Article

# A strategically located Arg/Lys residue promotes correct base paring during nucleic acid biosynthesis in polymerases

Vito Genna, Paolo Carloni, and Marco De Vivo

*J. Am. Chem. Soc.*, **Just Accepted Manuscript** • DOI: 10.1021/jacs.7b12446 • Publication Date (Web): 09 Feb 2018

# A Strategically Located Arg/Lys Residue Promotes Correct Base Paring During Nucleic Acid Biosynthesis in Polymerases

Vito Genna,[1,2] Paolo Carloni[2] and Marco De Vivo[*, 1,2]

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

[2] Computational Biophysics, German Research School for Simulation Sciences, and Computational Biomedicine, Institute for Advanced Simulation IAS-5 and Institute of Neuroscience and Medicine INM-9, Forschungszentrum Jülich, 52425 Jülich, Germany

Corresponding author:

Dr. Marco De Vivo

Email: marco.devivo@iit.it

Phone: +39 01071781577

**Abstract**

Polymerases (Pols) synthesize the double-stranded nucleic acids in the Watson-Crick (W-C) conformation, which is critical for DNA and RNA functioning. Yet, the molecular basis to catalyze the W-C base pairing during Pol-mediated nucleic acids biosynthesis remains unclear. Here, through bioinformatics analyses on a large dataset of Pol/DNA structures, we first describe the conserved presence of one positively charged residue (Lys or Arg), which is similarly located near the enzymatic two-metal active site, always interacting directly with the incoming substrate (d)NTP. Incidentally, we noted that some Pol/DNA structures showing the alternative Hoogsteen base pairing were often solved with this specific residue either mutated, displaced or missing. We then used quantum and classical simulations coupled to free-energy calculations to illustrate how, in human DNA Pol-η, the conserved Arg61 favors W-C base pairing through defined interactions with the incoming nucleotide. Taken together, these structural observations and computational results suggest a structural framework in which this specific residue is critical for stabilizing the incoming (d)NTP nucleotide and base pairing during Pol-mediated nucleic acid biosynthesis. These results may benefit enzyme engineering for nucleic acid processing and encourage new drug discovery strategies to modulate Pols function.

## Introduction

More than 25 DNA and RNA polymerases (Pols) with different specializations catalyze nucleic acid extension within the cell. This is an essential process for storing, retrieving, and transmitting genetic information.[1;2] For this reason, Pols are often effective therapeutic targets to treat pathogenic infections, neurodegenerative diseases, and cancer onset.[3-5] Importantly, Pols' function allows the elongation of DNA and RNA by progressively incorporating the incoming (d)NTP nucleotide into the growing primer strand. This vital reaction leads to the formation of antiparallel nucleic acid strands, adopting the canonical Watson-Crick (W-C) base-pair conformation (G•C and A•T base pairs). The W-C base pairing generates optimal electrostatic fingerprints for protein-nucleic-acid recognition,[6] while its conformational plasticity permits nucleosome positioning[7] and epigenetic-driven structural rearrangements.[8]

Nevertheless, alternative base pairing models exist in physiological conditions. One example is the Hoogsteen (HG) base pairing, in which a purine base is flipped by 180° (3.14 rad) around the glycosidic bond (N9–C1′) through the so-called *anti*-to-*syn* transition.[9] HG geometry, however, has prevalently been observed in structures such as triple-helical or G-quadruplex nucleic acids, or has been identified as a transient base-pair conformation during certain DNA-repair mechanisms.[10;11] Notably, experiments have demonstrated that both W-C and HG base pairs exist as transient conformations in thermal equilibrium when the DNA is freely dispersed in solution, with spontaneous interconversion between these base-pair geometries.[12-14] Nonetheless, the molecular strategy by which Pols synthesizes and shapes the growing nucleic acid duplex in the W-C conformation remains unknown.

To shed light on the molecular strategy adopted by Pols, we performed bioinformatics analysis on a broad set of 28 ternary Pol/(R)DNA/(d)NTP complexes (see Fig.1, Supplementary text and Supplementary Tab. S1), which includes Pol members belonging to 5 different superfamilies. These cover all domains of life. First, we show that a key positively charged residue (Lys or Arg, hereafter referred to as "K(R)") is always spatially located close to the enzymatic two-metal active site in all the structures, thus further extending previous observations.[15] Moreover, we noted that K(R) establishes defined interactions with the incoming (d)NTP nucleotide, which have been reported to aid catalysis in Pols.[16;17] Incidentally, we also noted that K(R) is either displaced, missing or mutated in a few ternary Pol/DNA/NTP complexes in which DNA has been found in a distorted W-C conformations or even in HG base-pair model such as in the structures of low-fidelity human DNA Pol-ι[18] and the African swine fever virus Pol-X.[19] Then, quantum and classical molecular dynamics (MD) simulations, coupled to free-energy computations, were used to exemplify how the K(R) Arg61 helps shaping the canonical W-C double-helix during nucleic acid biosynthesis

operated by Y-family human DNA Pol-η.[16] Taken together, our structural observations and findings may help resolving some incongruities between W-C and HG geometries of debated Pols structures,[18;20;21] and further elucidate a unifying structural framework for correct base pairing during DNA and RNA Pols catalysis.

## Results

### A Conserved, Spatially Preserved, and Positively Charged Residue in the Active Site of Pols.

Through bioinformatics analysis performed across 28 representative X-ray structures of polymerases (Pols) in ternary Pols/(R)DNA/(d)NTP complexes, we describe a conserved, aliphatic and positively-charged amino acid (lysine or arginine, hereafter referred to as "K(R)") identically located in the first coordination-shell of the canonical two-metal-ion (2M) architecture (see Fig. 1, Supplementary text and Supplementary Tab. S1).[22;23] K(R) is invariantly positioned nearby the two catalytic $Mg^{2+}$ ions, with the latter regularly chelated by a set of conserved carboxylate groups, i.e. the DEDD-motif. This 2M-centered enzymatic structure also includes second-shell basic amino acids and monovalent cations that have been recently identified at structurally conserved positions.[24] In this respect, K(R) α-carbons are always positioned along the so-called O-Helix (or its structural homologue), which is a conserved and functional portion of the fingers subdomain involved in substrate discrimination and replication fidelity.[25;26]

Interestingly, all the analyzed DNA and RNA Pols, albeit innate differences in the overall shape, molecular size and biological role, display the K(R) α-carbon on top of the catalytic metals, at a distance ranging from ~7.5 Å to ~13 Å (see Fig. 1). The presence of this residue has already been reported for RdRp from poliovirus (PV, PDBid 1RA6),[27] the reverse transcriptase (RT) from human immunodeficiency virus type 1 (HIV, PDBid 1RTD),[28] the DdDp from bacteriophage RB69 (PDBid 1IG9),[29] and the DdRp from bacteriophage T7 (PDBid 1S76),[30] as representatives of the four Pols superfamilies.[15] For those Pols, K(R) was proposed to act as general acid during Pols catalysis, thus mediating proton-transfer events during nucleotide incorporation.[15] Here, we noted that the conserved K(R) is present in 5 different enzymatic classes, thus further extending previous observations,[15] as follows: i) *DNA-dependent DNA-Polymerases* (E.C. 2.7.7.7), ii) *DNA nucleotidylexotransferase* (E.C. 2.7.7.31), iii) *DNA-directed RNA-Polymerases* (E.C. 2.7.7.6), iv) *RNA-directed RNA-Polymerases* (E.C. 2.7.7.48) and v) *RNA-directed DNA-Polymerases* (E.C. 2.7.7.49). These enzymatic classes include, for example, the A-family bacteriophage ϕ29 DNA Pol (PDBid 2PYJ),[31] the B-family *E.coli* DNA Pol-II (PDBid 3K59),[32] the C-family *G. kaustophilus* DNA Pol-III (PDBid 3F2B)[33] and the human DNA Pol-η (PDBid 5KFZ),[34] thus covering all kingdoms of life (see Supplementary text and Supplementary Tab. S1 for a complete overview).

In addition, the short distance between K(R) α-carbon and the two catalytic metals always permits the K(R) side-chain to interact directly with the incoming substrate (d)NTP. This occurs through two main conformations of K(R) side-chain, namely: 1) the so-called "A-conf", observed in 22 X-ray structures (*vide infra*), which is characterized by the formation of a bifurcated H-bond

between K(R) and the *pro*-R oxygen of the β-phosphate of the incoming (d)NTP, and 2) the so-called "C-conf", found in 4 X-ray structures where K(R) forms one or two H-bonds with the nitrogen atoms of the (d)NTP aromatic moiety or, alternatively, remains solvent exposed (PDBids 4KLF, 4QZ9, 2E9R and 5TYD).[35-38] Notably, in the remaining 2 structures K(R) sidechain has not been solved (PDBids 4U6P and 2AQ4).[39;40]

**K(R) is critical for incoming (d)NTP nucleotide binding and correct base pairing**

Intrigued by this structural observation, we extended our analysis to consider the interaction geometry between K(R) and the surrounding structural environment. We hypothesized that K(R) may assist the formation of correct base pairs in Pols through dNTP stabilization, i.e. via the K(R)-promoted and dNTP-mediated base-pair formation (see Fig. 2). On this basis, we analyzed the base-pairing geometries formed between the incoming (d)NTP and its cognate nucleotide in all ternary Pols/(R)DNA/(d)NTP complexes forming our dataset (see Supplementary Tab. S1 and Methods section). Among them, we noted few ternary Pol/DNA/NTP complexes in which the DNA double-helix is distorted or in the alternative and unusual Hoogsteen (HG) base-pairing architecture rather than the expected Watson-Crick (W-C) geometry.[18;19] Contextually, most of these structures have K(R) either mutated, missing or slightly displaced (see more below). Thus, our bioinformatics and structural analyses suggest K(R) as a key player in promoting the correct (d)NTP:(d)N base pairing during Pol-mediated nucleic acid biosynthesis.

