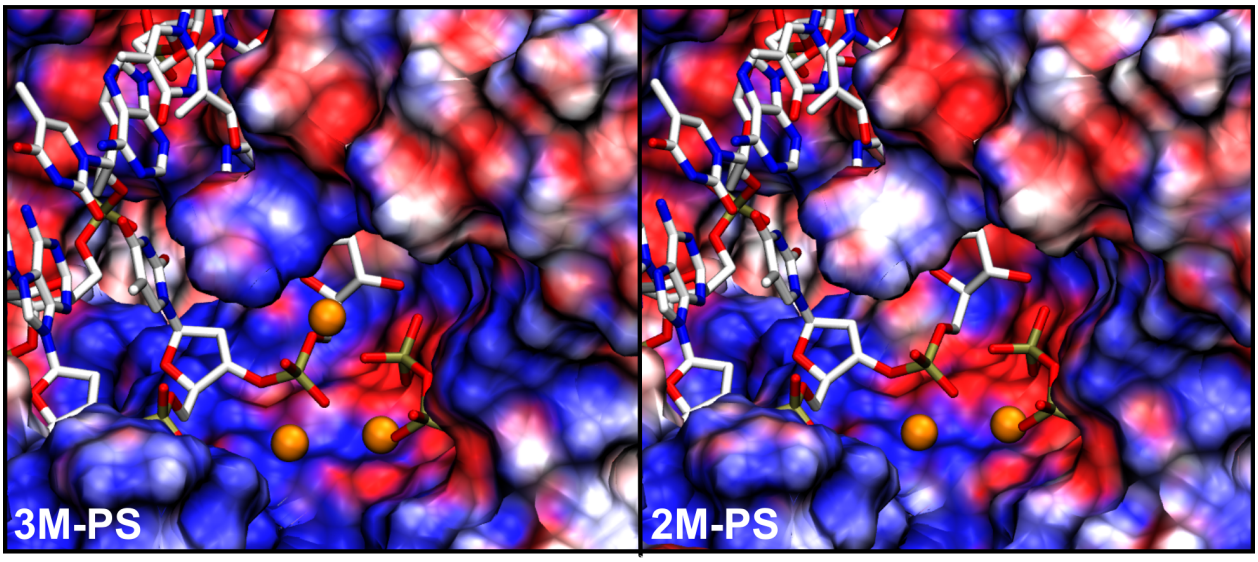


Figure S8. Electrostatic surface of Pol-η active site in postreactive state. Left, 3M-PS: Pol-η in complex with dsDNA, PPi and three Mg^{2+} ions. **Right**, 2M-PS: Pol-η in complex with dsDNA, PPi and two Mg^{2+} ions. The lack of the third metal, in 2M-PS, determines the formation of a more negatively charged environment which serves as driving force to recruit the third metal from the bulk solution.

