

A self-activated mechanism for nucleic acid polymerization catalyzed by DNA/RNA polymerases

Vito Genna^{1,2}, Pietro Vidossich², Emiliano Ippoliti², Paolo Carloni^{2*} and Marco De Vivo^{1,2*}

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

² IAS-5 / INM-9 Computational Biomedicine and JARA-HPC, Forschungszentrum Jülich, Wilhelm-Johnen-Straße 52428 Jülich, Germany

*Corresponding Authors

Prof. Paolo Carloni

E-mail: p.carloni@fz-juelich.de

Dr. Marco De Vivo

E-mail: marco.devivo@iit.it

Supporting Information

Table of contents

Supplementary Text

Car-Parrinello Molecular Dynamics (QM/MM).....	page S3
Role of metal ions in SAM.....	page S4

Supplementary Figures

Figure 1. Force profiles and FES for DNA translocation investigated with no bias on PT for nucleophile activation.....	page S5
Figure 2. Superimposition between QM/MM and X-ray structures.....	page S6
Figure 3. Free energy surface for SAM in human DNA Pol- η with transient MgC.	page S7
Figure 4. Free energy surface for pyrophosphorolysis reaction in presence of transient MgC.....	page S8
Figure 5. Free energy surface for SAM with R61 adopting C-conf.....	page S10
Figure 6. Structural and energetic detail of WMSA mechanism investigated in Pol- η	page S11
Figure 7. Graph reporting convergence of ab initio metadynamics simulations.....	page S12

Supplementary tables

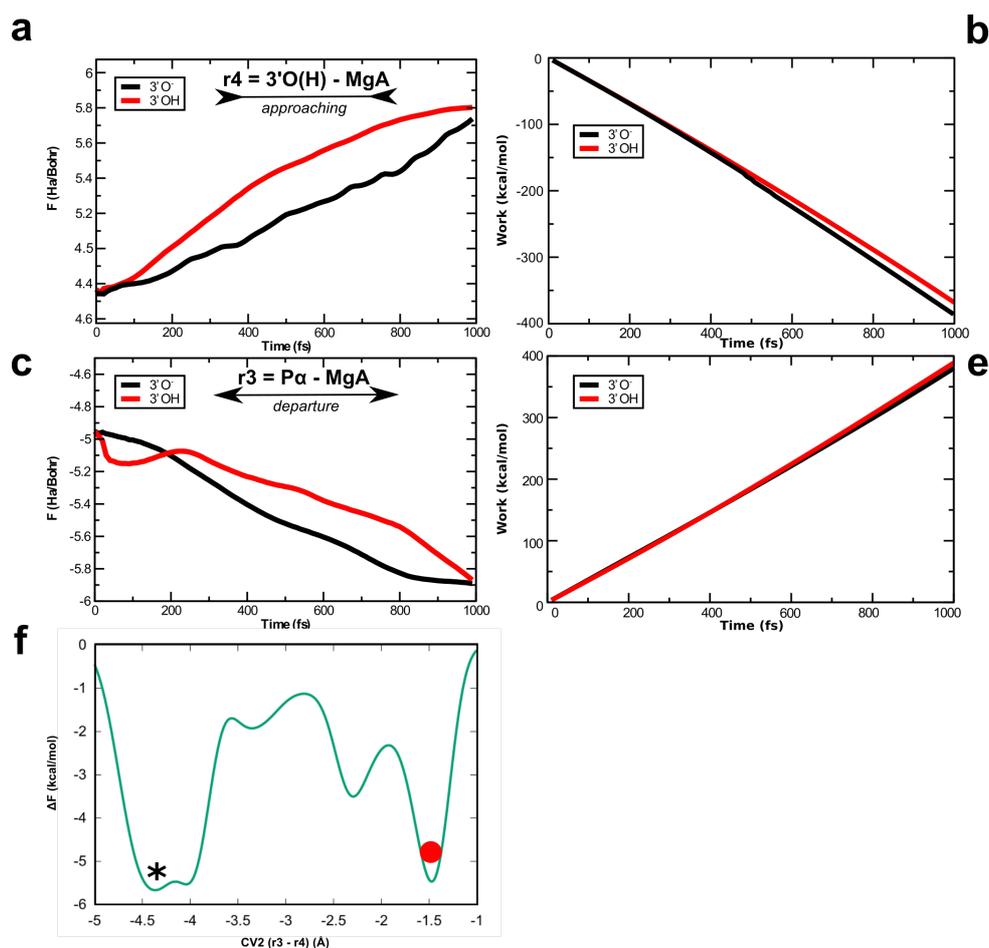
Table 1. Value of d-PT measured in different Pols ternary complexes.....	page S13
---	----------