To further examine this hypothesis, we first considered the torsion angles describing the rotation of the N-glycosidic bond of the incoming nucleotide dNTP and its 5'-templating base in all ternary Pol/(R)DNA/(d)NTP structures. Precisely, we refer to $\omega_{Pu}$ to identify the N-glycosidic bond of purines (i.e. A and G, atoms O4'-C1'-N9-C8) in the nascent 3'-base pair, whereas $\omega_{Py}$ identifies the N-glycosidic bond of the templating pyrimidines (i.e. C and T, atoms O4'-C1'-N1-C6) (see Fig. 3B and Methods section). By analyzing $\omega_{Pu}$ and $\omega_{Py}$ values in the Pol/(R)DNA/(d)NTP X-ray structures in our dataset, we observed that the correct W-C base-pairing shows ~0.1 < $\omega_{Pu}$ < ~0.7 and ~0.6 rad < $\omega_{Py}$ < ~1.1 rad (see Fig. 3B and Supplementary Tab. S2). On this basis, we further divided our dataset into those structures where K(R) adopts A-conf and those where K(R) adopts C-conf. When K(R) adopts A-conf it forms a bifurcated H-bond with the phosphate tail of the incoming (d)NTP and a π-cation interaction with its aromatic ring(s). In this subset, our analysis returned an averaged $\omega_{Pu}$ and $\omega_{Py}$ values respectively of ~0.6 and ~1 rad thereby indicating a well-structured W-C base-pairing mode (see Supplementary Tab. S2). In contrast, in those structures where K(R) adopts C-conf – featured by the formation of one/two H-bond with the

incoming nucleobase and/or with its templating nucleotide – the nascent base-pair geometry at the growing 3'-terminal deviates from the iconic W-C arrangement. A rotation of $\omega_{Pu}$ and $\omega_{Py}$ values, which are ~1.5 and ~0.5 rad, respectively, suggests a destabilization of the forming nucleic acid double-helix. Interestingly, a similar rotation has been observed, for example, in *human* DNA Pol-η bound to dsDNA during a functionally programmed misincorporation of dGTP opposite a templating dT (PDBid 4J9K).[41] Here, K(R) (i.e. Arg61) surprisingly adopts C-conf. The Arg61 guanidine group is rotated and H-bonds the O4 atom of the templating dT. Concomitantly, Arg61 interacts with the 3'end nucleobase via an additional H-bond established with the O6 atom of the dGTP substrate. As a consequence, the 3'-terminal nucleotides get distorted, as shown by the $\omega_{Pu}$ = ~0.9 rad, which is ~0.2 rad higher with respect to its value in the stable W-C base-pair geometry.

We also observed similar distortions in the nascent 3'-terminal base-pair solved in the active site of the *prokaryotic* Dpo4 enzyme operating 7,8-dihydro-8-oxodeoxyguanosine (PDBid 2C2R)[42] translesion synthesis, one of the most common DNA lesions resulting from reactive oxygen species (ROS) metabolism.[43] Here, K(R) (i.e. K56) assumes an unusual solvent-exposed conformation that prevents its interaction with the incoming dNTP. Again, the nascent 3'terminal base-pair architecture is wobble and this is well shown by $\omega_{Pu}$ and $\omega_{Py}$ values of ~1.4 and ~0.8 *rad,* thus sensibly deviating from the $\omega_{Pu}$ and $\omega_{Py}$ values of the W-C geometry (i.e. ~0.6 and ~1.1, respectively). This is likely because of the damaged templating 7,8-dihydro-8-oxodeoxyguanosine or the unusual backbone distortion at the 3'-end of the primer strand.[44]

Even more severe structural alterations were observed in the forming 3'-terminal base-pair geometry solved in the active site of the ternary complex formed by the Y-family human DNA Pol-ι, and the error-prone DNA Pol-X from the deadly African swine fever virus (PDBid 1T3N and 2M2W, respectively).[18;19] Importantly, both complexes were obtained in the absence of K(R). Specifically, K(R) (i.e. K77) is mutated into a leucine in the ternary Pol-ι/DNA/dNTP complex, while K(R) is replaced by a valine (i.e. V120) in the ternary Pol-X/DNA/dGTP complex. In particular, human DNA Pol-ι has been crystallized in both wild-type form, where K(R) is K77 (PDBid 1ZET),[21] and K77L mutant form (PDBid 1T3N).[18] As expected, the wild-type form, solved in the prereactive state, shows the conserved K(R), K77, adopting the so-called A-conf that is mainly described by the presence of H-bonds interactions with the β-phosphate of the incoming dTTP. These interactions stabilizes the newly formed dTTP:dA W-C base-pair geometry, as demonstrated by the $\omega_{Pu}$ and $\omega_{Py}$ values of ~0.7 and ~0.7 rad that match the values describing a proper W-C geometry. Moreover, the stability of this nascent base-pair is also confirmed by the interbase H-bond pattern, which depicts an averaged donor-acceptor length of ~3 Å.

In contrast, in Pol-ι K77L mutant form, the incoming dTTP is found H-bonding its cognate nucleotide through HG base pairing, rather than the expected W-C base architecture. Here, the templating dA purine is rotated by 3.14 rad ($\omega_{Pu}$ = ~ –2.4 *rad*) with respect to its conformation in the wild-type complex. Consequently, dA exposes the so-called "Hoogsteen-edge", which features the HG geometry formed through two H-bonds with the incoming dTTP[10] (~3.2 Å, measured in between the $O4_{dTTP}$:$N6_{dA}$ and $N3_{dTTP}$:$N7_{dA}$ atoms). We stress again that the single mutation K77L is the only difference between the wild-type active site and the mutant isoform (see Supplementary Fig. S2).

More recent wild-type crystal structures of Pol-ι show the HG base-pair in presence of K77 (e.g., PDBid 2ALZ and 5KT2).[45-47] In some of these structures, K77 is slightly displaced so to interact with the γ-phosphate of the incoming dNTP, which suggests a lower K(R)-mediated stabilization of the incoming nucleotide. Nevertheless, these wild-type structures indicate a complex network of interactions for the regulation of base paring in Pol-ι. Notably, in all these Pol-ι structures, the HG edge is formed by the rotation of the template base, which is more distant from K(R). The incoming nucleotide is aligned for catalysis, here likely stabilized by the narrow catalytic site of Pol-ι, which preferentially accommodates Mn (over Mg, as in most Pols) as the metal ion cofactor.[48;49]

**K(R) and the Nascent Nucleic Acid Strand in *human* DNA Pol-η.**

Intriguingly, it was proposed that Arg61 may somehow prevent Hoogsteen base pairing in *human* DNA Pol-η,[16] a trans-lesion Pol that shares all the common structural determinants of DNA and RNA Pols.[50;51] Thus, Arg61 would help correct base pairing in addition to its established role in assisting the catalytic steps in Pol-η.[17;50;52] However, Arg61 mutants have been actually solved with DNA in W-C geometry,[34] which clearly indicates that Arg61 is not the only factor at play. To test the mechanistic role of K(R) in promoting correct base-pair formation via dNTP stabilization, here we performed extensive classical and first-principle molecular dynamics simulations coupled with metadynamics-based free-energy calculations of *human* DNA Pol-η in its wild-type and Arg61Ala mutant form.

Through two independent classical MD simulations (~500 ns each) of the wild-type ternary Pol-η/dsDNA/dATP prereactive complex, we examined the evolution over time of two torsional angles, $\omega_{Pu}$ •(i.e. O4'-C1'-N9-C8 of the incoming dATP, purine) and •$\omega_{Py}$ (i.e. O4'-C1'-N1-C6 of the templating dT, pyrimidine), that well describe the rotation of the N-glycosidic bonds of the nascent base-pairing and the consequent stability of the Watson-Crick (W-C) base-pairing geometry (see

Fig. 3). During the entire MD simulations Arg61 maintained its native A-conf while $\omega_{Pu}$ and $\omega_{Py}$ shown a remarkable stability with an averaged values of 0.62 ± 0.29 rad and 0.95 ± 0.29 rad (see Fig. 3), thus well matching the X-ray values of (0.61 and 1.3 rad, respectively).[52] As a consequence, the W-C geometry is well structured and the two characteristics H-bond typical of A:T pair show an averaged value 3.11 ± 0.25 Å. Different results were obtained through two independent simulations (~500 ns each) of the prereactive Arg61Ala-mutant ternary Pol-η/dsDNA/dATP complex. Here, the torsional $\omega_{Pu}$ and $\omega_{Py}$ angles show a value respectively of 0.76 ± 0.27 and 1.15 ± 0.25 rad, indicating a less stable W-C geometry generated by the incoming dATP and its cognate dT (see Fig. 3). These computations, coupled with previous findings,[51] further indicate K(R) as one of the key structural determinants in shaping the correct base-pair geometry in the nascent nucleic acid duplex.