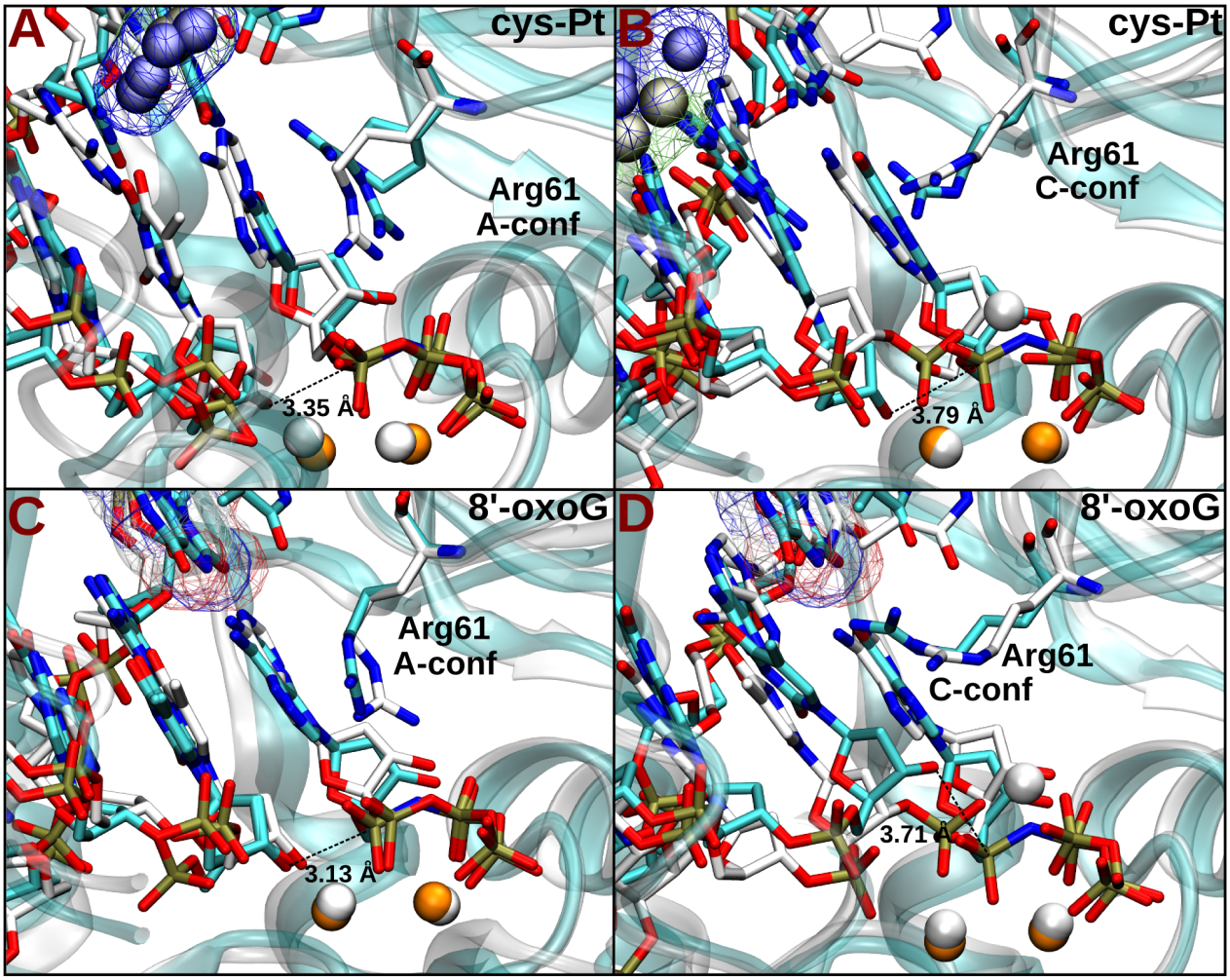


Figure S9. Structural superimpositions of Pol-η in complex with damaged or undamaged dsDNA. White carbons and metals indicate snapshot coming from MD simulations while cyan identify crystallographic structures; *d-newbond* is indicated by a black dashed line. (**A**) X-ray structure (PDBid 4DL4) of a prereactive configuration of Pol-η enzyme in complex with DNA containing a cis-Pt adduct (van der Waals representation). Crystallographic **A-conf** (cyan) well match **A-conf** described by our MD simulation (white). (**B**) X-ray structure (PDBid 4DL6) of a post-insertion configuration of Pol-η enzyme in complex with DNA containing a cis-Pt adduct (van der Waals representation). Crystallographic **C-conf** (cyan) well matches **A-conf** described by our MD simulation (white); here *d-newbond* adopts higher value with respect system reported in A. (**C**) X-ray structure (PDBid 4O3O) of a pre-reactive configuration of Pol-η enzyme in complex with DNA containing a 8-hydroxyguanine (8-oxoG) lesion (wireframe representation). Crystallographic **A-conf** (cyan) well matches **A-conf** described by our MD simulation (white). (**D**) X-ray structure (PDBid 4O3Q) of a pre-reactive configuration of Pol-η enzyme in complex with DNA containing a 8-hydroxyguanine (8-oxoG) lesion Crystallographic **C-conf** (cyan) well match **C-conf** described by our MD simulation (white).