References	page S14
-------------------------	----------

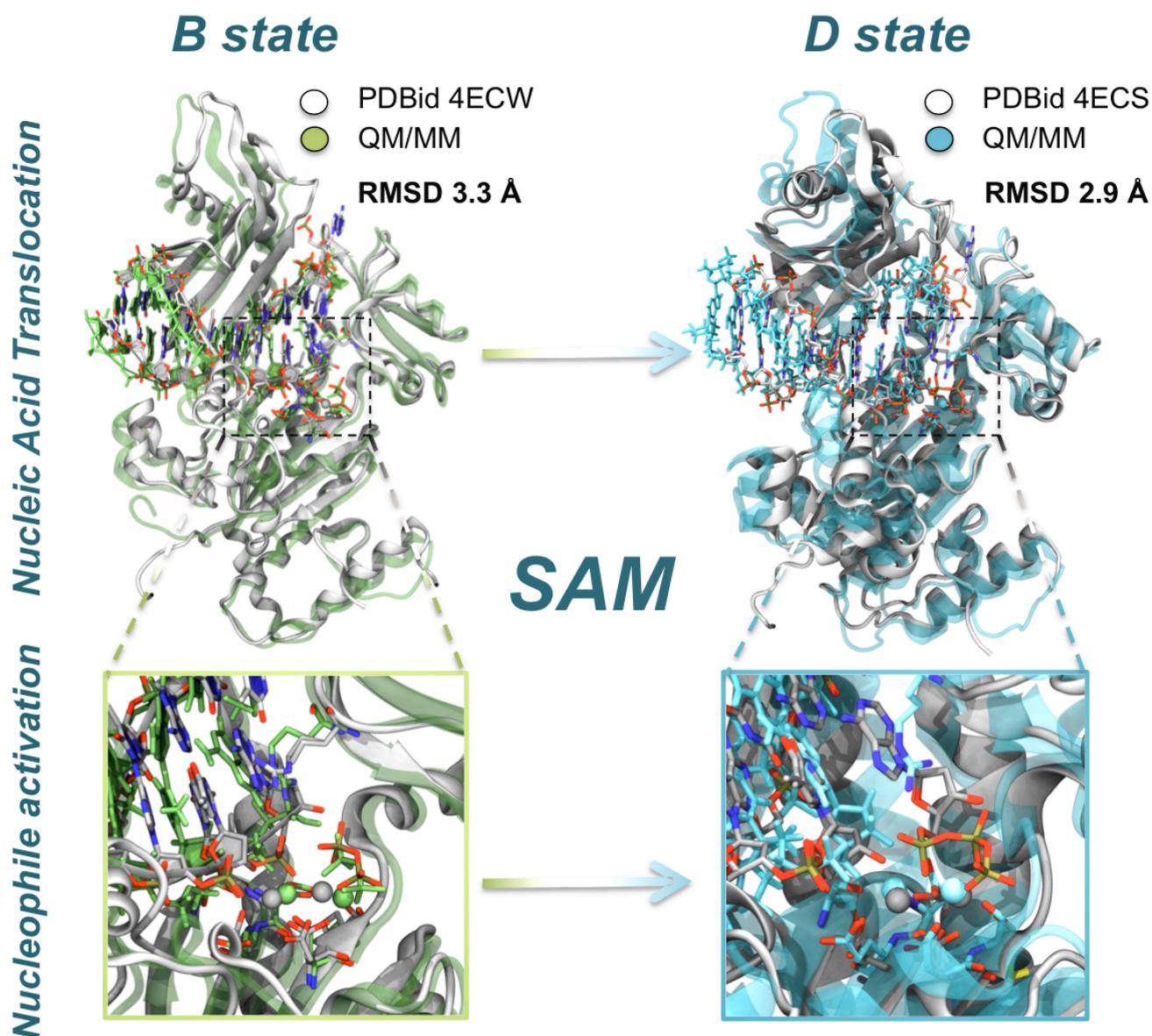
Car-Parrinello Molecular Dynamics (QM/MM MD). Car-Parrinello MD is an established *ab initio* method that does not rely on empirically determined potentials¹. Instead, internuclear forces are determined on-the-fly from electronic structure calculations using the Kohn-Sham formulation of the Density Functional Theory (DFT)². Such approach, despite being much more computationally demanding than classical and semiempirical methods, has been demonstrated to be robust and informative when used to study chemical reactivity, including systems based on metals and their complexes. Due to its high computational cost, a full QM treatment of the whole system (about ~70,000 atoms) is not tractable. Thus, we rely on a hybrid QM/MM potential³ to investigate SAM for Pol- η catalysis. According to this approach, the QM description is reserved to the enzyme's active site (including the two catalytic metals and a relevant portion of the DNA substrate and incoming nucleotide), while the remaining part of the system is treated at classical level. In detail, the reactive region of the complex – 183 atoms in total, including MgA and MgB and the M17, R61, D113, D115, E116 residues, five water molecules and key atoms of the nucleic acid involved in the reaction – was treated at the DFT level using the BLYP functional^{4;5}. This level of theory has been shown to describe well a variety of enzymatic reactions, with a fairly good (semi-qualitative) estimation of the energetics for catalysis⁶⁻⁹.