Then, we used well-tempered metadynamics[53] to energetically characterize the W-C stabilization in the wild-type and the Arg61Ala mutant form.[14] Toward this aim, we reconstructed the free-energy surface (FES) as a function of two distinct collective variables (CVs) that well discriminate between W-C and alternatives base-pairing modes (such as HG). Specifically, these CVs – represented by torsion angles $\omega_{Pu}$ and $\omega_{Py}$ (see Fig. 3) – respectively describe the N-glycosidic bond of dATP (O4'-C1'-N9-C8) and that of its templating dT (O4'-C1'-N1-C6), allowing therefore the exploration of the conformational space of nascent base-pair. Notably, $\omega_{Pu}$ and $\omega_{Py}$, adopt a value of ~0.6 and ~1.1 rad when the nucleobases assume the *syn* conformation (i.e. W-C pairing).

The calculated FES for the K(R)-containing ternary Pol-η/dsDNA/dATP complex shows two distinct minima representing the two alternative W-C and HG base-pairing geometries (see Fig. 4). The most stable minimum is defined by an energetic basin of ~ -20 kcal/mol and is located at ~0.1 < $\omega_{Pu}$ < ~0.8 and ~1 rad < $\omega_{Py}$ < ~1 rad. Here, we found an ensemble of dATP:dT conformations depicting the W-C architecture with their typical interbase H-bonds of ~3 Å. The W-C geometry is here further stabilized by the Arg61 side chain, which through its A-conf establishes π-stacking interactions with the dATP aromatic moiety while concomitantly coordinates the phosphate groups of the incoming dATP. These conformations resemble the crystallographic pre-reactive Michaelis-Menten complex[50] (RMSD = ~3.5 Å versus PDBid 4ECS).

To further test the stability of the W-C geometry in Pol-η, we evaluated the free energy barrier to convert the W-C geometry into the HG base pair. This transition returned a barrier of ~10 kcal/mol that is characterized by the so-called Hoogsteen-edge, in which the dATP H-bonds its templating counter-base. In this way, the system falls in the second minima located at ~-1.5 rad < $\omega_{Pu}$ < ~-3 rad and ~-1 rad < $\omega_{Py}$ < ~2 rad (see Fig. 4) where the characteristic H-bond pattern typical of the HG geometry is generated (i.e. $O4_{dTTP}$:$N6_{dA}$ and $N3_{dTTP}$:$N7_{dA}$), as confirmed by the H-bond averaged length of 3.1 Å. The Arg61 side chain is now rotated into C-conf, forming stable H-bond

interactions with the O6 atom of the templating dT and with the N6 atom of the aromatic base of the incoming dATP. Notably, the computed energetic barrier for reconstructing the W-C base-pair geometry is only ~2.5 kcal/mol, confirming that the W-C base-pair conformation is more favored.

For the ternary Pol-η/dsDNA/dATP complex that lacks K(R) (i.e. a ternary complex with the Arg61Ala mutant form of Pol-η),[16] the free-energy landscape again shows two distinct minima that identify W-C and HG base-pairing geometries. This time, however, the deepest energetic basins, located at ~-1.5 rad < $\omega_{Pu}$ < ~-3 rad and ~-1 rad < $\omega_{Py}$ < ~1 rad for W-C and HG base-pairing geometries are almost isoenergetic, with an energy difference of only ~2 kcal/mol in favor of the HG base pairing (see Fig. 4). Interestingly, the possibility of HG base-pairing geometries was suggested by Ling et al. in their report on the wild-type structure of Dpo4, an archaeal Pol-η homologue.[54]

For comparison, we also evaluated the W-C ↔ HG interconversion energy in the same DNA duplex freely dispersed in solution. Here, using the same enhanced sampling approach that returned a mechanism for W-C ↔ HG interconversion rather similar to that observed in Pol-η, the two base-pairing geometries are isoenergetic, suggesting an equilibrium ensemble of structures that can equally interconvert between W-C and HG models (see Supplementary Fig. S3). This is in good agreement with NMR studies, which recently indicated a thermal equilibrium between W-C and HG geometries in the DNA duplex freely dispersed in solution.[12;14]

Finally, Car-Parrinello quantum-mechanics/molecular-mechanics (QM/MM) simulations of both Arg61Ala and wild-type forms (with Arg61 adopting either A- or C-conf) of human Pol-η were used to investigate the direct participation of K(R) as a general base in the chemical step for nucleotide incorporation (see Methods). We found that, when Arg61 adopts the A-conf, the catalytic reaction has a barrier of ~12 kcal/mol, with Arg61 that helps catalysis through its electrostatic contributions. This reaction barrier is ~4 kcal/mol lower than the one determined for the other systems (see Supplementary Fig. S4).[55;56]

## Discussion

Polymerases (Pols) synthesize double-helix nucleic acids by progressively elongating a primer strand through a sequence of nucleotide [(d)NTP] incorporation events. Importantly, Pols catalyze nucleic acid biosynthesis in the Watson-Crick (W-C) base pairing, although they can be significantly different in size, shape, and specific biological role. Yet, the exact molecular basis used by Pols to shape nucleic acid duplexes remains unclear.

Here, through structural and bioinformatics studies, we first describe the conserved and positively charged residue (a lysine or arginine, hereafter referred to as "K(R)"), located within the two-metal-aided (2M) active site in a broad and multispecies set of 28 representatives ternary Pols/(R)DNA/(d)NTP (see Fig.1, Supplementary text and Supplementary Tab. S1). Notably, this result sensibly expands previous analyses on the conservation of this specific residue.[15] In particular, we found that K(R) α-carbons preserve a similar spatial localization on 'top' of the catalytic metals, at a distance ranging from ~8.5 Å to ~12.7 Å (see Fig. 1 and Fig. 2). This is illustrated well by the structural comparison between two phylogenetically distant enzymes – i.e. *T. thermophilus* RNA Pol-I (PDBid 2CW0)[57] and *Human* DNA Pol-η (PDBid 4ECS).[52] These Pols exhibit striking 3D structural divergences. The RNA Pol-I is formed by five subunits (i.e. $\beta'\beta\alpha^{I}\alpha^{II}\omega$), ~3361 amino acids, and its mass is ~375 kDa.[58] The DNA Pol-η comprises a single monomer of ~450 residues and its molecular weight is ~78 kDa.[59] Nonetheless, K(R) is in a similar location in the active site of both these Pols structures. In detail, K(R) identifies Arg1239 in *T. thermophilus* RNA Pol I and Arg61 *Human* DNA Pol-η, respectively. Importantly, mutagenesis studies have shown reduced catalytic activity in both Pols when K(R) was either mutated or missing (see Supplementary Tab. S1).[16;60] Hence, these observations suggest a convergent evolution of Pols to strategically locate K(R) in their active site, likely because of its functional role for nucleotide incorporation during nucleic acid extension.[17;50;52]

Incidentally, we noted here that the nascent Watson-Crick (W-C) base-pairs were often found wobble or adopting the uncommon HG conformation when K(R) was found displaced, missing or mutated. This is clear in the prereactive complex of Dpo4 operating 7,8-dihydro-8-oxodeoxyguanosine (PDBid 2C2R)[61] translesion synthesis, where K(R) (i.e. Lys56) unexpectedly assumes an unusual solvent-exposed conformation that prevents any interaction with the incoming dNTP. In this complex, the incoming dCTP is slightly distorted, as shown by the partial rotation of the base with respect to the sugar moiety (i.e. dihedral angle taken along the N-glycosidic bonds $\omega_{Pu}$ and $\omega_{Py}$) of ~1.4 rad and ~0.8 rad, sensibly deviating from the W-C values (~0.6 rad and 1.1 rad, respectively. See Results section).

In two additional Pols X-ray structures where K(R) is missing or has been mutated for experimental needs, we observed more severe distortions in the nascent 3'-terminal base-pair geometry. One example is the Lys77Leu mutant form of *human* DNA Pol-ι crystallized in complex with dsDNA and incoming dNTP (PDBid 1T3N).[18] Surprisingly, this structure revealed the dTTP:dA pairing geometry at the 3'terminal base-pair adopting the HG geometry. The templating dA is rotated of 3.14 rad ($\omega_{Pu}$= ~-2.4 rad while $\omega_{Py}$= ~0.8 rad), sensibly deviating from the $\omega_{Pu}$ and $\omega_{Py}$ values typical of W-C base pairing (i.e. ~0.6 rad and ~1.1 rad for $\omega_{Pu}$ and $\omega_{Py}$, respectively).

