nature catalysis

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41929-019-0227-5

In the format provided by the authors and unedited.

Intrinsic cleavage of RNA polymerase II adopts a nucleobase-independent mechanism assisted by transcript phosphate

Carmen Ka Man Tse[1,8], Jun Xu[2,8], Liang Xu[2,3], Fu Kit Sheong[1], Shenglong Wang[4], Hoi Yee Chow[5], Xin Gao[6], Xuechen Li[5], Peter Pak-Hang Cheung[1*], Dong Wang[2*], Yingkai Zhang[4,7*] and Xuhui Huang[1*]

[1]Department of Chemistry, Centre of Systems Biology and Human Health, State Key Laboratory of Molecular Neuroscience, The Hong Kong University of Science and Technology, Kowloon, Hong Kong. [2]Department of Cellular and Molecular Medicine, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA, USA. [3]Department of Chemistry, Sun Yat-Sen University, Guangzhou, China. [4]Department of Chemistry, New York University, New York, NY, USA. [5]Department of Chemistry, State Key Lab of Synthetic Chemistry, The University of Hong Kong, Pok Fu Lam, Hong Kong. [6]Computational Bioscience Research Centre, CEMSE Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. [7]NYU-ECNU Centre for Computational Chemistry at NYU Shanghai, Shanghai, China. [8]These authors contributed equally: Carmen Ka Man Tse, Jun Xu. *e-mail: ppcheung@ust.hk; dongwang@ucsd.edu; yz22@nyu.edu; xuhuihuang@ust.hk

Supplementary Information for

Intrinsic Cleavage of RNA Polymerase II Adopts a Nucleobase-independent Mechanism Assisted by Transcript Phosphate

Carmen Ka Man Tse[1,+], Jun Xu[2,+], Liang Xu[2,4], Fu Kit Sheong[1], Shenglong Wang[3], Hoi Yee Chow[5], Xin Gao[7], Xuechen Li[5], Peter Pak-Hang Cheung[1*], Dong Wang[2*], Yingkai Zhang[3,6*], Xuhui Huang[1*]

[1]Department of Chemistry, Centre of Systems Biology and Human Health, State Key Laboratory of Molecular Neuroscience, The Hong Kong University of Science and Technology, Kowloon, Hong Kong

[2]Department of Cellular and Molecular Medicine, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California 92093, United States

[3]Department of Chemistry, New York University, New York, New York 10003 United States

[4]Department of Chemistry, Sun Yat-Sen University, Guangzhou, China

[5]Department of Chemistry, State Key Lab of Synthetic Chemistry, The University of Hong Kong, Hong Kong

[6]NYU-ECNU Centre for Computational Chemistry at NYU Shanghai, Shanghai 200062, China

[7]Computational Bioscience Research Centre (CBRC), CEMSE Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

*Email: ppcheung@ust.hk or dongwang@ucsd.edu or yz22@nyu.edu or xuhuihuang@ust.hk

[+] These authors contribute equally to this work.

Supplementary Note 1. Obtaining initial structure from classical MD simulation. All the MD calculations were performed using GROMACS 4.5.4[1]. The protein was described by Amber ff03[2] as previous Pol II studies[3,4]. Modified nucleic backbone parameters[5] were used. Improved Mg^{2+} L-J parameters were used[6] to obtain better description between Mg^{2+} ion, phosphate group and water molecule. Harmonic restraints (b_0=0.242 nm, k_b=10000 kJ mol^{-1} nm^{-2}) were added between Zn^{2+} ions and the corresponding ligands, with b_0 as the average coordination distance calculated from crystal structure and k_b arbitrarily assigned. Dodecahedron box was used to minimise the size of the system. Box size was set by defining the boundary distance as 0.7 nm, a value large enough to avoid image interaction. The system was solvated using explicit TIP3P water and 81 Na^+ ions were added to neutralise the system. The final system contains 372,194 atoms. After solvation, the system was subjected to 5000 steps of energy minimization using steepest descent. Since the loosely bound Mg_B is absent, it was modelled by replacing a bridging water molecule found in the active site after minimization (Supplementary Fig. 15). Justification could be provided from the study on reaction *in crystallo* in *D. mobilis* homing endonuclease I-Dmol, where the metal binding sites was found to be pre-occupied by water molecules[7].

To further equilibrate the system, MD simulation with heavy atoms restrained at the initial position (using harmonic restrain with force constant=1000 kJ mol^{-1} nm^{-2}) was performed for 2 ns under NPT ensemble. Integration time step of 1 fs was used. The system temperature was adjusted to 310 K by the V-rescale algorithm[8] and pressure to 1 bar by the Berendsen algorithm[9], with relaxation time constant of 0.1 ps and 0.5 ps, respectively. The system was partitioned into two temperature coupling groups: (1) Proteins, nucleic acids, Mg^{2+} ions and Zn^{2+} ions; (2) Water molecules and Na^+ ions. The initial velocities were assigned from the Maxwell-Boltzmann distribution at 310K. For non-bonded cutoffs, neighbour list was set to 1.2 nm, and PME (rcoulomb=1.2 nm) and switching L-J potential (rvdw=1.1 nm and rvdw_switch=1.0 nm) were used. All the bonds were constrained with SHAKE algorithm[10].

The final frame was used as the starting point for 10-ns production run using the same settings with integration time step of 2 fs. The system was annealed from 50 K to 310 K in the first 1 ns. System configurations were saved every 1000 steps. Totally, 30 trajectories each containing 5000 conformations were generated from different initial velocities to improve the sampling on the initial conformation. Reactant-like structure was manually picked up by measuring three distances that describe the suitability of the conformation as a reactant in the two-metal mechanism, which are (1) distance between the bridging oxygen (i.e. pro-Sp oxygen of U11) and tightly bound Mg_A (Mg_A-O_{U11_Sp}); (2) distance between the bridging oxygen and loosely bound Mg_B (Mg_B-O_{U11_Sp}); and (3) distance between the general base (*Rp* oxygen of G12) and the attacking nucleophile (O_{G12_Rp}-H_{W1}). Distance plots for the 30 trajectories were shown in Supplementary Fig. 16(a) and (b). The chosen structure was first

subjected to MM energy minimizations (AMBER12)[11], followed by QM/MM optimization (QChem/Amber)[12].

From the proposed two-metal-ion mechanism, a symmetric coordination between the central phosphate oxygen and the two catalytic metal ions would be expected. The 30 distances plots showed that in most of the trajectory, the coordination between the pro-Sp oxygen of U11 and the modelled Mg_B ion was stable during the 10 ns simulation. The coordination of Mg_A ion was also highly preserved. Therefore, the modelling of Mg_B with the mentioned approach was considered to be reasonable.

Supplementary Note 2. Validation of the initial catalytically active conformation. Analyses were performed to examine the stability of the modelled structure and the coordination of Mg_B in our MD simulations. As an indicator for the stability of our two-metal complex, calculations of our model structure were performed to determine the coordination of the bridging U11 pro-Sp oxygen (Sp-O) with both of the two Mg^{2+} ions. Both of these two coordination distances (Sp-O:Mg_A and Sp-O:Mg_B) were well preserved in the majority of MD trajectories (22 out of 30). Among these stable trajectories, the coordination of the two metal ions as well as the critical Mg_A-Mg_B distance were maintained in good condition. In particular, Mg_A coordination with all distances with nearby atoms well with 2.3 Å (Left panel of Supplementary Fig. 1(b)), all distances of Mg_B coordination with nearby atoms are well within 2.2 Å (Middle panel of Supplementary Fig. 1(b)), and Mg_A-Mg_B well within 4 Å critical for two-metal catalysis (Right panel of Supplementary Fig. 1(b)).

In addition, a systematic comparison of the coordination of Mg_B with an alternative structural model proposed by Sosunova E *et al.*[13] was performed. In this alternative model, Mg_B is in direct coordination with Pol II residue D483, which is drastically different from our model. Our simulations starting from this alternative model clearly indicate that the coordination of Mg_B is not stable with d_B (the distance between U11 pro-Sp oxygen and Mg_B) quickly increasing from 2 to 3.2Å (Supplementary Fig. 2(d)). More strikingly, our simulations further show that the incorrect binding model of Mg_B could induce instability of Mg_A binding (Supplementary Fig. 2(c)) and the distance between the two Mg^{2+} ions (Supplementary Fig. 2(b)). The above results provide additional support for our modelled structure containing Mg_B.

Supplementary Note 3. Alternative Mg_B Position. An alternative model of Mg_B Position had been proposed[13], where the major difference resides on the coordination mode of Rpb1-D483. In this model, D483 coordinates directly with Mg_B. Inspection on the provided coordinates suggested that Mg_A and the loop spanning from D481 to D485 was optimised, while all the nucleotides had their coordinates preserved in 3GTG. The coordinates of Mg_B was applied to the energy-minimised, one-metal-ion 3GTG structure by aligning the atoms of nucleotides. The system was then solvated with the same water box as used in our proposed

model to minimise the difference. The solvated system was subjected to 5000 steps of energy minimization with steepest descent, followed by 5000 steps with conjugate gradient. The latter step was added because the structure resulted from the former step was not stable in the 2 ns position restrained MD. The resulted equilibrated structure was shown with comparison to our model in Supplementary Fig. 17. Setup for position restrained and production MD were the same as described in Supplementary Note 1. Totally, 20 trajectories were produced. Distance plots for the two Mg^{2+} ions were shown in Supplementary Fig. 16(a) and (c). Since the coordination of both Mg_A and Mg_B to the pro-Sp oxygen of U11 were less stable, this alternative model was considered to be less favourable.