The remaining part of the system, including all the other water molecules and counterions, was treated at the AMBER parm99 force field level¹⁰. In more details, the wave functions are expanded in a plane wave basis set up to 70 Ry in a QM cell of proper dimensions, taking as reference our previous studies^{6;11;12}. Only the valence electrons are treated explicitly (in the case of Mg atoms, electrons in the $n = 5$ shell are also included in the valence), while the core electrons will be described using norm-conserving pseudopotentials of the Martins–Troullier type¹³. An adapted monovalent carbon pseudopotential is employed to saturate the dangling bonds in between the QM and MM regions. Isolated system conditions¹⁴ is imposed in the QM part by employing the Martyna-Tuckerman scheme¹⁵. Notably, this approach has been shown to accurately describe a variety of enzymatic systems and protein/DNA complexes^{9;16;17}. All simulations were performed with the code CPMD (www.cpmc.org) version 4.1.1, developed at IBM Research Division, Zürich and the Max-Planck Institute, Stuttgart. This program features an efficient QM/MM interface developed by Prof. Röthlisberger and co-workers³ based on the Gromos MD engine¹⁸.

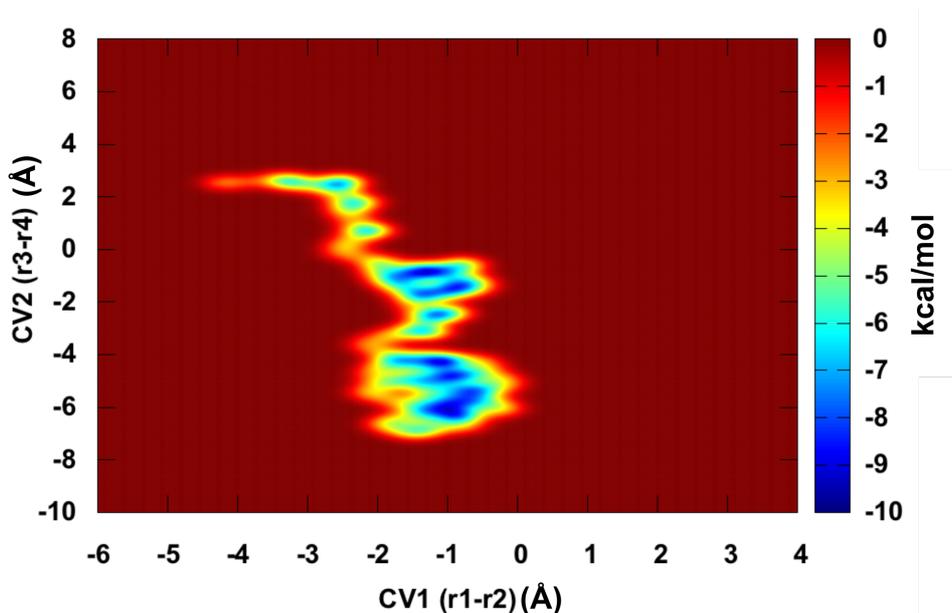
Role of Metal ions in SAM. The 2M-containing catalytic pocket actively assists this self-assisted and orchestrated mechanism. In fact, we observed fluctuations in the MgA-MgB distance and variation in the MgA coordination geometry along the different steps of the reaction. In **A**, metals are 3.35 Å apart with MgA showing a coordination vacancy in its apical position. This vacancy is then fulfilled by the shifting of the 3'O(H) group, with final formation of the octahedral coordination centered on MgA, in **B**. In **B**, the inter-metal distance adopts a value of ~4.00 Å, which agrees well with ternary complexes of postreactive states of different polymerases.¹⁹⁻²¹ Then, the two ions slowly return to their initial internuclear distance of ~3.5 Å, moving from **C** to **D-E-A**, to catalyze the phosphoryl transfer along the catalytic cycle. Indeed, the two ions have been reported to get closer (from ~4 to ~3.35 Å) during the phosphoryl transfer for nucleotide addition, as found in several other similar two-metal-ion enzymes for nucleic acid processing (see refs 7, 8, 35, 36, 46, 47 in the revised manuscript).