However, a second X-ray structure of human DNA Pol-ι depicts the wild-type form of the ternary Pol-ι/DNA/dTTP (PDBid 1ZET)[21] with the nucleic acids in the expected W-C base-pairing geometry.[21] In this case, the incoming dTTP is perfectly aligned with the templating dA and, accordingly, displays the proper W-C $\omega_{Pu}$ and $\omega_{Py}$ values and H-bond interactions. Indeed, in contrast to the Lys77Leu mutant form, the 3'-terminal nucleotide here shows $\omega_{Pu}$ and $\omega_{Py}$ values of ~0.7 rad, which fall within the appropriate range of values of a well-structured W-C base-pairing geometry. Importantly, the K(R) mutation is the only difference between these two Pol-ι crystals (see Supplementary Fig. S2). Recent wild-type X-ray structures of human Pol-ι, however, captured the HG base-pair in presence of K77, although here this flexible residue is slightly displaced so to interact more closely with the γ-, and not the α-/β-phosphate of the incoming dNTP.[45;46] Anyhow, these structures clearly denote a complex network of interactions to regulate base paring in Pol-ι, which necessarily involves other structural factors in addition to K(R). In this respect, it is worth noting that all these structures with HG base-pairing in Pol-ι are formed through the rotation of the template base, opposite the incoming nucleotide. Indeed, Pol-ι is known to have a restricted active-site, which differs from the larger and solvent-accessible active site of other Pols, which allow binding and bypass of bulky DNA lesions, like for UV-induced cyclobutane thymine dimers (T-T) processed by Pol-η.[62;63] Instead, the narrow catalytic site of Pol-ι, and its preference for Mn (*vs* Mg)[48;49] at the catalytic site, are suggested to be critical factors for the kinetics and fidelity of polymerase catalysis.[64] These distinctive structural features are likely part of the complex network of interactions that involve K(R) for incoming nucleotide binding and base-pair stabilization during catalysis in Pol-ι.

The K(R) residue is also missing from the error-prone DNA Pol-X of the deadly African swine fever virus (PDBid 2M2W),[19] a mutagenic DNA polymerase which introduces dG:dGTP mismatch through non-Watson-Crick incorporation.[19] Notably, here Val120, which has been recently suggested to be important for misincorporation activity of Pol-X,[65] replaces K(R). In the Pol-X active site, the γ-phosphate group of the incoming dGTP forms a single H-bond interaction with Asn48. The nascent base-pair here adopts the HG geometry favored by two purines (the susceptible nucleobases prone to WC to HG transition). Indeed, while the templating dG adopts the *anti* conformation (with $\omega_{Pu1}$ = ~1.3 rad), the incoming nucleotide dGTP is rotated ~3.1 rad, as shown by $\omega_{Pu2}$ value of ~-2.3 rad. It remains to be rationalized how these observations and mechanistic hypotheses may relate to lower fidelity upon K(R) mutation in Pols.[66] It is tempting to hypothesize that the lack of K(R) may represent an evolutionary strategy to produce mutagenic Pols essential for life.[67;68] For example, lack of K(R) may lower Pols fidelity to favor recruiting wrong nucleotides for

generating the functional mismatches (e.g. somatic hypermutation process).[41;69;70] This has been observed in the X-family member DNA Pol-β where R183A mutation drastically reduces enzyme's processivity,[71] dNTP affinity (decreased up to ~12 fold)[71] and pyrophosphate release, which concludes the catalytic cycle.[72] Lastly, the lack of K(R) may permit the proper accommodation of the mismatched or hybrid base-pair,[16;73;74] by enlarging Pols' active site.

Then, we tested our mechanistic hypothesis on the role of K(R) in promoting W-C formation via (d)NTP-stabilization in Pols catalysis through a series of classical and first-principle molecular dynamics simulations (MD) coupled with free-energy calculations. We considered *human* DNA Pol-η , which is experimentally well-characterized.[16;17;70;75;76] In *human* DNA Pol-η, Arg61 is suggested to somehow prevent Hoogsteen base pairing,[16] in addition to its role for enzyme efficiency and fidelity.[16;17;50] Our simulations of the wild-type Pol-η confirms that Arg61 forms defined interactions with the incoming nucleotide dNTP, favoring W-C base pairing. However, when K(R) is mutated (i.e. Arg61Ala), the relative stability of the global minima is comparable, with the HG base pairing model that becomes competitive to the W-C geometry (Fig. 4). Notably, this situation resembles the conformational equilibria of the DNA freely dispersed in solution, in which W-C and HG geometry were computed to be isoenergetic, in agreement with NMR studies.[14]

The recurrent presence of a positively charged residue located in a similar area close to the 2M catalytic site has been noted in four Pols superfamilies (see Results). It has been proposed that it may directly participate as a general base in the chemical step for nucleotide incorporation.[77] In this respect, our Car-Parrinello QM/MM simulations show that the lowest catalytic pathway occurs with a barrier of ~12 kcal/mol (see Supplementary Fig. S6) when Arg61 adopts the A-conf. In this conformation, this residue aids catalysis through its electrostatic contributions, with no active role in proton transfer events in the chemical step for nucleotide addition (see Supplementary Fig. S4).[15;55;56;78]

In summary, we propose that K(R) contributes in shaping the base pairing of the growing DNA during Pol catalysis. This is based on our structural observations, bioinformatics analyses, and simulations, which show (i) how the K(R) residue is conserved and spatially preserved in eukaryotic, prokaryotic, and viral Pols, and (ii) how the K(R) residue establishes defined interactions with the incoming (d)NTP nucleotide promoting its stability, which in turns may assist correct base pairing of the growing nucleic acid strand during Pols catalysis.

## Conclusions

Our results suggest that the conserved and positively charged residue (Lys or Arg), strategically located near the enzymatic two-metal active site in a broad dataset of polymerase enzymes, helps promoting correct base pairing during nucleic acids biosynthesis. We illustrated the role of this specific residue in human DNA Pol-η, where we found that the conserved Arg61 favors W-C base pairing, relative to Hoogsteen, through defined interactions with the incoming nucleotide. Although K(R) is clearly not the only element at play in shaping the base-pairing in Pols, our results indicate that this residue is critical for dNTP binding and stabilization, which in turn favors correct base-pairing of the nascent nucleic acid strand. Taken together, our findings describe and further clarify a unifying structural framework for Pols catalysis in which K(R) promotes correct base pairing during nucleic acid biosynthesis. These findings may have direct implications for enzyme engineering to process nucleic acids and drug discovery.

## Methods

### Bioinformatics and Structural Models

For the bioinformatics and structural analyses, we considered the available X-ray structures deposited in the Protein Data Bank of ternary complexes of wild-type-Pols bound to undamaged (R)DNA and the (d)NTP nucleotide, captured forming a prereactive state with an integer two-metal-ion active site. These comprise five enzymatic classes, for a total of 28 different and representatives structures chosen among a total of 3,307 analyzed complexes, namely: i) *DNA-dependent DNA-Polymerases* (E.C. 2.7.7.7), ii) *DNA nucleotidylexotransferase* (E.C. 2.7.7.31), iii) *DNA-directed RNA-Polymerases* (E.C. 2.7.7.6), iv) *RNA-directed RNA-Polymerases* (E.C. 2.7.7.48) and v) *RNA-directed DNA-Polymerases* (E.C. 2.7.7.49). In all these structures, K(R) was found to be similarly located nearby the catalytic site. In Figure 1, we created a subset of 28 complexes formed by those with the best resolution, per each family member, covering all kingdoms of life (see Fig. 1, Supplementary text and Tab. S1 for further details).

### Classical Molecular Dynamics Simulations

The ternary Pol-η/DNA/dATP complex used for MD simulations is based on the crystallographic structure of the enzyme before completion of the nucleotidyl-transfer reaction and consequent formation of products (PDBid 4ECS).[52] To verify our mechanistic hypothesis, we first performed classical MD simulations on three systems. The first system was the wild-type ternary

Pol-η/DNA/dATP complex where K(R) is Arg61. We then investigated the Arg61Ala mutant form of the same complex, aided by the wealth of mutagenesis and kinetics data on this specific mutation.[16] We also performed MD simulations on a dsDNA freely dispersed in solution. The dsDNA was extracted from the ternary Pol-η/DNA/dATP complex. The all-atom AMBER/parm99SB-ILDN[79] force field and parmbsc0 corrections[80] were adopted for the Pol-η in complex with dsDNA, whereas dATP was treated with the general AMBER force field.[81] The atomic charges were derived by fitting the electrostatic potential according to the Merz-Singh-Kollman scheme, also known as the RESP fitting procedure (see Supplementary Tab. S3). The length of all covalent bonds, including hydrogen atoms, was set using the LINCS algorithm, allowing a time-integration step of 2 fs. All simulations were performed using Gromacs 4.6.5 code.[82] Long-range electrostatic interactions were calculated with the particle mesh Ewald method with a real space cut-off of 12 Å. Periodic boundary conditions in the three directions of Cartesian space were applied. Constant temperature (310 K) was imposed using Langevin dynamics[83] with a damping coefficient of 1 ps. A constant pressure of 1 atm was maintained with Langevin-Piston dynamics with a decay period of 200 fs and a time constant of 50 fs. The metal active site was treated with a flexible non-bonded approach based on the "atoms in molecules" partitioning scheme of the DFT-BLYP electronic density of the active site (see Supplementary Tab. S3).[84] We could thus consider the charge-transfer interactions between $Mg^{2+}$ ions and their ligands, permitting possible structural rearrangements at the active site during the MD simulations. All the simulated systems were hydrated using TIP3P[85] water molecules. A total of 7 $Mg^{2+}$ ions were added to each system to reach a final concentration of ~1 mM, while $Na^{+}$ and $Cl^{-}$ ions were added to neutralize the total charge. The size of the final systems was approximately 116 × 94 × 92 Å, with ~12 000 water molecules and a total of ~102 000 atoms for both wild-type and Arg61Ala-mutant form of Pol-η/DNA/dATP complex. It was 71 × 56 × 62 Å, with ~6 000 water molecules and a total of ~23 000 atoms for the system in which the dsDNA is freely dispersed in solution. These equilibrated systems (~150 ns of simulations time, in total) were propaedeutic for the enhanced sampling simulations.