Supplementary Note 4. Obtaining minimum energy pathway with reaction coordinate driving method. In this method, a few variables were chosen as reaction coordinates to represent the slow motion of the reaction of interest. To drive the reaction across the high energy barrier, biased harmonic potential was imposed on the chosen reaction coordinates, while other degrees of freedom were minimised. The reactant state was first driven to the intermediate state, then to the product state. The same calculation was repeated in the reverse direction (i.e. from product state to reactant state). This '1D path-scan' calculation was repeated back-and-forth until it converged to confirm that a particular set of reaction coordinates is reasonable to describe the reaction. When 2D reaction coordinates are used, the corresponding 2D potential energy surface would be explored to identify the minimum energy pathway connecting the two energy minima. To construct the 2D surface, the y-axis was first generated by increasing y while restraining x at the minimum value. For all the generated points, x was increased while restraining y at the starting value. After obtaining the 2D potential energy surface, free energy calculation was performed along the minimum energy pathway. In this study, 1810 and 2862 points were used to construct the 2D surface for step 1 and 2, respectively.

Supplementary Note 5. Non-ideal phosphorane transition state geometry in two-metal mechanism. We discovered from our aiQM/MM-MD simulations that the phosphorane TS1 deviates from ideal trigonal bipyramidal geometry: The average value of apical O-P-O angle is 163.0±2.7° in TS1, compared with 180° for an ideal geometry. Similar structures were identified in the vanadate complex of several alkaline phosphatases (Supplementary Fig. 18). Largely distorted structure was found in the *Xac* nucleotide pyrophosphatase/phosphodiesterase (NPP)[14], where the angle formed by the vanadium and the two apical oxygen atoms is 157.5°. This distortion is seemingly caused by the topological restrain induced by simultaneous binding to two metal ions, because a value of 171.7° was found in *E. coli* alkaline phosphatase (AP)[15], where one of the apical oxygen locates slightly away from the coordinating Zn^{2+} ion. In the Autotaxin (ATX) complex[16], the O-V-O angle further increases to 179.35° when the apical oxygen is free from the Zn^{2+} ion.

Vanadate is a powerful transition state analogue for phosphoryl transfer reaction[17]. The above observations together with our results suggest that the transition state geometry may be altered in enzymatic reaction due to the catalytic metal ions involved. The power of vanadate may be attributed to the ability of forming distorted trigonal bipyramidal structures.

Apart from the apical O-P-O angle, deviation from ideal value also appears in the planarity between the equatorial oxygen atoms and the central phosphorus. The average torsion of the PO_3 group ($O5'$-O_{Sp}-P-O_{Rp}) of U11 is 160.26°, showing that inversion of configuration takes place before TS1, and this can be explained by the late-transition state nature of TS1 in our proposed reaction mechanism, in accordance with the Hammond-Leffler postulate[18].

Supplementary Note 6. Protonation of the leaving group by Mg^{2+}-coordinating water molecule. The A10 3'-oxyanion is protonated by a nearby metal-coordinating water molecule W3. In TS2, proton transfer from W3 to A10 O3' is in mid-way, where the average O_{A10}-H_{W3} and H_W-O_{W3} distances are 1.15±0.03 Å and 1.32±0.04 Å, respectively. The increasing hydroxyl character of W3 attracts the neighbouring W2 and triggers the subsequent proton transfer processes, leading to the final product containing neutral water molecules and unprotonated terminal phosphate.

Supplementary Note 7. Comparison with resolved product structure from another endonuclease. With advance crystallographic techniques, direct comparison with the true reaction product instead of a structure analogue is now possible. We selected the product structure captured in DNA/DNA TtAgo complex, which also contains 3'-OH and 5'-phosphate for comparison because it was crystallised with Mg^{2+} ions instead of the commonly used Mn^{2+}. Mn^{2+} ions are known to be more flexible than Mg^{2+} in coordination number and distance; therefore, comparison would be difficult due to the complexity. The average value of 3.72±0.05 Å in the unrestrained product structure is well consistent with the Mg-Mg distance of 3.69 Å in the pure cleavage product obtained in TtAgo (Supplementary Fig. 19). The bridging pro-Sp oxygen on the scissile phosphate still coordinates with both Mg_A and Mg_B in the identified PS, and the corresponding Mg-O distances of 2.11±0.07 and 2.15±0.10 Å also agree with the 1.97 Å and 2.27 Å found in TtAgo.

Supplementary Note 8. MD Simulation with Protonated Asp 485. The same set up and procedure were used as Supplementary Note 1, except that the protonation state of D485 was predefined as singly protonated.

Supplementary Note 9. MD Simulation with 2'-5' modification. The energy minimised one-Mg^{2+} structure obtained in the original model was used to construct the model with 2'-5' linkage. To link G12 to the 2' oxygen of U11, the whole G12 nucleotide was shifted to the

position where the P atom overlays the U11 2'-OH atom. To carefully relieve the steric clashes, only the bridge helix (BH) residues (Rpb1 812 to 844) and trigger loop (TL) residues (Rpb1 1067 to 1106) were relaxed using 50,000-step steepest descent plus 50,000-step conjugate gradient. The position of Mg_B was modelled using the same procedure as the original model. The solvated system was again subjected to 50,000-step energy minimization using steepest descent and conjugate gradient, respectively. The system was slowly relaxed by reducing the harmonic position-restraint step-by-step. The force constant was reduced from 1000 to 100, finally to 10 kJ mol^{-1} nm^{-2} in 2 ns interval. The final frame was used as the starting point for 10-ns production run using the same settings as above. Totally, 20 trajectories were generated.

Based on these MD simulations, we have computed the distances to examine if the 2'-5' linkage induces noticeable changes in the active site architecture. As shown in Supplementary Fig. 10(a), we plotted a number of important distances in the active site including distances related to the coordination of Mg_A and Mg_B, Mg_A-Mg_B distance, and water coordination. As shown in Supplementary Fig. 10(b), the only major change is the distance of the phosphate to the nucleophilic water (W1-Rp-O), which in turn marginally affect the Mg_A-Mg_B distance (<0.1 Å). Our results strongly suggest that the active-site architecture remains nearly intact with the change from the 3'-5' to 2'-5' linkage.

Supplementary Note 10. Calculating the average distance between active site atoms for the original 3'-5' model and the 2'-5' modification model. Only reactant-like structures were included in the calculation. Reactant-like structures were those preserving both Mg_A-O_{U11_Sp} and Mg_B-O_{U11_Sp} coordination, using 2.5 Å as the cut-off value. 22 out of 30 trajectories and 14 out of 20 trajectories were used for analysis from the original model and the 2'-5' modification model, respectively. The first 7 ns data was considered as equilibration and only the last 3 ns data was used here (see Supplementary Note 9 for MD simulation details).

Supplementary Note 11. Validation of energy maxima. Unrestrained aiQM/MM-MD simulation was used to validate the identified TS1 and TS2. The idea is similar with that of intrinsic reaction coordinate (IRC) calculation: a meaningful transition state should collapse into one minimum with forward transition vector and into the other minimum with reverse vector. If the structure sampled in aiQM/MM-MD simulation locates near TS region, relaxation (i.e. without imposing restraints) should direct the system to one state with forward velocity and to another with reverse velocity. To validate the identified TS, the simulation on TS window was extended by 1ps, while checkpoint files containing coordinates and velocities were saved every 10 fs. 20 frames were randomly selected for validation by performing 1 ps aiQM/MM-MD simulation without restraints. There are 6 and 3 trajectories successfully collapsed into the two minima in step 1 and step 2, respectively. Projection on the 2D PMF