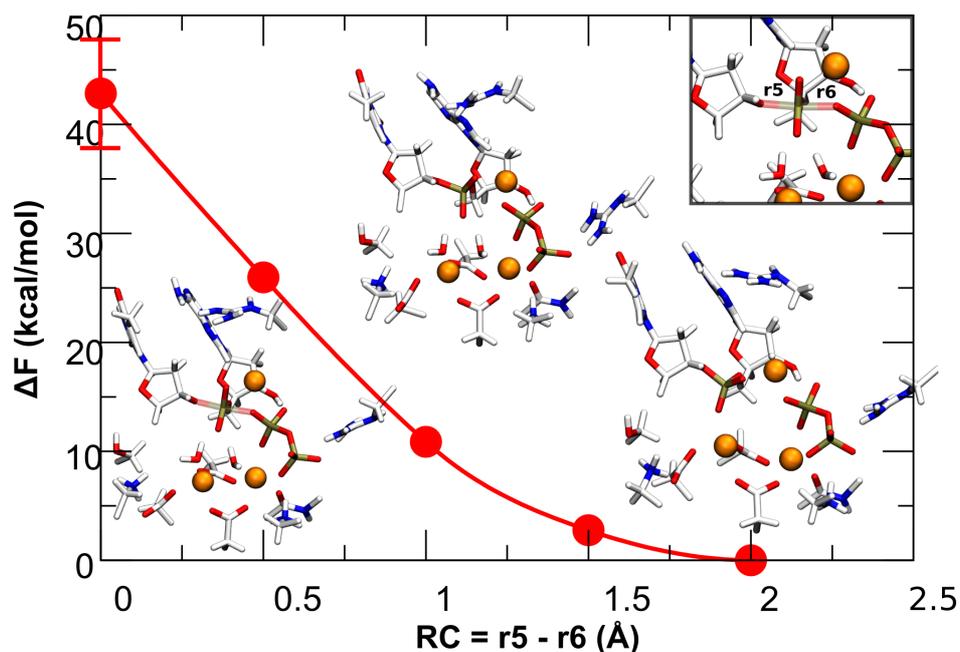
Supplementary Figure 1. **A** and **C**, force profiles derived from *ab initio* CP QM/MM steered MD simulations (SMD) along r_3 and r_4 (CV2) components to determine DNA translocation. The plot shows the resulting mean values by averaging the force profiles from SMD runs using either 3'OH or 3'O. **B** and **E**, external work for DNA translocation calculated from profiles in **A** and **C**, respectively. **F**, free energy surface for SAM in human DNA Pol- η obtained by means of CP QM/MM metadynamics, considering CV2 (i.e. $r_3 - r_4$), only. The red dot refers to the starting point (see Fig.3 point B in the manuscript). From here, the system explores additional minima, including a large well at about CV \sim 4-4.5 (indicated by an asterisk). Notably, the free energy landscape in **F** was investigated by employing the same computational protocol applied for SAM where we coupled CV1 with CV2 (hills height and width were set to $0.05 \text{ kcal mol}^{-1}$ and 0.01 \AA , respectively). A total of 250 gaussian hills were deposited. The resulting profile in **F** indicates that DNA translocation – described by only CV2 - is highly unfavored compared to the case where CV2 is coupled to CV1, i.e. nucleophile deprotonation (see Fig.3 in the manuscript). Although only qualitative, these computational results, taken together, indicate a key role in SAM of the coupling between the chemical (CV1, i.e. proton-transfer) and physical (CV2, i.e. nucleic acid translocation) steps for nucleophile formation and nucleic acid translocation.