**Free-Energy Metadynamics Simulations**

We used well-tempered metadynamics[53] to estimate the energy for the WC ↔ HG interconversion in the presence/absence of K(R). Calculations were performed using the Gromacs 4.6.5[82] package as a molecular engine coupled to Plumed code[86] to enhance the sampling of our simulations (see Supplementary Text for further detail). The free energy was determined as a

function of two selected collective variable (CVs) that discriminate between W-C and HG base-pairing geometries. Specifically, these CVs are the angles $\omega_{Pu}$ and $\omega_{Py}$ (see Fig. 3), which describe the same dihedral angle formed by the atoms O4'-C1'-N9-C8 (i.e. N-glycosidic bond) in purines and atoms O4'-C1'-N1-C6 in pyrimidines in the nascent 3'-terminal base pair. While both $\omega_{Pu}$ and $\omega_{Py}$ show a value of ~0.6 and ~1.1 rad when the nucleobases adopt the *syn* conformation (i.e. W-C pairing) mainly $\omega_{Pu}$ changes to ~-2.5 rad when W-C ↔ HG interconversion takes place. The fictitious temperature associated with the CVs was set to 1550 K while the Gaussian function deposition rate was set to 1 ps. The initial hills height and width were set to 0.05 kcal $mol^{-1}$ and 0.01 kcal $mol^{-1}$, respectively. The well-tempered simulations were carried out until their convergence (~350 ns), i.e. the progressive stabilization of the energetic minima on the free-energy surface (see Supplementary Figs. S3, S5 and S6). All other parameters correspond to those used for the plain MD simulations, described above.

**Car-Parrinello Quantum Mechanics/Molecular Mechanics Simulations**

Car-Parrinello (CP) simulations,[87] in the quantum mechanics/molecular mechanics (QMMM) implementation,[88;89] were performed to dissect the contribution of K(R) (i.e. Arg61) during Pol-η catalysis. In particular, we ran CP QM/MM simulations on three distinct Pol-η/DNA/dATP complexes. In the first complex, Arg61 adopted the so-called A-conf. In the second complex, Arg61 was in C-conf. In the third complex, Arg61 was mutated into an alanine according to experimental data. The nucleophile (i.e. 3'-$O^-$) initiating reaction is considered as ready to react (i.e. deprotonated)[51] and is ~3.6 Å from the $P^{\alpha}$, in agreement with crystallographic data (PDBid 4ECS, 3.3 Å).[52] The reactive region of the ternary complex was treated at the DFT/BLYP level and includes the $Mg^{2+}$ coordination sphere (DED motif: D13, E115, D118, M14), part of the DNA dATP:dT, $dT_{-1}$ nucleotides, Arg61, and solvation water molecules (for a total of 183 QM atoms, see Supplementary Figure S4). The remaining part of the complex (~101 820 atoms) was treated using the AMBER force field.[79] The valence electrons are described by a plane wave basis set up to a cut-off of 70 Ry. A 20 × 20 × 18 $Å^3$ cell includes the QM part of the system. The interactions between valence electrons and ionic cores are described with norm-conserving Martins–Troullier pseudopotentials.[90] CP QM/MM dynamics is carried out with a time step of 0.12 fs (for a total simulation time of ~250 ps, including plain and constrained QM/MM simulations) and a fictitious electron mass of 500 au. Constant temperature simulations are achieved by coupling the system with a Nose′–Hoover thermostat at 500 $cm^{-1}$ frequency.[91;92] The interactions between the MM and QM regions are coupled in a Hamiltonian scheme as discussed by Laio et al.[88] Notably, a rigorous Hamiltonian treatment of the electrostatic interaction between QM and MM regions is used, as in

ref. 74. The approach has been shown to accurately describe a variety of metal-dependent enzymes and, specifically, protein–DNA complexes.[89;93-98]

The QM/MM protocol includes an initial equilibration of the configuration produced by MD simulations, followed by a run where only the MM part is free to move, while the QM part is kept frozen (~2 ps). Then, the whole system is allowed to move and heat up to 300K (~2 ps). Trajectories are then collected for analysis. Configurations from the equilibrated QM/MM simulations are used for free-energy calculations.[50] The $S_N2$-type phosphoryl-transfer reaction is described with a reaction coordinate (RC) defined as the difference between the length of the forming bond (-3'O—$P^{\alpha}$) and that of the breaking bond ($P^{\alpha}$—$O_b$ where b stands for bridging). This RC is well-suited for $S_N2$-like reactions. "Blue-Moon" ensemble simulations are carried out adiabatically, constraining the RC, while leaving all other degrees of freedom free to evolve. The free-energy surface (FES) of the reaction is obtained by thermodynamic integration.[99] The pathway from reactants to products is divided into nine steps, with a resolution of 0.2 Å. This has been shown to lead to a reasonable compromise between accuracy and CPU-time consumption (additional steps with 0.1 Å resolution are considered in the vicinity of the intermediate states). Each step is simulated for at least 5 ps, or until the force on the constraint is equilibrated (i.e. the running averages over 1 ps windows varies less than 5%). The free-energy profile is then obtained by integrating the force profile, leading to the FES description. The error associated with the critical points of FES is calculated by propagating the error on forces at every step, using the propagation of error formula for linear functions. The free-energy values should be considered approximate. This is because of the still-limited sampling accessible to first-principles-based DFT calculations, the choice of a 1-D RC, and the limitations of current GGA XC functionals.

**Acknowledgements**

M.D.V. thanks the Italian Association for Cancer Research (AIRC) for financial support ['MFAG n. 18883']. P.C. acknowledges financial support from the Human Brain Project www.humanbrainproject.eu and BioExcel (https://bioexcel.eu/) EU projects. We thank Grace Fox for proofreading and copyediting the manuscript.

**Keywords:** Nucleic Acid Polymerization • Polymerase • DNA conformation and synthesis • Molecular simulations and QM/MM • Metalloenzyme

**Supporting Information**:

The Supporting Information is available free of charge on the ACS Publications website.

Details on the PDB analysis and sequence alignments, classical metadynamics simulations, and ab initio Car−Parrinello Molecular Dynamics in Supplementary, Figures S1− S6, and Supplementary Table 1-3 (PDF ).

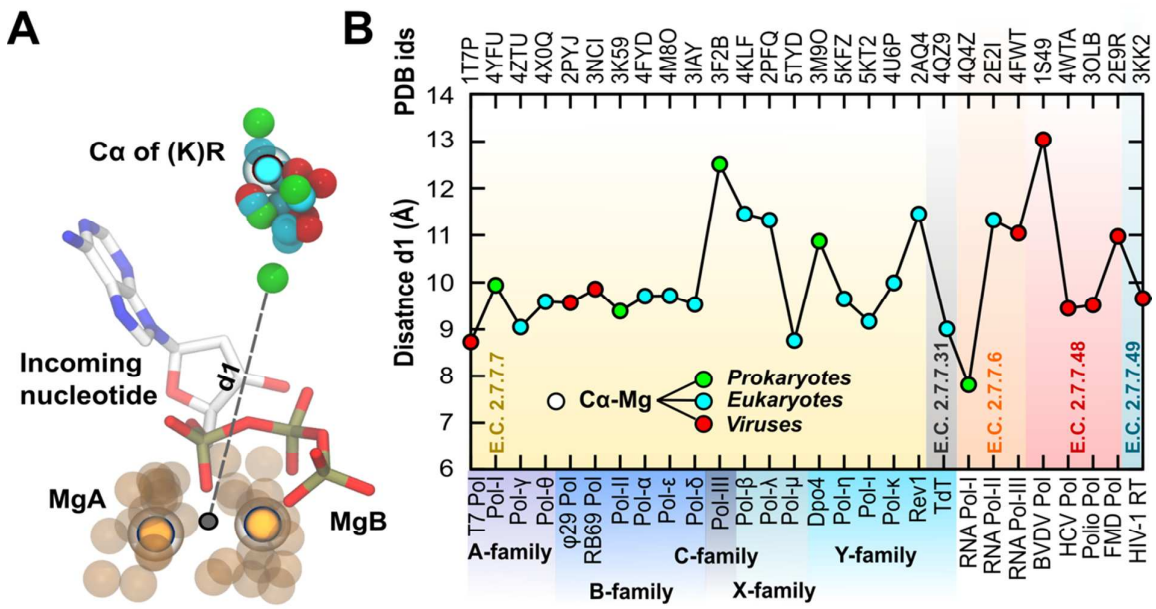


**Figure 1.** Graph reporting the spatial localization of **K(R)** α-carbons (**A**) and their distance from the two catalytic metals measured in all the currently available ternary Pol/(R)DNA/(d)NTP complexes (**B**).