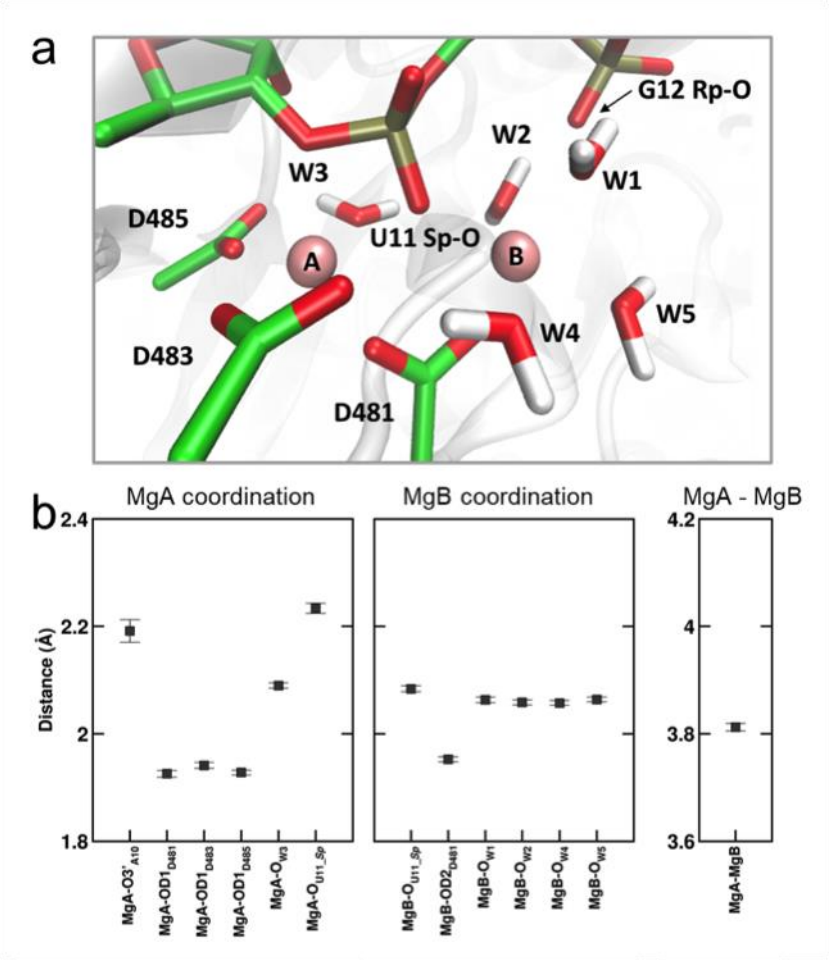
surfaces shows that these trajectories follow faithfully along the predicted energy landscape (Supplementary Fig. 5). This observation validates the identified TS.

Supplementary Note 12. Validation of energy minima. Unrestrained aiQM/MM-MD simulation was used to validate the identified RS, IS and PS. The idea is that a true minimum should be stable in the corresponding sampling region upon relaxation. The same procedure as above was used. Projection on the 2D PMF surfaces shows that all the trajectories are stable around the sampling region (Supplementary Fig. 6). This strongly suggests that the energy minima we identified are reliable.

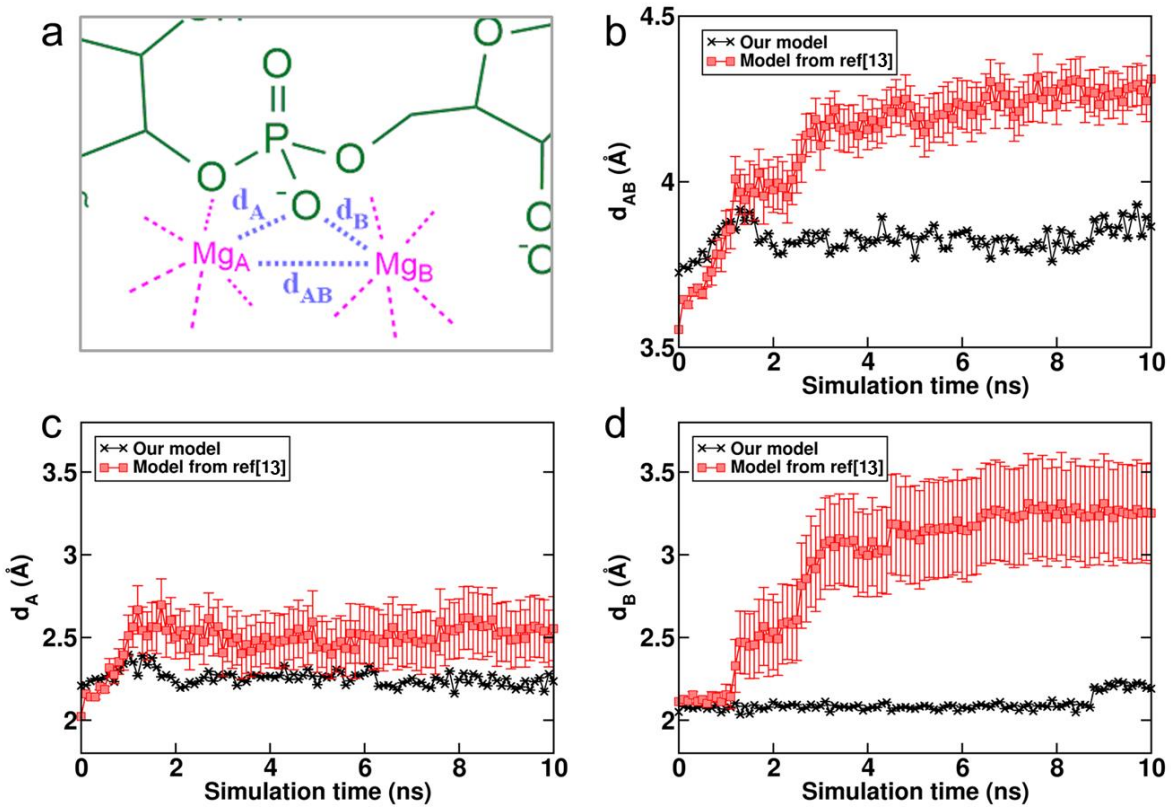
Supplementary Note 13. Exclusion of other plausible general bases. Atoms in close proximity that can potentially serve as the alternative general base to activate the attracting nucleophile were systematically surveyed (Supplementary Fig. 8). We define general bases as electron rich atoms close by the nucleophile in our system. These electron-rich atoms are: (a) terminal nucleotide Rp-O, (b) N7 atom in G12 nucleotide, (c) U11 Rp-O and (d) G12 O6. Supplementary Fig. 8(a) shows the minimum distance and Supplementary Fig. 8(b) displays the coordination between the attacking water molecule and G12 pro-Rp oxygen in comparison with three alternative electronegative atoms in the active site. Our QM/MM calculation results clearly indicate that all these three alternative bases (N7, O6, and U11 pro-Rp) are substantially less favourable in potential energy for endonucleolytic cleavage than our proposed general base bearing the G12 pro-Rp oxygen (Supplementary Fig. 9).

*Exclusion an alternative general base: N7 atom in the terminal G12 nucleotide***:** We performed QM/MM calculations to determine the potential energy profiles of the endonucleolytic cleavage involving G12 N7 in comparison with G12 Rp-O to exclude the possibility of this N7 atom at G12 to serve as general base. The initial reactant-like structure with N7 serving as general base was chosen using the similar criteria as stated in Supplementary Note 1: (1) Distance between the bridging oxygen (i.e. pro-Sp oxygen of U11) and tightly bound Mg_A (Mg_A-O_{U11_Sp}); (2) distance between the bridging oxygen and loosely bound Mg_B (Mg_B-O_{U11_Sp}); and (3) distance between the general base (N7 atom of G12) and the attacking nucleophile (O_{G12_N7}-H_{W1}). As shown in Supplementary Fig. 9(a), N7 atom serving as general base will increase the activation energy barrier by 20 kcal/mol, rendering the system inactive. A stable/metastable intermediate state could not be located with unrestrained optimisation, indicating that G12 N7 has a weak ability to deprotonate the attacking nucleophile. Furthermore, we have performed experiments involving N7 substitution to 7-deaza-G in intrinsic cleavage assay to show that removal of N7 atoms did not affect cleavage rate at 5 mM Mg^{2+} condition, suggesting that N7 does not participate in intrinsic cleavage (Supplementary Fig. 12).