Supplementary Figure 2. Before and after SAM. Superimposition between X-ray structures (in white) and snapshots extracted from our QM/MM metadynamics (see Fig. 6 in the manuscript). **B-state** represents the starting model system, in green, superimposed with the X-ray structure (PDBid 4ECW²², in white) of Pol- η 's ternary complex in the post-reactive state. Bottom: close view of the ternary complex showing the protonated 3'OH. **D-state** represents the superimposition between a snapshot extracted from our QM/MM simulations and the X-ray structure (PDBid 4ECS²² in cyan; see also Fig. 6 in the manuscript). Here, the DNA strand is translocated of one base. Bottom: close view of the active site ready to accommodate a new incoming NTP.

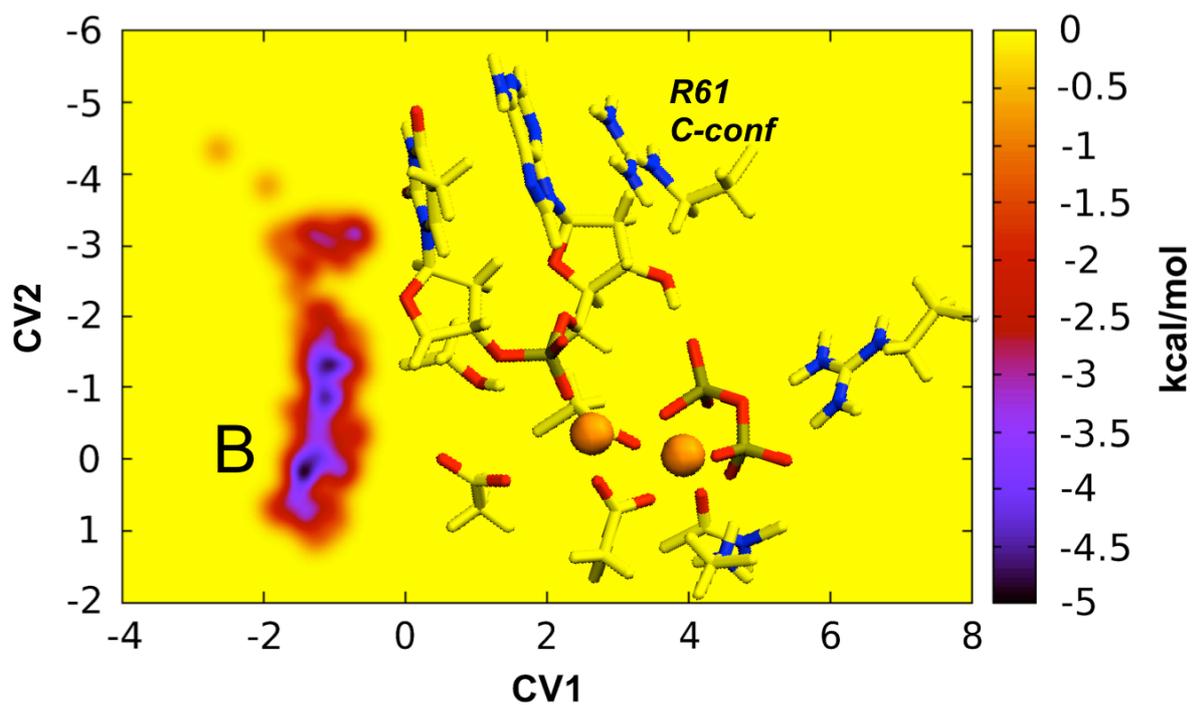


Supplementary Figure 3. Free energy surface for SAM in human DNA Pol- η with third Mg^{2+} ion bound to the pre-translocation complex. The ensemble of global minimum refers to the starting point (see reaction scheme and point B in Fig. 5 in the manuscript). Neither proton-transfer nor DNA translocation was observed with MgC bound in this conformation further corroborating the structural evidence that a third metal ion cannot be placed in the reactant enzyme-substrate complex, mainly because of steric clashes.²³ The free energy landscape was investigated by employing the same computational protocol previously applied for SAM (in the absence of a transient metal ion bound to the active site). Hills height and width were set to $0.05 \text{ kcal mol}^{-1}$ and 0.01 \AA , respectively). Also the same collective variables were used (CV1 and CV2, see Fig. 3 in the manuscript) for the proton transfer under consideration, defined as the difference between the lengths of the $3'\text{O}-\text{H}$ (r_a) and the $\text{H}-\text{O}_{\text{wat}}$ bonds (r_b). A total of 250 gaussian hills were deposited.