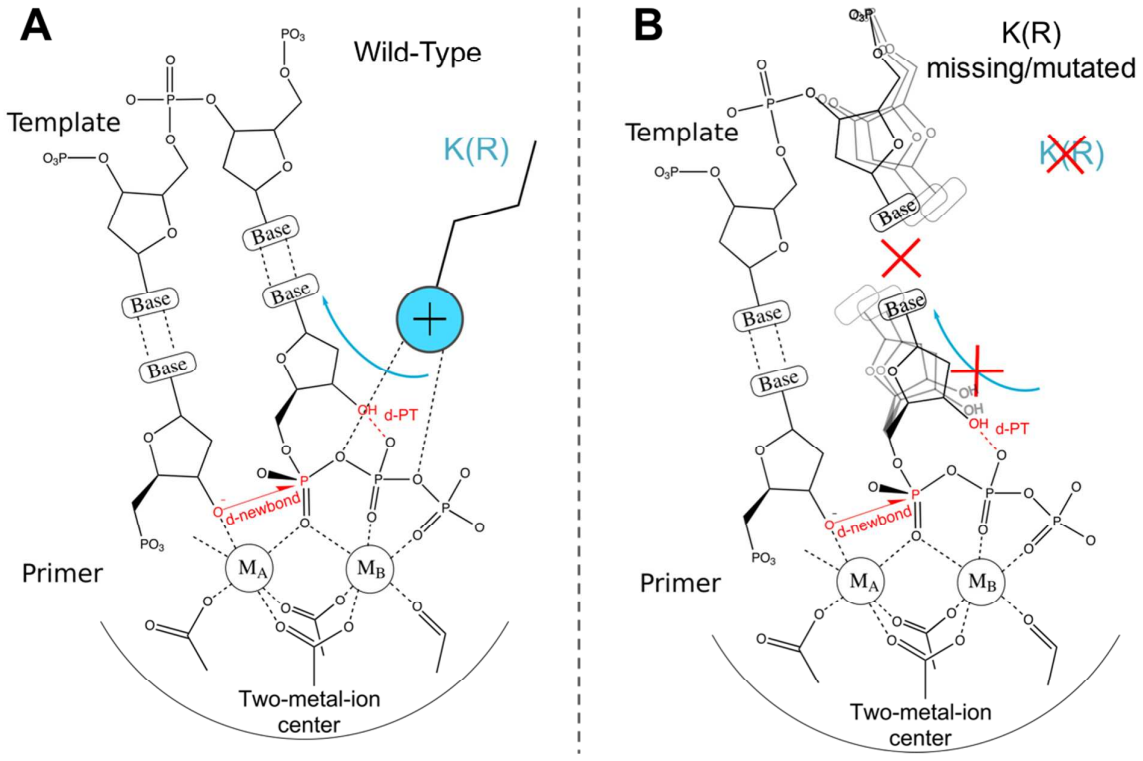


**Figure 2.** Scheme of the mechanism for K(R)-mediated base-pair stabilization during nucleic acid biosynthesis. (**A**) The positively charged side-chain of K(R) (i.e. cyan sphere) assists W-C geometry by stabilizing the phosphate tail of the incoming nucleotide dNTP. Concurrently, K(R) favors the formation of the Michaelis-Menten complex (i.e. key geometrical descriptors are reported in red) as discussed in our previous studies.[50;51] (**B**) The lack/mutation of the positively charged K(R) side-chain provokes the disruption of the H-bond framework characteristic of a stable W-C base-pairing.

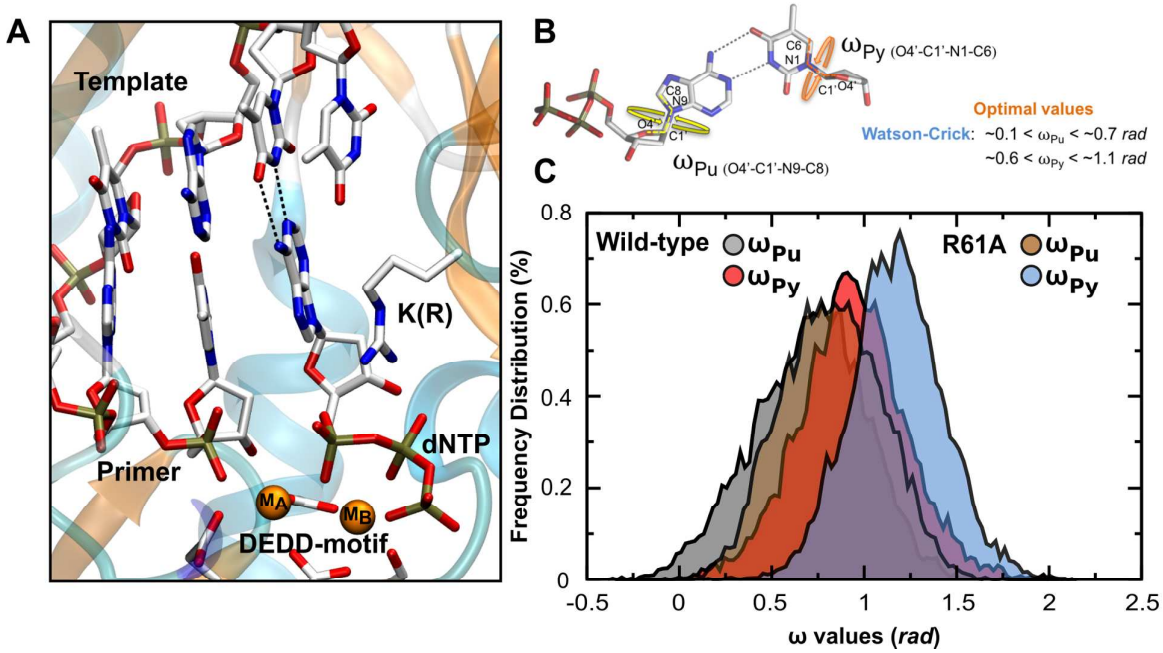


**Figure 3.** (**A**) Active site of human DNA Polymerase-η (Pol-η) in complex with dATP and dsDNA substrates. Orange spheres represent the two catalytic $Mg^{2+}$ ions; dATP and dsDNA are depicted in licorice; the protein is represented in cartoon. The image shows K(R) (i.e. Arg61) interacting with the α- and β-phosphate group of the incoming dATP. Black dashed lines indicate the H-bond typical of W-C base-pair geometry. (**B**) $\omega_{Pu}$ (i.e. O4'-C1'-N9-C8) and $\omega_{Py}$ (i.e. O4'-C1'-N1-C6) dihedral angles were monitored along plain MD simulations to investigate W-C base-pair stability in both wild-type and R61A mutant systems. Optimal values refer to the $\omega_{Pu}$ and $\omega_{Py}$ values measured in Pol-η/dsDNA/dATP X-ray structure (PDBid 4ECS).[52] (**C**) Frequency distribution of $\omega_{Pu}$ and $\omega_{Py}$ values in both wilt-type and R61-mutant MD simulations.

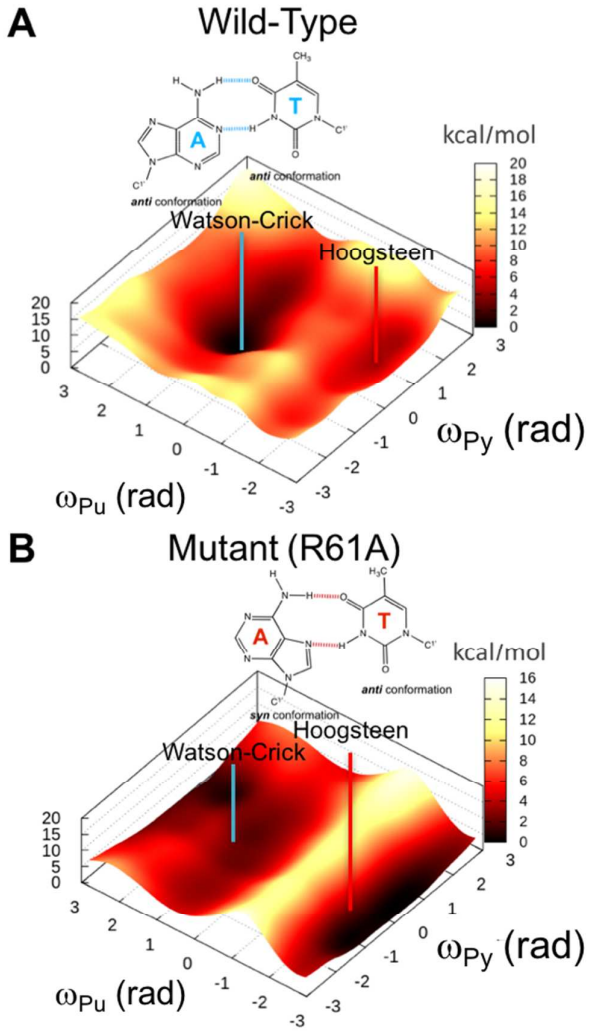


**Figure 4.** Free energy simulations of Watson-Crick base pairing (W-C) stability in presence and absence of K(R) side chain (i.e. Arg61 in Pol-η). Both free energies surfaces (FESs) were reconstructed using $\omega_{Pu}$ and $\omega_{Py}$ (see Fig. 3B and 'Materials and Methods' section) as collective variables to energetically describe dATP:dT base-pair stability. (**A**) FES of wild-type system showing two distinct energetic basins with different $\omega_{Pu}$ and $\omega_{Py}$ values. The deepest corresponds to Watson-Crick base pairing (W-C) while the relative one to the Hoogsteen base-pair (HG). (**B**) FES of R61A mutant system showing large and deep energetic basins with different $\omega_{Pu}$ and $\omega_{Py}$ values. Here, the deepest energetic minimum represents an ensemble of dATP:dT architectures far from the canonical W-C one. Among them, HG is the most sampled. On the contrary, a relative minimum identifies a cluster of architectures close to the iconic W-C one.