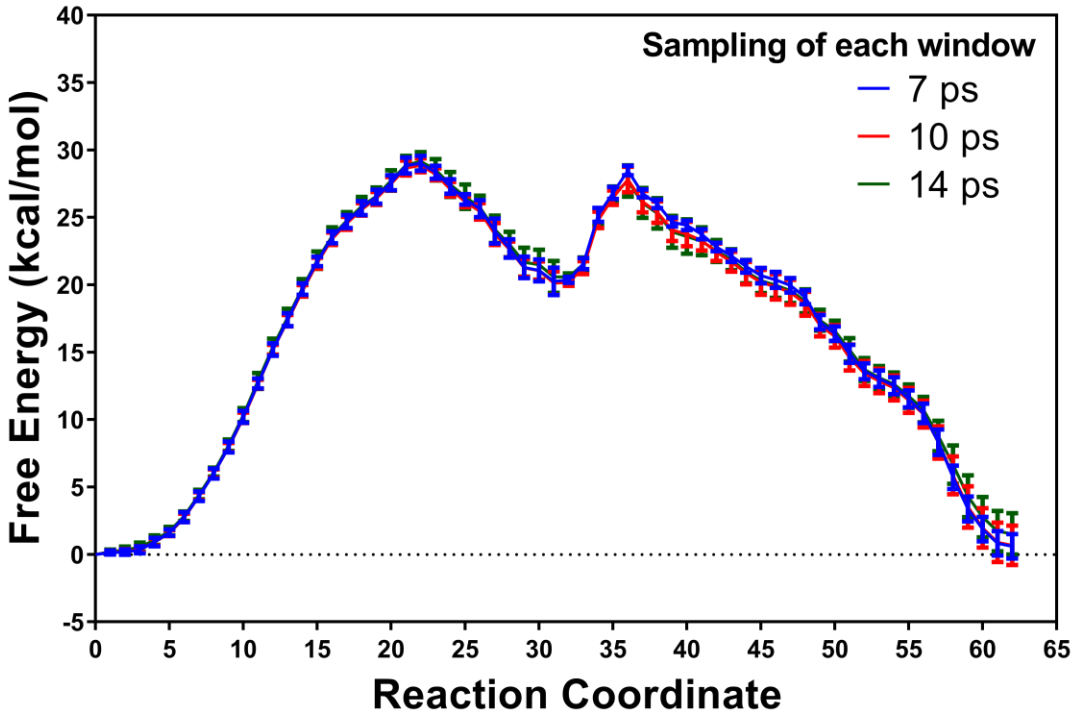
Exclusion two more alternative general bases: U11 Rp-O and G12 O6. We performed QM/MM calculations to obtain energy profiles of U11 Rp-O and G12 O6 of the terminal nucleobase serving as general base. The initial reactant-like structure with O6 serving as general base was chosen using the similar criteria as stated in Supplementary Note 1: (1) Distance between the bridging oxygen (i.e. pro-Sp oxygen of U11) and tightly bound Mg_A (Mg_A-O_{U11_Sp}); (2) distance between the bridging oxygen and loosely bound Mg_B (Mg_B-O_{U11_Sp}); and (3) distance between the general base (O6 atom of G12) and the attacking nucleophile (O_{G12_O6}-H_{W1}). The initial reactant-like structure with U11 Rp-O as general base is the same as G12 Rp-O. Our results show that potential energy barriers with U11 Rp-O and G12 O6 serving as alterative general base are much higher compared to G12 Rp-O by ~20 kcal/mol (see Supplementary Fig. 9(b)) and 15 kcal/mol (see Supplementary Fig. 9(c)), respectively. For U11 Rp-O, our results show that ΔE of the unrestrained intermediate state for U11 Rp-O is +47.88 kcal/mol compared with +30.66 kcal/mol of G12 Rp-O. This further supports that G12 Rp-O is a stronger general base despite it has a larger average distance with the attacking nucleophile than U11 Rp-O. Similar with G12 N7, a stable/metastable intermediate state could not be located for G12 O6.



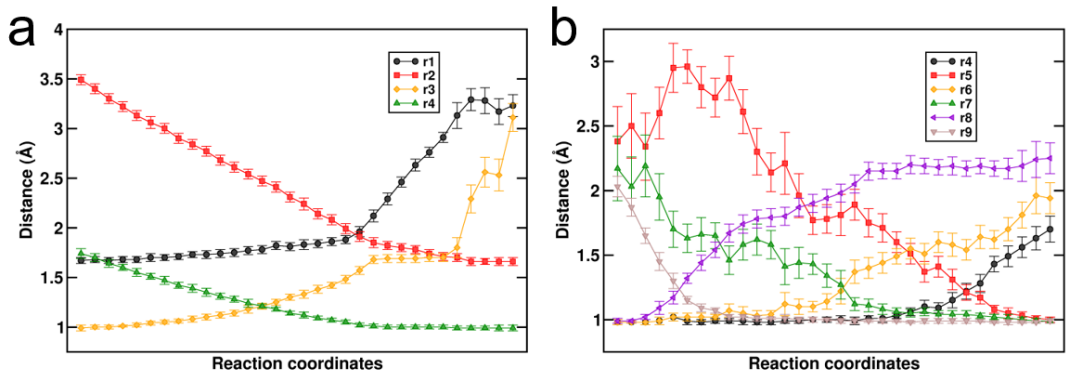
Supplementary Figure 1. The modelled reactant-like structure is stable in MD simulations based on Mg_A, Mg_B, and Mg_A-Mg_B coordination. (a) Nomenclature of the active-site atoms. (b) Coordination distances measured based on reactant-like structures (defined as structures retaining coordination of the bridging U11 pro-Sp oxygen with both Mg_A and Mg_B). Although not being included in the selection criteria, all the other critical coordination is stable in our two-metal model. The Mg_A-Mg_B distance is within the proposed 4Å to enable the general two-metal catalytic reactions. The error bars are reported as sample standard deviations obtained from distances computed from an ensemble of MD conformations as described in Supplementary Note 2.



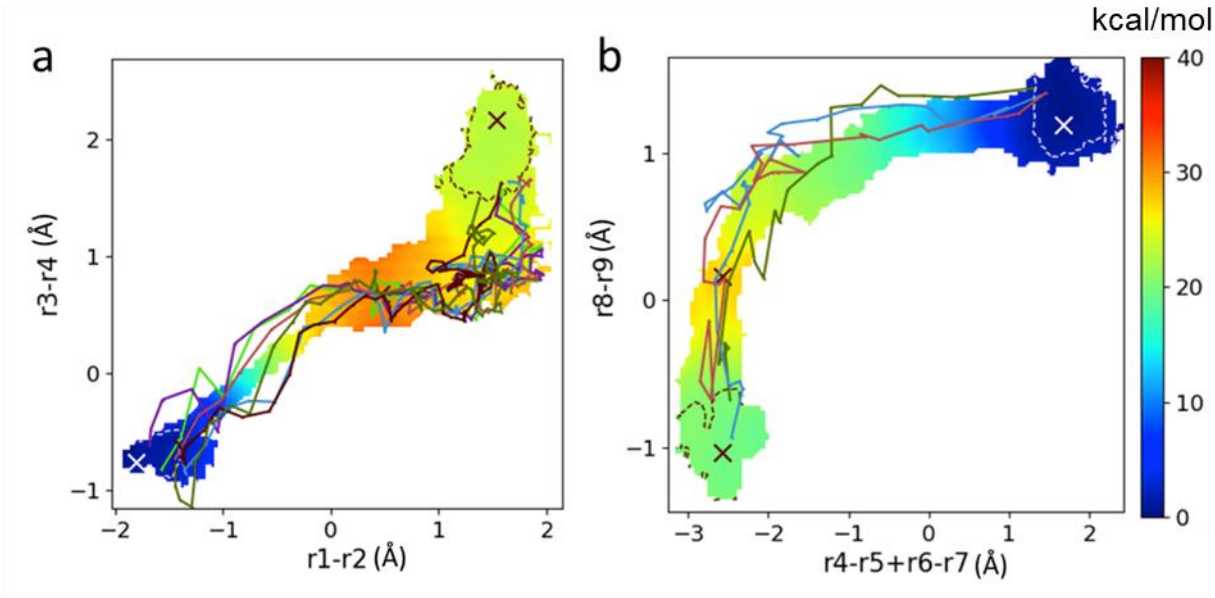
Supplementary Figure 2. Our model greatly outperforms an alternative model from Sosunova E *et al*. [Ref 13] for the initial structural modelling of MgB. (a) A reactant-like structure is indicated by stable coordination of the bridging U11 pro-Sp oxygen with both Mg_A and Mg_B (d_A and d_B) as well as ~4 Å distance between the two metal ions (d_{AB}). (b-d) Distance plot of our model versus alternative model sampled at 100 ps interval. Totally 30 and 20 trajectories were used for analysis from the original model and the alternative model, respectively. The error bars refer to the sample standard deviation computed for the average distance based on 30 (original model) and 20 (alternative model) MD simulation trajectories. Our model demonstrates higher stability for d_A, d_B and d_{AB} compared with reference[13]. Note that the average d_{AB} of the alternative model proposed by reference[13] exceeds 4 Å at the early stage of the MD simulations.



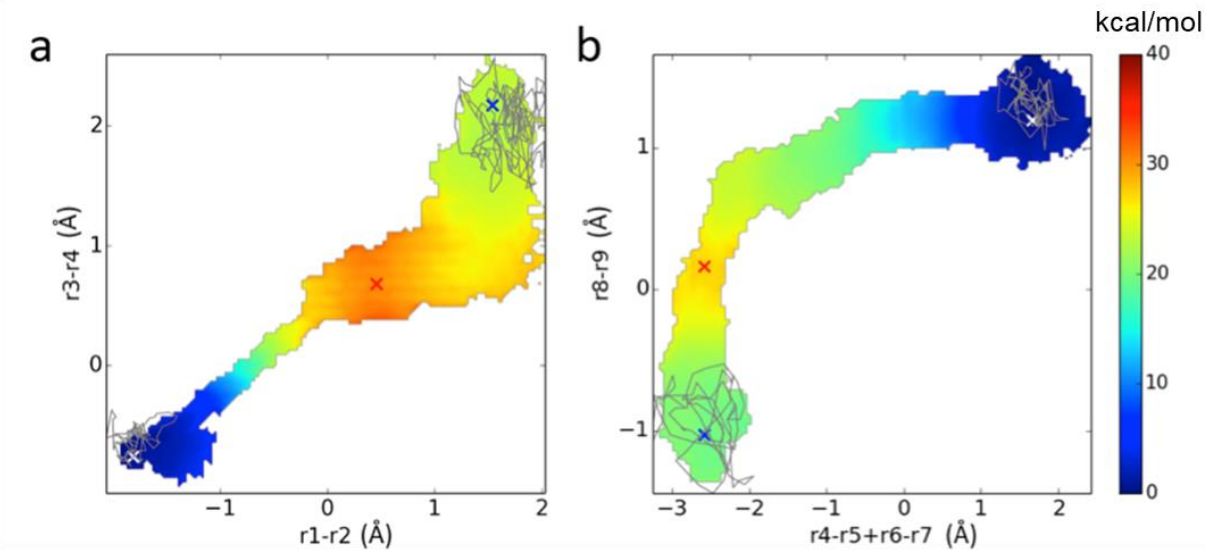
Supplementary Figure 3. The free energy profiles obtained from the umbrella sampling of aiQM/MM-MD simulations with various amount of sampling at each window are converged. Free energy profile computed from umbrella sampling using aiQM/MM-MD simulations at the length of 7 ps (4 to 11 ps), 10 ps (4 to 14 ps), and 14 ps (4 to 18 ps) at each window yield consistent results. For each data set, the simulation data was dissected with 1-ps segment and the standard deviation of the resulting free energy profiles was used to compute the error bars. The same reaction coordinate as in Figure 2a is adopted to plot the free energy profiles.