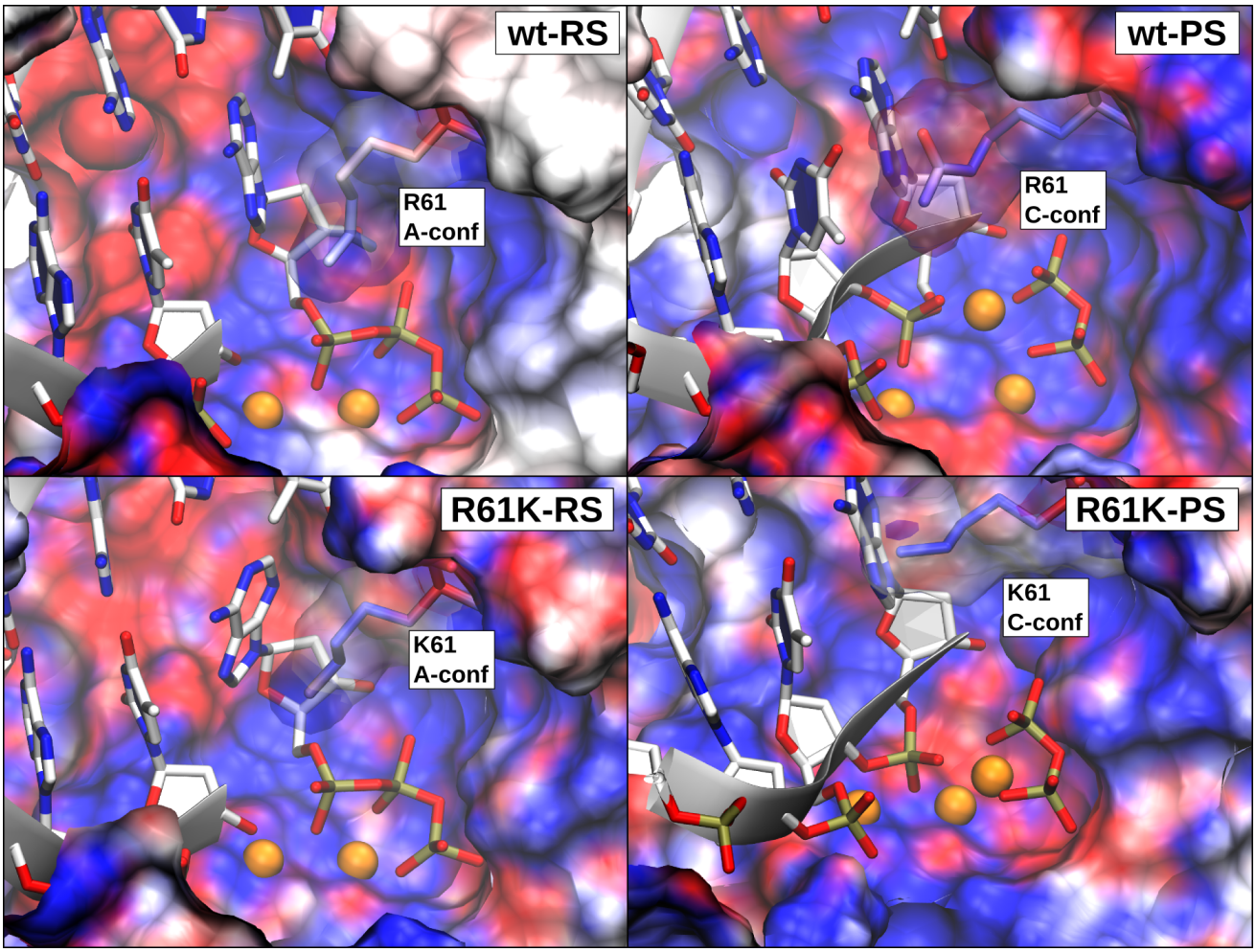
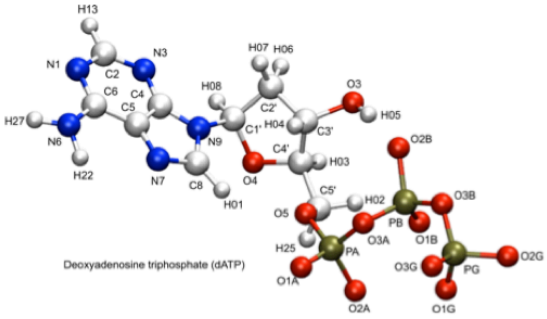
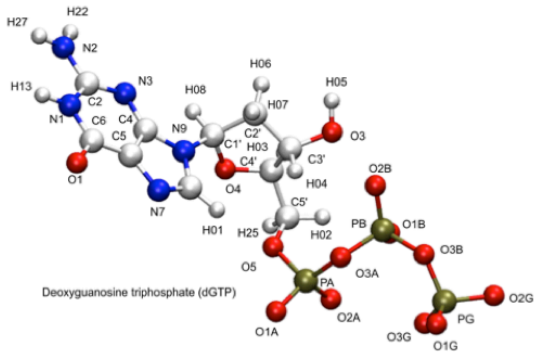