## References


(1) Loeb, L. A.; Monnat, R. J.; *Nat. Rev. Genet.* **2008**, *9*, 594.

(2) Cramer, P.; Armache, K. J.; Baumli, S.; Benkert, S.; Brueckner, E.; Buchen, C.; Damsma, G. E.; Dengl, S.; Geiger, S. R.; Jaslak, A. J.; Jawhari, A.; Jennebach, S.; Kamenski, T.; Kettenberger, H.; Kuhn, C. D.; Lehmann, E.; Leike, K.; Sydow, J. E.; Vannini, A.; *Annu. Rev. Biophys.* **2008**, *37*, 337.

(3) Lange, S. S.; Takata, K.; Wood, R. D.; *Nat. Rev. Cancer.* **2011**, *11*, 96.

(4) Bae, B.; Nayak, D.; Ray, A.; Mustaev, A.; Landick, R.; Darst, S. A.; *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, E4178.

(5) Warren, T. K.; Wells, J.; Panchal, R. G.; Stuthman, K. S.; Garza, N. L.; Van Tongeren, S. A.; Dong, L.; Retterer, C. J.; Eaton, B. P.; Pegoraro, G.; Honnold, S.; Bantia, S.; Kotian, P.; Chen, X.; Taubenheim, B. R.; Welch, L. S.; Minning, D. M.; Babu, Y. S.; Sheridan, W. P.; Bavari, S.; *Nature.* **2014**, *508*, 402.

(6) Rohs, R.; Jin, X. S.; West, S. M.; Joshi, R.; Honig, B.; Mann, R. S.; *Annu. Rev. Biochem.* **2010**, *79*, 233.

(7) Jiang, C. Z.; Pugh, B. F.; *Nat. Rev. Genet.* **2009**, *10*, 161.

(8) Liu, P. F.; Carvalho, C. M. B.; Hastings, P. J.; Lupski, J. R.; *Curr. Opin. Genet. Dev.* **2012**, *22*, 211.

(9) Hoogsteen, K.; *Acta. Cryst.* **1963**, *16*, 9.

(10) Nikolova, E. N.; Zhou, H. Q.; Gottardo, F. L.; Alvey, H. S.; Kimsey, I. J.; Al-Hashimi, H. M.; *Biopolymers.* **2013**, *99*, 955.

(11) Goobes, R.; Cohen, O.; Minsky, A.; *Nucleic Acids Res.* **2002**, *30*, 2154.

(12) Kimsey, I. J.; Petzold, K.; Sathyamoorthy, B.; Stein, Z. W.; Al-Hashimi, H. M.; *Nature.* **2015**, *519*, 315.

(13) Alvey, H. S.; Gottardo, F. L.; Nikolova, E. N.; Al-Hashimi, H. M.; *Nat. Commun.* **2014**, *5*.

(14) Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M.; *Nature.* **2011**, *470*, 498.

(15) Castro, C.; Smidansky, E. D.; Arnold, J. J.; Maksimchuk, K. R.; Moustafa, I.; Uchida, A.; Gotte, M.; Konigsberg, W.; Cameron, C. E.; *Nat. Struct. Mol. Biol.* **2009**, *16*, 212.

(16) Biertumpfel, C.; Zhao, Y.; Kondo, Y.; Ramon-Maiques, S.; Gregory, M.; Lee, J. Y.; Masutani, C.; Lehmann, A. R.; Hanaoka, F.; Yang, W.; *Nature.* **2010**, *465*, 1044.

(17) Su, Y.; Patra, A.; Harp, J. M.; Egli, M.; Guengerich, F. P.; *J. Biol. Chem.* **2015**, *290*, 15921.

(18) Nair, D. T.; Johnson, R. E.; Prakash, S.; Prakash, L.; Aggarwal, A. K.; *Nature.* **2004**, *430*, 377.

(19) Wu, W. J.; Su, M. I.; Wu, J. L.; Kumar, S.; Lim, L. H.; Wang, C. W. E.; Nelissen, F. H. T.; Chen, M. C. C.; Doreleijers, J. F.; Wijmenga, S. S.; Tsai, M. D.; *J. Am. Chem. Soc.* **2014**, *136*, 4927.

(20) Aggarwal, A.; Nair, D.; Johnson, R.; Prakash, L.; Prakash, S.; *Nature.* **2005**, *437*, E7.

(21) Wang, J. M.; *Nature.* **2005**, *437*, E6.

(22) Steitz, T. A.; Steitz, J. A.; *Proc. Natl. Acad. Sci.* **1993**, *90*, 6498.

(23) Steitz, T. A.; *Curr. Opin. Struct. Biol.* **1993**, *3*, 31.

(24) Genna, V.; Colombo, M.; De Vivo, M.; Marcia, M.; *Structure.* **2018**, *26*, 40.

(25) Ogawa, M.; Tosaka, A.; Ito, Y.; Yoshida, S.; Suzuki, M.; *Mutat. Res.* **2001**, *485*, 197.

(26) Joyce, C. M.; Steitz, T. A.; *Annu. Rev. Biochem.* **1994**, *63*, 777.

(27) Thompson, A. A.; Peersen, O. B.; *Embo J.* **2004**, *23*, 3462.

(28) Huang, H.; Chopra, R.; Verdine, G. L.; Harrison, S. C.; *Science.* **1998**, *282*, 1669.

(29) Franklin, M. C.; Wang, J.; Steitz, T. A.; *Cell.* **2001**, *105*, 657.

(30) Yin, Y. W.; Steitz, T. A.; *Cell.* **2004**, *116*, 393.

(31) Berman, A. J.; Kamtekar, S.; Goodman, J. L.; Lazaro, J. M.; de Vega, M.; Blanco, L.; Salas, M.; Steitz, T. A.; *Embo J.* **2007**, *26*, 3494.

(32) Wang, F.; Yang, W.; *Cell.* **2009**, *139*, 1279.

(33) Evans, R. J.; Davies, D. R.; Bullard, J. M.; Christensen, J.; Green, L. S.; Guiles, J. W.; Pata, J. D.; Ribble, W. K.; Janjic, N.; Jarvis, T. C.; *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20695.

(34) Gao, Y.; Yang, W.; *Science.* **2016**, *352*, 1334.

(35) Freudenthal, B. D.; Beard, W. A.; Shock, D. D.; Wilson, S. H.; *Cell.* **2013**, *154*, 157.

(36) Gouge, J.; Rosario, S.; Romain, F.; Poitevin, F.; Beguin, P.; Delarue, M.; *Embo J.* **2015**, *34*, 1126.

(37) Ferrer-Orta, C.; Arias, A.; Perez-Luque, R.; Escarmis, C.; Domingo, E.; Verdaguer, N.; *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9463.

(38) Jamsen, J. A.; Beard, W. A.; Pedersen, L. C.; Shock, D. D.; Moon, A. F.; Krahn, J. M.; Bebenek, K.; Kunkel, T. A.; Wilson, S. H.; *Nat. Commun.* **2017**, *8*.

(39) Jha, V.; Bian, C. B.; Xing, G. X.; Ling, H.; *Nucleic Acids Res.* **2016**, *44*, 4957.

(40) Nair, D. T.; Johnson, R. E.; Prakash, L.; Prakash, S.; Aggarwal, A. K.; *Science.* **2005**, *309*, 2219.

(41) Zhao, Y.; Gregory, M. T.; Biertumpfel, C.; Hua, Y. J.; Hanaoka, F.; Yang, W.; *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 8146.

(42) Zang, H.; Irimia, A.; Choi, J. Y.; Angel, K. C.; Loukachevitch, L. V.; Egli, M.; Guengerich, F. P.; *J. Biol. Chem.* **2006**, *281*, 2358.

(43) Bruskov, V. I.; Malakhova, L. V.; Masalimov, Z. K.; Chernikov, A. V.; *Nucleic Acids Res.* **2002**, *30*, 1354.