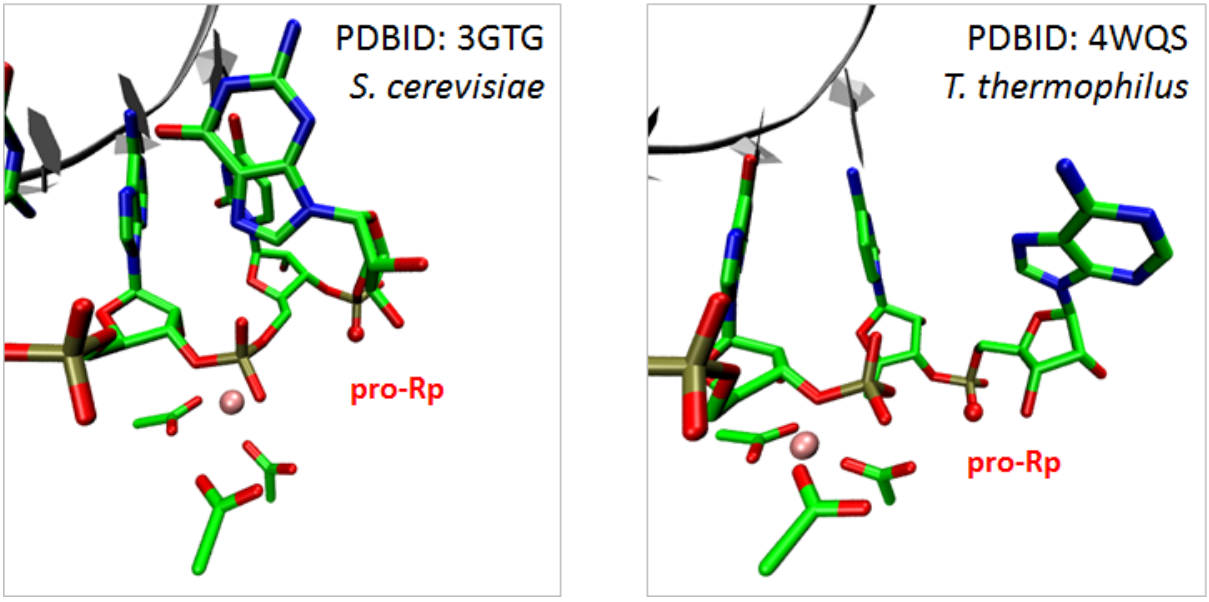
Supplementary Figure 4. Monotonic bond distance changes along the reaction coordinates of step 1: endonucleolytic cleavage (a) and step 2: proton transfer (b) suggest that the complex reaction coordinates do not cause degeneracy problem. The average value is calculated from the conformation sampled with 4-11 ps in aiQM/MM-MD simulations along 1D PMF reaction pathway. The error bars refer to the standard deviation of the corresponding data set used for calculating the average distance.



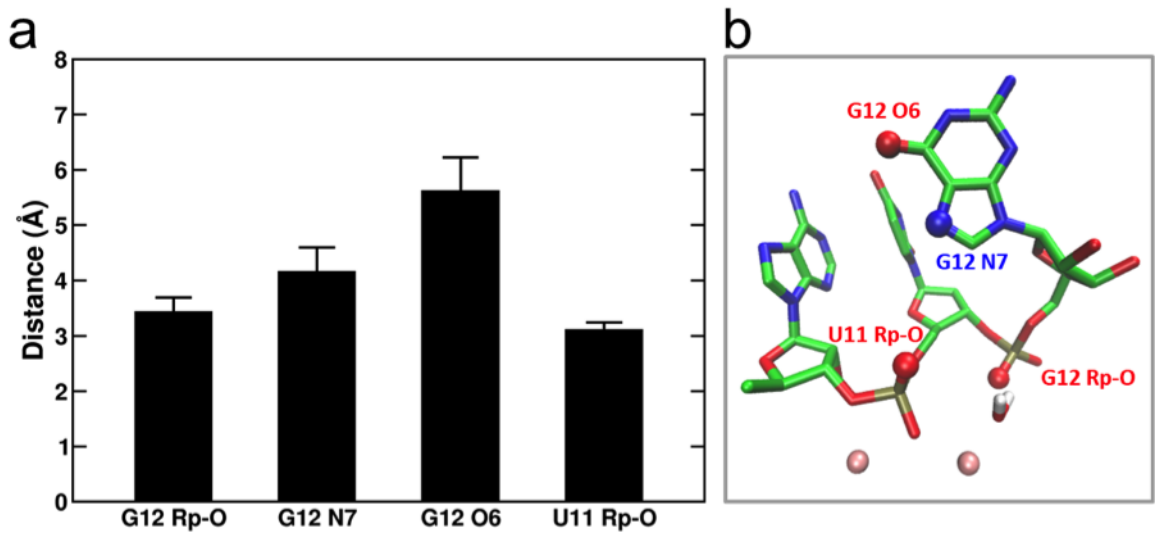
Supplementary Figure 5. Validation of TS by unrestrained aiQM/MM-MD simulation. All validation simulations are initialised around the TS region. The crosses indicate the positions of the minima and transition states. The dash line is the energy contour line that is 1 kcal above the corresponding crosses, calculated based on the 2D PMF. For visualization purpose, the projection is truncated once the trajectory reaches the region enclosed by the dash line, where we consider the local energy minimum is reached. (a) For TS1 in the endonucleolytic cleavage (step 1), 6 out of 20 initial structures reach the two minima in forward and reverse directions within 1 ps, and their corresponding traces are shown. (b) For TS2 in the proton transfer (step 2), 3 out of 20 initial structures reach the two minima in forward and reverse direction within 1 ps, and their corresponding traces are shown.



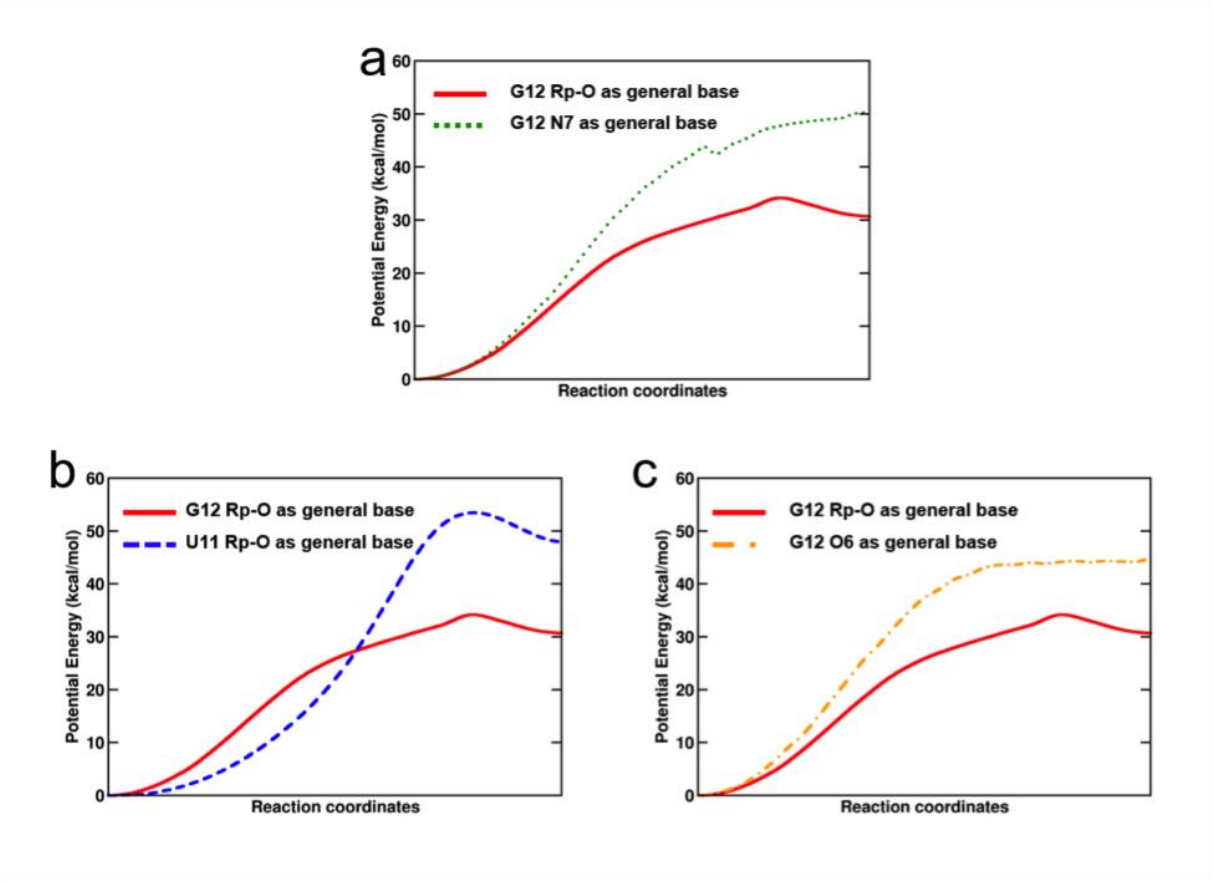
Supplementary Figure 6. Validation of RS, IS and PS by unrestrained aiQM/MM-MD simulation. The trajectories are projected onto the 2D PMF of (a) step 1: endonucleolytic cleavage and (b) step 2: proton transfer.