Figure S10. Electrostatic distribution of Pol-η active site in wild type and R61K mutant system. wt-RS, wild type pre-reactive state with Arg61 adopting **A-conf**. **wt-PS**, wild type post-reactive state with Arg61 adopting **C-conf**. **R61K-RS**, pre-reactive state of a system in which Arg61 was mutated to lysine. **R61K-PS**, active site electrostatic distribution of a post-reactive system affected by the mutation R61K. Nucleic acid is drawn in licorice and ribbon style. Mg^{2+} ions are depicted as orange spheres. Incoming nucleotide and PPi are represented in licorice. The lack of the third metal, in 2M-PS, determines the formation of a more negatively charged environment which serves as driving force to recruit the third metal from the bulk solution, as also reported in Fig. S8. Hydrogen atoms were removed for clarity.

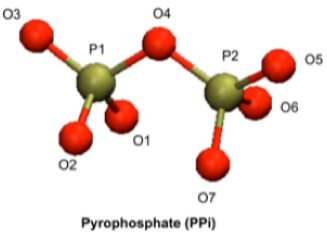
Label	Charge	Label	Charge
PA	1.630320	O3G	-0.883397
PB	1.752583	C4	0.430437
PG	1.631219	C5	-0.267981
C5'	0.145097	C6	0.685211
O5	-0.653778	N6	-1.042088
C4'	0.080100	N7	-0.584816
O4	-0.436726	C8	0.346741
C3'	0.079202	N9	-0.312571
O3	-0.651155	H01	0.060952
C2'	0.114262	H02	0.068952
C1'	0.217376	H03	0.083337
N1	-0.864491	H04	0.058165
O1A	-0.885583	H05	0.387645
O1B	-0.880119	H06	0.063558
O1G	-0.897715	H07	0.063558
C2	0.622012	H08	0.135477
O2A	-0.895529	H22	0.400770
O2B	-0.855092	H13	0.051333
O2G	-0.846348	H25	0.043781
N3	-0.797823	H27	0.406164
O3A	-0.899244		
O3B	-0.903616		



Label	Charge	Label	Charge
PA	1.174653	O3G	-0.970609
PB	1.237063	C4	0.068111
PG	1.320224	C5	0.209695
C5'	0.035607	C6	0.402502
O5	-0.604728	N2	-1.072920
C4'	0.160974	N7	-0.567726
O4	-0.443649	C8	0.205828
C3'	0.499087	N9	-0.040049
O3	-0.752877	O1	-0.665659
C2'	-0.0149334	H01	0.197091
C1'	0.165278	H02	0.116475
N1	-0.455129	H03	-0.026766
O1A	-0.794761	H04	-0.013690
O1B	-0.788383	H05	0.407593
O1G	-0.970609	H06	0.018874
C2	0.770401	H07	0.083470
O2A	-0.794761	H08	0.018874
O2B	-0.788383	H22	0.442659
O2G	-0.970609	H13	0.344874
N3	-0.582274	H25	0.116475
O3A	-0.406107	H27	0.442659
O3B	-0.659542		



Label	Charge
P1	1.493900
P2	1.493900
O1	-1.021500
O2	-1.021500
O3	-1.021500
O4	-0.855800
O5	-1.021500
O6	-1.021500
O7	-1.021500



Atom	δq	q^{eff}	q^{*}_{AMBER}
MgA-MgB	-0.25	1.75	2.00
Oδ1-2 (Asp13)	0.05	-0.74	-0.80
Oδ1-2 (Asp115)	0.07	-0.73	-0.80
Oε1-2(Glu116)	0.05	-0.78	-0.82
O (Met14)	0.06	-0.50	-0.56

Table S1. Calculated RESP charges. Upper left, dATP. **Upper center**, dGTP. **Upper right**, PPi leaving group. The QM electrostatic potential (ESP), required during the RESP fitting procedure, was calculated at the HF/6-31G* level of theory by using Gaussian 09. **Bottom**, Corrected charges for metal ions and ligands with respect to AMBER RESP charges.