(44) Zang, H.; Irimia, A.; Choi, J. Y.; Angel, K.; Loukachevitch, L.; Egli, M.; Guengerich, F. P.; *J. Biol. Chem.* **2006**, *28*, 2358.

(45) Nair, D. T.; Johnson, R. E.; Prakash, L.; Prakash, S.; Aggarwal, A. K.; *Structure.* **2005**, *13*, 1569.

(46) Choi, J. Y.; Patra, A.; Yeom, M.; Lee, Y. S.; Zhang, Q. Q.; Egli, M.; Guengerich, F. P.; *J. Biol. Chem.* **2016**, *291*, 21063.

(47) Kirouac, K. N.; Ling, H.; *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3210.

(48) Frank, E. G.; Woodgate, R.; *J. Biol. Chem.* **2007**, *282*, 24689.

(49) Pence, M. G.; Blans, P.; Zink, C. N.; Hollis, T.; Fishbein, J. C.; Perrino, F. W.; *J. Biol. Chem.* **2009**, *284*, 1732.

(50) Genna, V.; Gaspari, R.; Dal Peraro, M.; De Vivo, M.; *Nucleic Acids Res.* **2016**, *44*, 2827.

(51) Genna, V.; Vidossich, P.; Ippoliti, E.; Carloni, P.; De Vivo, M.; *J. Am. Chem. Soc.* **2016**, *138*, 14592.

(52) Nakamura, T.; Zhao, Y.; Yamagata, Y.; Hua, Y. J.; Yang, W.; *Nature.* **2012**, *487*, 196.

(53) Barducci, A.; Bussi, G.; Parrinello, M.; *Phys. Rev. Lett.* **2008**, *100*, 020603.

(54) Ling, H.; Boudsocq, F.; Plosky, B. S.; Woodgate, R.; Yang, W.; *Nature.* **2003**, *424*, 1083.

(55) Warshel, A., Levitt, M.; *J. Mol. Biol.* . **1976**, *103*, 227.

(56) Warshel, A.; *Abstr. Pap. Am. Chem. Soc.* **2013**, *246*.

(57) Tuske, S.; Sarafianos, S. G.; Wang, X. Y.; Hudson, B.; Sineva, E.; Mukhopadhyay, J.; Birktoft, J. J.; Leroy, O.; Ismail, S.; Clark, A. D.; Dharia, C.; Napoli, A.; Laptenko, O.; Lee, J.; Borukhov, S.; Ebright, R. H.; Arnold, E.; *Cell.* **2005**, *122*, 541.

(58) Zhang, G. Y.; Campbell, E. A.; Minakhin, L.; Richter, C.; Severinov, K.; Darst, S. A.; *Cell.* **1999**, *98*, 811.

(59) Glick, E.; Vigna, K. L.; Loeb, L. A.; *Embo J.* **2001**, *20*, 7303.

(60) Malinen, A. M.; Turtola, M.; Parthiban, M.; Vainonen, L.; Johnson, M. S.; Belogurov, G. A.; *Nucleic Acids Res.* **2012**, *40*, 7442.

(61) Zang, H.; Irimia, A.; Choi, J. Y.; Angel, K. C.; Loukachevitch, L. V.; Egli, M.; Guengerich, F. P.; *J. Biol. Chem.* **2008**, *283*, 13500.

(62) Masutani, C.; Kusumoto, R.; Yamada, A.; Dohmae, N.; Yokoi, M.; Yuasa, M.; Araki, M.; Iwai, S.; Takio, K.; Hanaoka, F.; *Nature.* **1999**, *399*, 700.

(63) Yoon, J. H.; Prakash, L.; Prakash, S.; *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 18219.

(64) Vashishtha, A. K.; Wang, J.; Konigsberg, W. H.; *J. Biol. Chem.* **2016**, *291*, 20869.

(65) Chen, Y.; Zhang, J.; Liu, H.; Gao, Y.; Li, X.; Zheng, L.; Cui, R.; Yao, Q.; Rong, L.; Li, J.; Huang, Z.; Ma, J.; Gan, J.; *Plos. Biol.* **2017**, *15*, e1002599.

(66) Katafuchi, A.; Sassa, A.; Niimi, N.; Gruz, P.; Fujimoto, H.; Masutani, C.; Hanaoka, F.; Ohta, T.; Nohmi, T.; *Nucleic Acids Res.* **2010**, *38*, 859.

(67) Lee, Y. S.; Gao, Y.; Yang, W.; *Nat. Struct. Mol. Biol.* **2015**, *22*, 298.

(68) Wu, W. J., Yang, W. and Tsai, M. D.; *Nat. Rev. Chem.* **2017**, *1*, 16.

(69) Roberts, S. A.; Gordenin, D. A.; *Nat. Rev. Cancer.* **2014**, *14*, 786.

(70) Ucisik, M. N.; Hammes-Schiffer, S.; *J. Am. Chem. Soc.* **2015**, *137*, 13240.

(71) Kraynov, V. S.; Showalter, A. K.; Liu, J.; Zhong, X.; Tsai, M. D.; *Biochemistry.* **2000**, *39*, 16008.

(72) Perera, L.; Freudenthal, B. D.; Beard, W. A.; Shock, D. D.; Pedersen, L. G.; Wilson, S. H.; *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, E5228.

(73) Yang, W.; *Biochemistry.* **2014**, *53*, 2793.

(74) Su, Y.; Egli, M.; Guengerich, F. P.; *J. Biol. Chem.* **2017**, *292*, 18044.

(75) McCulloch, S. D.; Kokoska, R. J.; Masutani, C.; Iwai, S.; Hanaoka, F.; Kunkel, T. A.; *Nature.* **2004**, *428*, 97.

(76) Patra, A.; Nagy, L. D.; Zhang, Q. Q.; Su, Y.; Muller, L.; Guengerich, F. P.; Egli, M.; *J. Biol. Chem.* **2014**, *289*, 16867.

(77) Castro, C.; Smidansky, E. D.; Arnold, J. J.; Maksimchuk, K. R.; Moustafa, I.; Uchida, A.; Gotte, M.; Konigsberg, W.; Cameron, C. E.; *Nat. Struct. Mol. Biol.* **2009**, *16*, 212.

(78) Warshel, A.; Russell, S. T.; *Q. Rev. Biophys.* **1984**, *17*, 283.

(79) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E.; *Proteins.* **2010**, *78*, 1950.

(80) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M.; *Biophys. J.* **2007**, *92*, 3817.

(81) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A.; *J. Comput. Chem.* **2004**, *25*, 1157.

(82) Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R.; *Comput. Phys. Commun.* **1995**, *91*, 43.

(83) Grest, G. S.; Kremer, K.; *Phys. Rev. A.* **1986**, *33*, 3628.

(84) Dal Peraro, M.; Spiegel, K.; Lamoureux, G.; De Vivo, M.; DeGrado, W. F.; Klein, M. L.; *J. Struct. Biol.* **2007**, *157*, 444.

(85) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L.; *J. Chem. Phys.* **1983**, *79*, 926.

(86) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G.; *Comput. Phys. Commun.* **2014**, *185*, 604.

(87) Car, R.; Parrinello, M.; *Phys. Rev. Lett.* **1985**, *55*, 2471.

(88) Laio, A.; VandeVondele, J.; Rothlisberger, U.; *J. Chem. Phys.* **2002**, *116*, 6941.

(89) Carloni, P.; Rothlisberger, U.; Parrinello, M.; *Acc. Chem. Res.* **2002**, *35*, 455.

(90) Troullier, N.; Martins, J. L.; *Phys. Rev. B.* **1991**, *43*, 8861.

(91) Nose, S.; *J. Chem. Phys.* **1984**, *81*, 9.

(92) Hoover, W. G.; *Phys. Rev. A.* **1985**, *31*, 1695.

(93) Palermo, G.; Cavalli, A.; Klein, M. L.; Alfonso-Prieto, M.; Dal Peraro, M.; De Vivo, M.; *Acc. Chem. Res.* **2015**, *48*, 220.

(94) Dreyer, J., Brancato, G., Ippoliti, E., Genna, V., De Vivo, M., Carloni, P., Rothlisberger, U. In *Simulating Enzyme Reactivity: Computational Methods in Enzyme Catalysis;* Tunon, I., Moliner, V.; Ed.; Royal Society of Chemistry: U.K., 2016, pp 294-339.

(95) Lonsdale, R.; Ranaghan, K. E.; Mulholland, A. J.; *Chem. Commun.* **2010**, *46*, 2354.

(96) Monard, G.; Merz, K. M.; *Acc. Chem. Res.* **1999**, *32*, 904.

(97) Campomanes, P.; Rothlisberger, U.; Alfonso-Prieto, M.; Rovira, C.; *J. Am. Chem. Soc.* **2015**, *137*, 11170.

(98) Dal Peraro, M.; Ruggerone, P.; Raugei, S.; Gervasio, F. L.; Carloni, P.; *Curr. Opin. Struct. Biol.* **2007**, *17*, 149.

(99) Ciccotti, G.; Ferrario, M.; Hynes, J. T.; Kapral, R.; *Chem. Phys.* **1989**, *129*, 241.

**Table of Content**