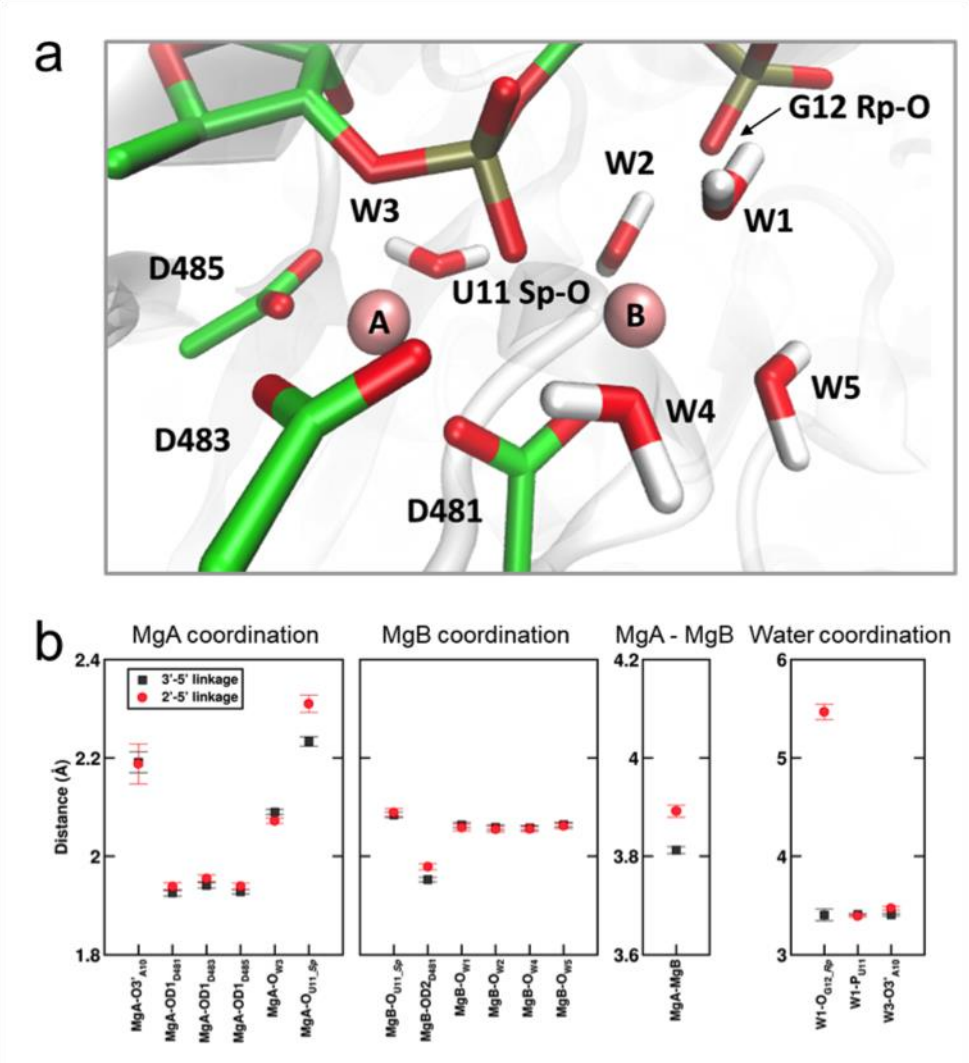
Supplementary Figure 7. The one-nucleotide-backtracked structures of *S. cerevisiae* Pol II[4] and *T. thermophilus* RNAP[19]. Both of them contain a twisted backbone that places the pro-Rp atom (highlighted sphere) of the terminal misincorporated nucleotide in close proximity to the active site. Mg_A in 4WQS is modelled.



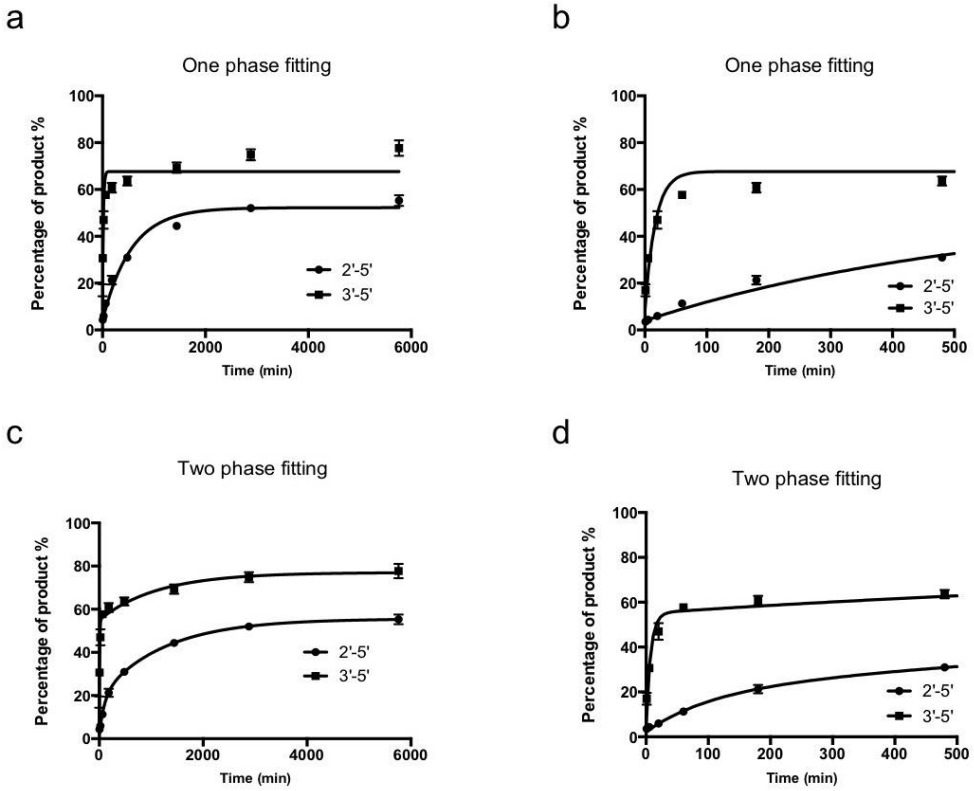
Supplementary Figure 8. Alternative general bases: (a) The minimum distance between the attacking water molecule and G12 pro-Rp oxygen in comparison with three alternative electronegative atoms in the active site based on reactant-like structures: G12 N7, G12 O6, and U11 pro-Rp O atoms. The error bars refer to sample standard deviations of distances obtained from an ensemble of MD conformations as described in Supplementary Note 13. (b) Structure of the active site with positions of G12 pro-Rp (Rp-O) and the three alternative general bases (G12 N7, U11 Rp-O and G12 O6) highlighted in spheres.