System/Distance	MgA-MgB	MgA-lig	MgB-lig
wt-RS	3.57 ± 0.09 Å	1.93 ± 0.05 Å	1.90 ± 0.05 Å
mut-RS	3.59 ± 0.09 Å	1.93 ± 0.05 Å	1.89 ± 0.04 Å
mp-RS	3.57 ± 0.08 Å	1.91 ± 0.05 Å	1.92 ± 0.07 Å
2M-PS	3.68 ± 0.17 Å	1.90 ± 0.05 Å	1.90 ± 0.05 Å
3M-PS	3.65 ± 0.11 Å	1.89 ± 0.04 Å	1.94 ± 0.05 Å

System	bi-dentate A-conf	Hb1 B-Conf	Hb1 C-Conf	Hb2 C-Conf	Hb3 C-Conf
wt-RS	2.48 ± 0.16 Å	2.51 ± 0.24 Å	2.23 ± 0.19 Å	3.44 ± 0.24 Å	
mp-RS			2.62 ± 0.19 Å	2.58 ± 0.22 Å	2.68 ± 0.17 Å
3M-PS			2.64 ± 0.30 Å	2.48 ± 0.22 Å	
2M-PS	2.52 ± 0.18 Å	2.57 ± 0.21 Å	2.64 ± 0.30 Å	2.48 ± 0.22 Å	

Table S2. Key distances considered for each of the simulated system. Average value was calculated for MgA-lig and MgB-lig. In **wt-RS** the average distance between MgA and the carboxylate groups was 1.93 ± 0.05 Å. The MgA-$O_{3'\text{-OH}}$ distance was 2.12 ± 0.05 Å, whereas the MgA-$O_{\alpha\text{-phosphate}}$ and MgA-O_{H2O} distances were 2.11 ± 0.04 Å and 2.02 ± 0.05 Å, respectively. MgB coordinated the carboxylate groups of Asp13 and Asp115, the backbone oxygen atom of Met14, and the non-bridging oxygen atoms of the α-, β-, and γ-phosphate groups of the incoming nucleotide. The average distance between MgB and the carboxylate groups (Asp13 and Asp115) was 1.91 ± 0.05 Å, while that to the oxygen atoms of the phosphate groups (α- β- and γ-phosphate groups) was 2.10 ± 0.05 Å. In **mut-RS** both metals maintained their native coordination as observed in wt-RS systems. Precisely, the averaged MgA-carboxylates length was 1.93 ± 0.05 Å. The MgA-$O_{3'\text{-OH}}$ distance was 2.14 ± 0.05 Å, whereas the MgA-$O_{\alpha\text{-phosphate}}$ and MgA-O_{H2O} distances were 2.10 ± 0.05 Å and 2.00 ± 0.05 Å, respectively. In **mp-RS** the metal geometry coordination is stable maintained as in the wt-RS system. Indeed, metal-carboxylates values are very similar to those reported for wt-RS system. Both postreactive state systems share the same value for metal-ligands interaction lengths. Indeed, for both of **3M-PS** and **2M-PS**, the averaged MgA-carboxylates length is 1.89 ± 0.04 Å while that of MgB-carboxylates is 1.92 ± 0.05 Å. Instead, the inter-metal length is slightly increased with respect to the prereactive state. Here, the averaged MgA-MgB distance is 3.67 ± 0.14 Å. The table below specifies the lengths of the H-bonds established by Arg61 via its different conformation along the collected trajectories. Orange background indicates H-bonds lengths detected for A-conf, blue indicate H-bond lengths for B-conf while tan indicate H-bond lengths corresponding to C-conf.

	wt-RS	mut-RS	mp-RS	2M-PS
	Kcal mol^{-1}	*Kcal mol^{-1}*	*Kcal mol^{-1}*	*Kcal mol^{-1}*
Arg61 A-conf	-9.50 ± 1	---	-9.00 ± 1	-7.00 ± 1
Arg61 B-conf	-7.00 ± 1	---	-5.50 ± 1	-4.00 ± 1
Arg61 C-conf	-6.00 ± 1	---	-5.50 ± 1	-4.00 ± 1

Table S3. Energetics of R61 conformations. Table reporting the energy values computed for each Arg61 conformation in the simulated systems.