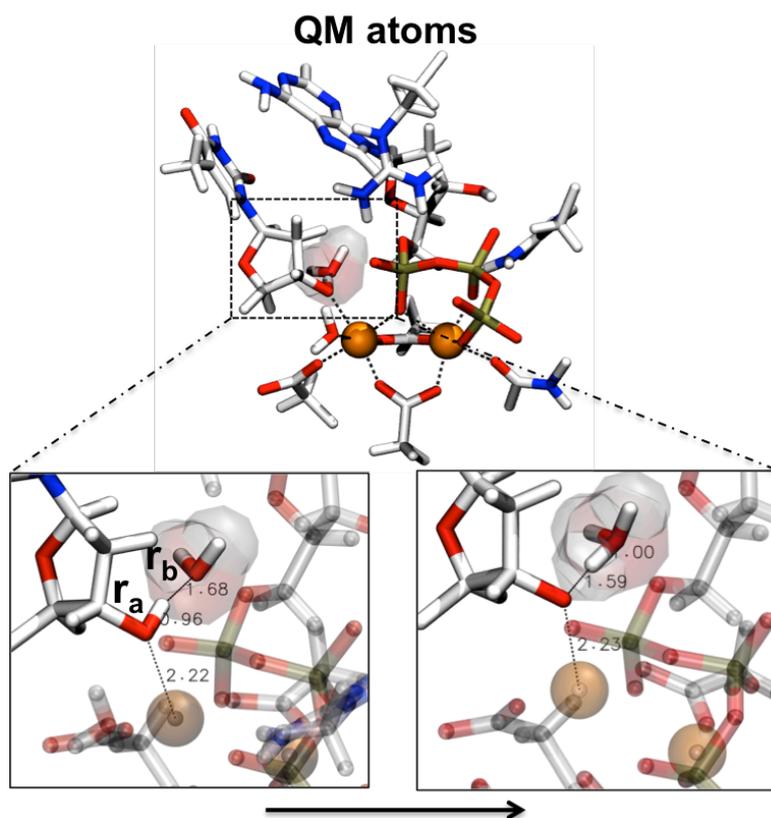


Supplementary Figure 4. Free energy profile and structural snapshots along the back-reaction (i.e. pyrophosphorolysis) in the presence of transient MgC in Pol- η . Each reaction pathway is studied using Car-Parrinello (CP) QM/MM dynamics: the reactive region of the complex (183 atoms) is treated at the quantum level (DFT-BLYP) and includes the same atoms reported in the methods section of the manuscript, plus MgC. The remaining part of the system is treated using at the classical level. The valence electrons are described by plane wave basis set up to a cutoff of 70 Ry. CP QM/MM dynamics are carried out with a time step of 0.12 fs (for a total of run time of ~ 50 ps). The QM/MM protocol includes an initial equilibration of the configuration produced by MD simulations, followed by runs where only the MM part is free to move, while the QM-part is kept frozen. Then, the whole system is allowed to move and heat up to 300K (~ 2 ps); after that trajectories are collected for analysis. Configurations from the equilibrated QM/MM simulations are used for free energy calculations. The pyrophosphorolysis reaction is described with a reaction coordinate (RC) defined as the difference between the length of the breaking bond ($3'\text{O}-\text{P}^\alpha$, r_5) and that of the forming bond ($\text{P}^\alpha-\text{O}_{\text{ppi}}$). This RC is well suited for $\text{S}_{\text{N}}2$ -like reactions (e.g. see ref. 35 in the manuscript). “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. The pathway from reactants to products is divided in 4 steps, with a resolution of 0.5 \AA . 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 integration of the force profile. The error associated to 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