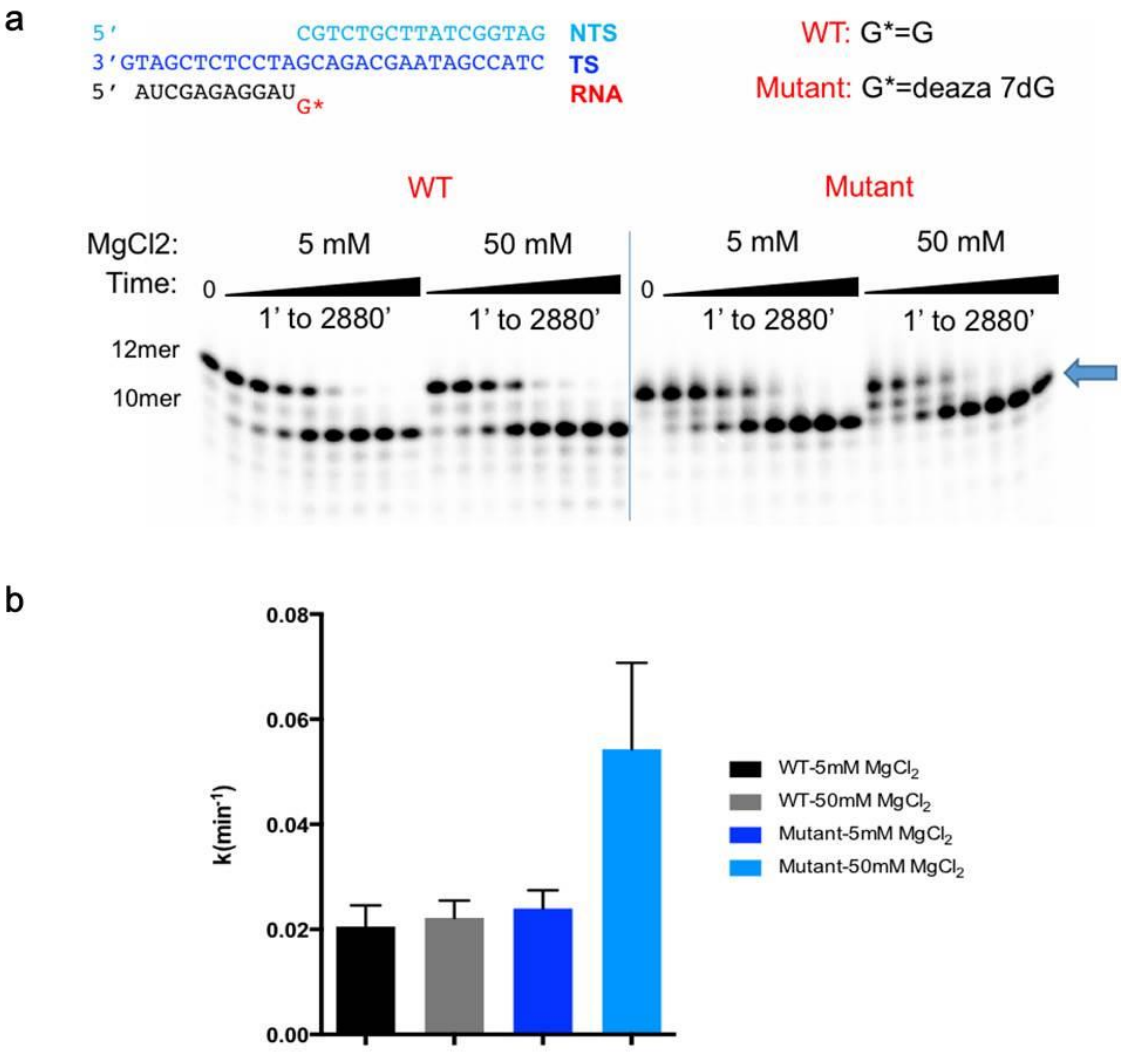
Supplementary Figure 9. QM/MM calculations to exclude three alternative general bases: (a). The potential energy curves for endonucleolytic cleavage (Step 1) are compared between calculations using general base G12 Rp-O (red) and G12 N7 (green). (b) The same as (a) except that G12 Rp-O (red) and U11 Rp-O (blue) are compared. (c) The same as (a) except that G12 Rp-O (red) and G12 O6 (orange) are compared. The reaction coordinates are chosen as RC = $d_{O(W1)-H}$-$d_{O(U11\_Rp)-H}$-r2 for U11 Rp-O, RC = d (r1-r2, r3-d_{H-O6}) for G12 O6, RC = d (r1-r2, r3-r4) for G12 Rp-O, and RC = d (r1-r2, r3-d_{H-N7}) for G12 N7. In the case of 2D RC used for G12 O6, G12 Rp-O and G12 N7, the two variables were increased with fixed step size, representing a diagonal reaction pathway on the respective 2D potential energy surface. Please refer to the Methods section for the definition of r_1 to r_4.



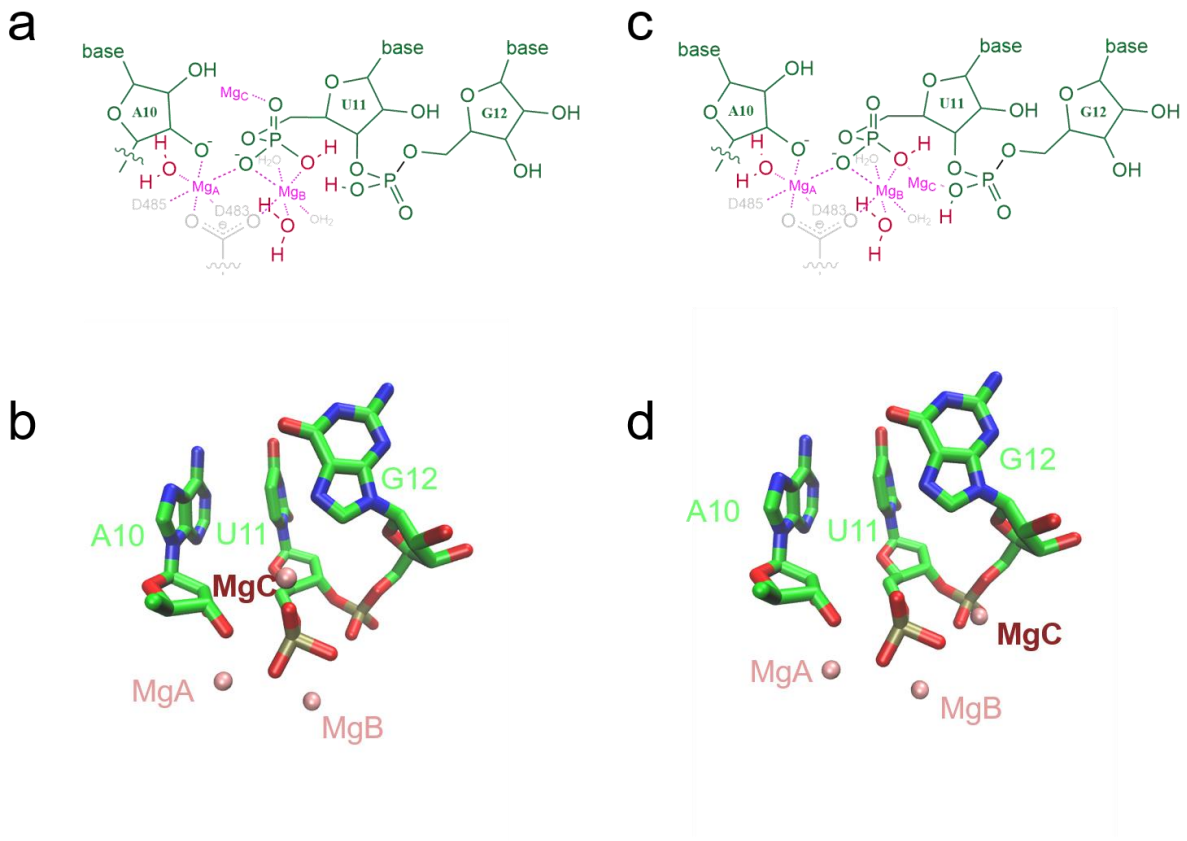
Supplementary Figure 10. 2'-5' modification causes limited changes in the active site architecture. (a). Nomenclature of the active-site atoms. Lower panel: Distances measured based on reactant-like structures (defined as structures retaining coordination of the bridging U11 pro-Sp oxygen with both MgA and MgB). (b). The average distance between the general base (G12 Rp-O) and the attacking nucleophile (W1) increases from 3.40 Å in 3'-5' linkage model to 5.47 Å in 2'-5' linkage model, while other distances show limited difference (<0.1 Å). The population of trajectories preserving both Mg_A and Mg_B bindings are 73.3% (22 out of 30) and 70% (14 out of 20) for 3'-5' and 2'-5' linkages, respectively. The error bars are reported as sample standard deviations obtained from distances computed from an ensemble of MD conformations as described in Supplementary Note 9.