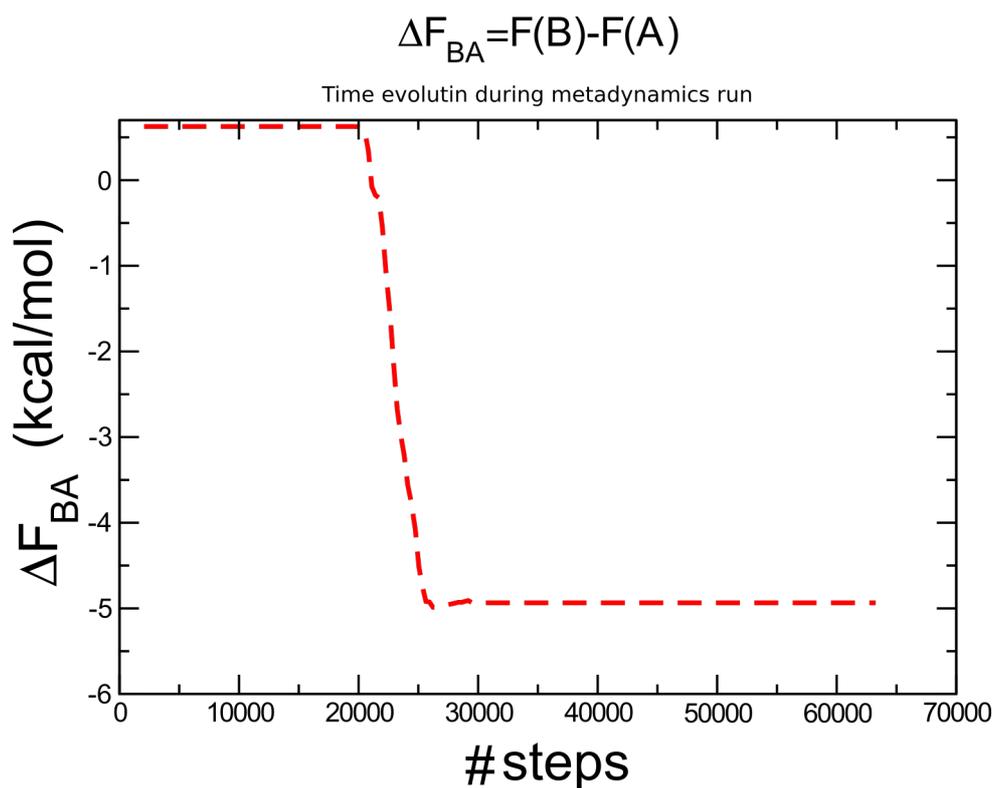
considered approximate due to the still limited sampling accessible to first principles DFT calculations, the choice of a 1-D RC, and the limitations of current GGA XC functionals. The catalytic pathways are then characterized in terms of the variation of critical bond lengths averaged over the equilibrated trajectory of each simulation step.



Supplementary Figure 5. Free energy surface for SAM in human DNA Pol- η with R61 adopting C-conf.²⁴ **B** identifies an ensemble of global minimum that refers to the starting point (see reaction scheme and point B in Fig. 3). Neither proton-transfer nor DNA translocation were observed with R61 in this conformation.



Supplementary Figure 6. We investigated nucleophile deprotonation in favor of a water molecule (Wat_N) conveniently located nearby the 3'OH group, as previously suggested by Nakamura et al. (Ref. 16 in manuscript). Notably, in this mechanism, the 3'OH group is already properly placed (PDBid 4ECS)²² to perform the in-line nucleophilic attack on the incoming nucleotide after DNA translocation. Thus, these simulations were started from a model system representing the reactant state of Pol- η as in PDBid 4ECS. The reaction pathway was investigated via the same computational protocol applied for SAM (hills height and width were set to $0.05 \text{ kcal mol}^{-1}$ and 0.01 \AA , respectively) but using a single collective variable (CV_{Wat}) for the proton transfer under consideration, defined as the difference between the lengths of the 3'O—H (r_a) and the H—O_{Wat} bonds (r_b). A total of 250 gaussian hills (in $\sim 50\,000$ steps) were deposited. In this case, the deepest minimum on the FES was the initial system conformation, where $r_a = 0.96 \text{ \AA}$ and $r_b = 1.68 \text{ \AA}$. Afterward, the system evolved to reach the transition state for nucleophile activation, where $r_a = r_b = 1.82 \text{ \AA}$, which returned a free-energy barrier of $\sim 7.2 \text{ kcal/mol}$. Subsequently, when $\text{CV}_{\text{Wat}} = \sim 0.6 \text{ \AA}$, the 3'-hydroxide and the hydronium ion were stably formed. A structured H-bond network built by surrounding water molecules stabilized the newly formed hydronium ion, although no spontaneous proton shuttle was observed. The higher energy cost for such mechanism for nucleophile deprotonation suggests that nucleophile activation via Wat_N seems unfavored compared to SAM for Pol- η catalysis.



Supplementary Figure 7. Free energy difference between state **B** and state **A** (see Fig. 6 in the manuscript for the definition of the states) during the progress of the metadynamics runs. Convergence is reached after $\sim 22,000$ steps. As evident from the graph, ΔF_{BA} , which is the relevant quantity for understanding if the 3'OH is more acidic than a PPi group bound to Mg^{2+} , appears well converged, at least within the metadynamics runs performed in our study.

Adenine		Guanine		Cytosine		Thymine	
PDBids	d-PT	PDBids	d-PT	PDBids	d-PT	PDBids	d-PT
4WC6	3.72 Å	3MAQ	3.09 Å	3OLB*	2.76 Å	4FS1	2.85 Å
4Q4Z	3.42 Å	3F2B	2.78 Å	5CWR	2.97 Å	1IG9	2.77 Å
3VNU	3.57 Å	2E2I	2.80 Å	4RPZ	2.97 Å	2E9R	3.46 Å
2Q66	2.70 Å	2Q66	2.70 Å	4IRK	3.15 Å		
4M8O	2.80 Å	4FWT	3.48 Å	4DQI	2.73 Å	Uracil	
4ECS	2.77 Å			3IAY	2.70 Å	PDBids	d-PT
3KK2	2.82 Å			3GQC	3.15 Å	4YD1	2.70 Å

Supplementary Table 1. Values of the intramolecular H-bond (d-PT) measured in the nucleotide triphosphate of different ternary complexes of Polymerases, across each domain of life.

Reference

- (1) Car, R.; Parrinello, M. *Phy. Rev. Lett.* **1985**, *55*, 2471.
- (2) Kohn, W.; Sham, L. J. *Phy. Rev. Lett.* **1965**, *140*, A1133
- (3) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941.
- (4) Becke, A. D. *Phys. Rev. A.* **1988**, *38*, 3098.
- (5) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B.* **1988**, *37*, 785.
- (6) De Vivo, M.; Dal Peraro, M.; Klein, M. L. *J. Am. Chem. Soc.* **2008**, *130*, 10955.
- (7) Dal Peraro, M.; Ruggerone, P.; Raugei, S.; Gervasio, F. L.; Carloni, P. *Curr. Opin. Struct. Biol.* **2007**, *17*, 149.
- (8) Iannuzzi, M.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 238302.
- (9) De Vivo, M.; Ensing, B.; Klein, M. L. *J. Am. Chem. Soc.* **2005**, *127*, 11226.
- (10) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
- (11) De Vivo, M.; *Front Biosci (Landmark Ed)*. **2011**, *16*, 1619.
- (12) De Vivo, M.; Ensing, B.; Dal Peraro, M.; Gomez, G. A.; Christianson, D. W.; Klein, M. L. *J. Am. Chem. Soc.* **2007**, *129*, 387.
- (13) Troullier, N.; Martins, J. L. *Phy. Rev. B.* **1991**, *43*, 1993.
- (14) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *J. Chem. Phys.* **2005**, *122*, 14113.
- (15) Martyna, G. J.; Tuckerman, M. E. *J. Chem. Phys.* **1999**, *110*, 2810.
- (16) De Vivo, M.; Cavalli, A.; Carloni, P.; Recanatini, M. *Chemistry.* **2007**, *13*, 8437.
- (17) Cavalli, A.; De Vivo, M.; Recanatini, M.; *Chem. Comm.* **2003**, 1308.
- (18) van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. Vdf Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, **1996**, 1.
- (19) Vaisman, A.; Ling, H.; Woodgate, R.; Yang, W. *Embo J.* **2005**, *24*, 2957.
- (20) Vyas, R.; Reed, A. J.; Tokarsky, E. J.; Suo, Z. C. *J. Am. Chem. Soc.* **2015**, *137*, 5225.
- (21) Palermo, G.; Cavalli, A.; Klein, M. L.; Alfonso-Prieto, M.; Dal Peraro, M.; De Vivo, M. *Acc. Chem. Res.* **2015**, *48*, 220.
- (22) Nakamura, T.; Zhao, Y.; Yamagata, Y.; Hua, Y. J.; Yang, W. *Nature.* **2012**, *487*, 196.
- (23) Gao, Y.; Yang, W. *Science.* **2016**, *352*, 1334.
- (24) Genna, V.; Gaspari, R.; Dal Peraro, M.; De Vivo, M.; *Nucleic Acids Res.* **2016**, *44*, 2827.