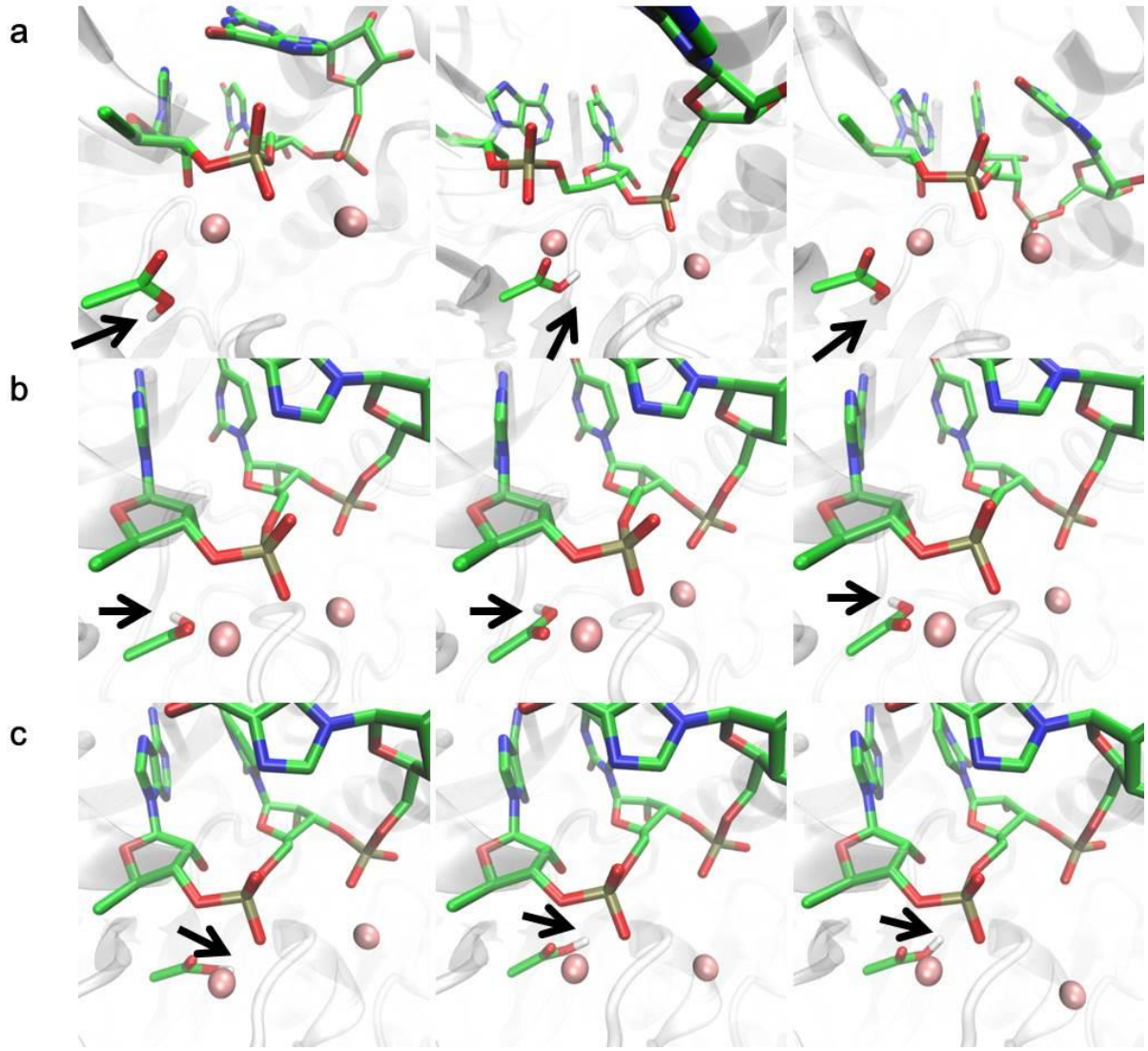
Supplementary Figure 11. Fitting curves of intrinsic cleavage product by two different fitting methods (One phase (a) vs. two phase (c)). Zoom in of the early time points are shown on the right (b, d). Mean and standard deviations (error bars) of 3 independent experiments are reported. We note that after labelling 5'-RNA, we observed a minor portion of RNA is resistant to intrinsic cleavage. Several factors could attribute to this: 1) A portion of Pol II adopted a different conformation that is resistant to intrinsic cleavage, 2) Not every RNA was assembled into Pol II complex.



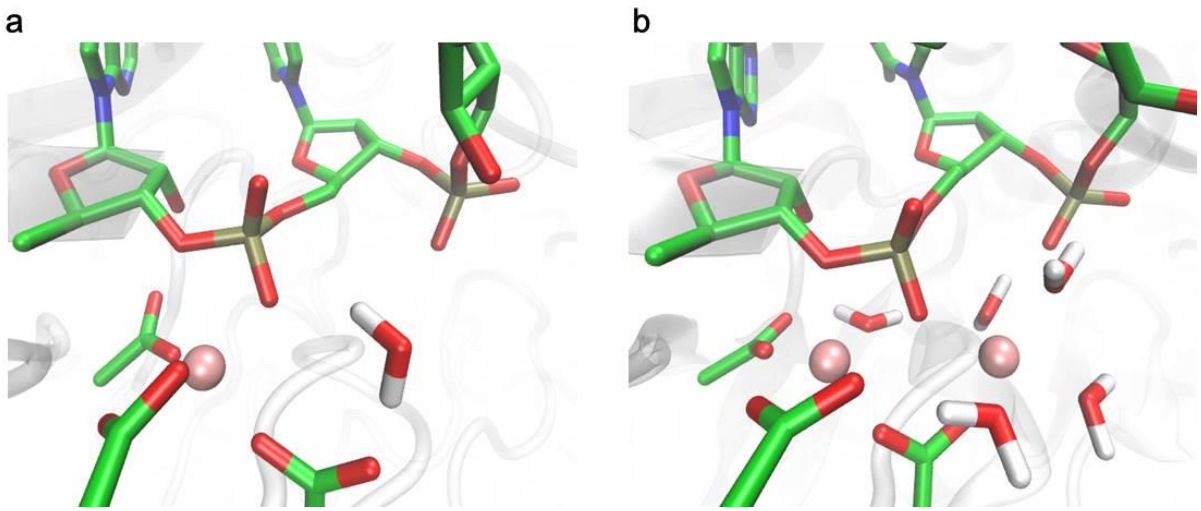
Supplementary Figure 12. (a) Sequence of the scaffold used in the intrinsic cleavage assay (top left). The results show methyl-substitution of N7 atom impose limited effect on cleavage rate. (b) The plot of RNA cleavage percentage against time. Mean and standard deviations (error bars) of 3 independent experiments are reported.



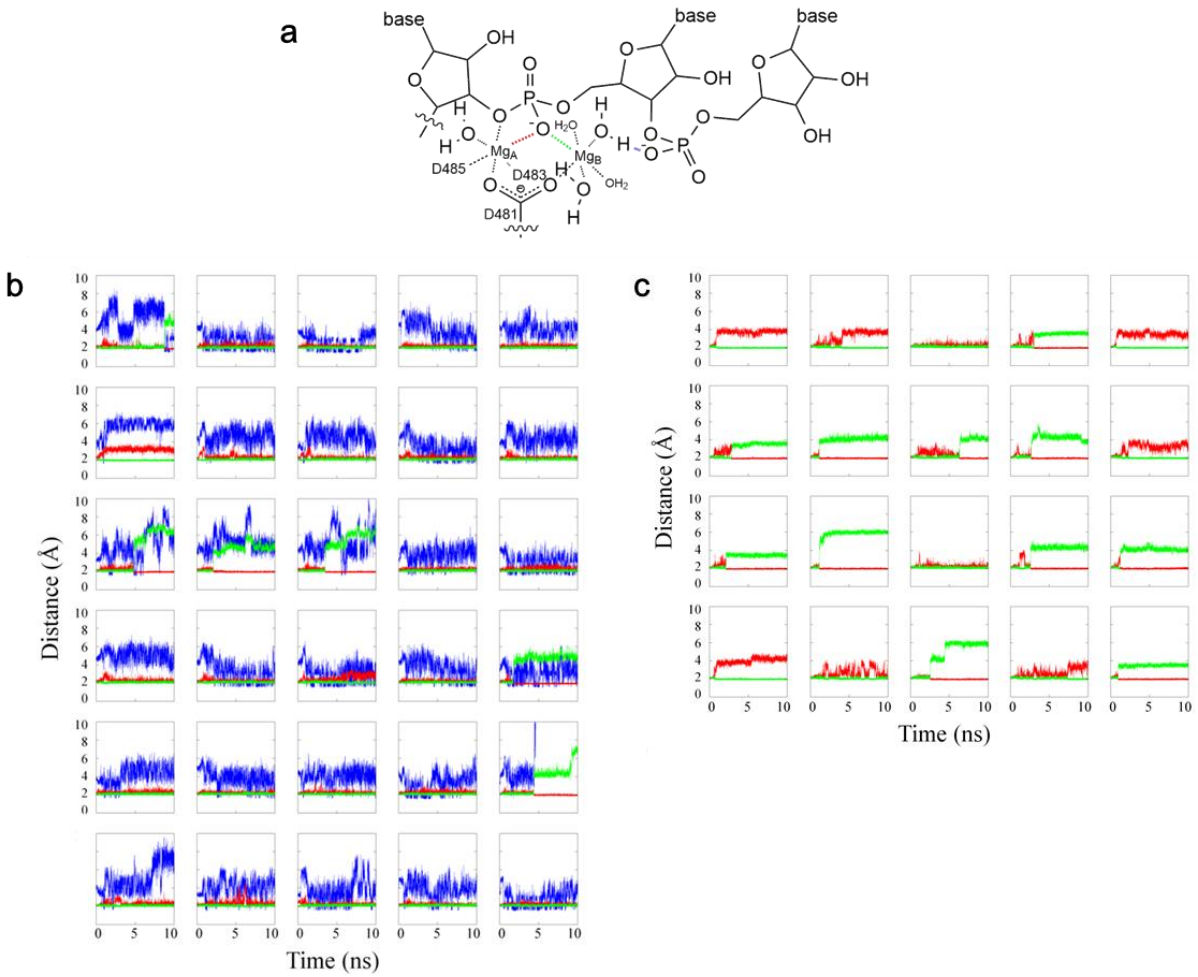
Supplementary Figure 13. Two potential binding sites of the third Mg^{2+} in the active site. Diagram and modelled active site structure of RNA hydrolysis catalysed by Pol II with Mg_C positioned at (a, b) the cleaved phosphate and (c, d) near the general base G12 Rp-O to activate the nucleophilic water. For (b), one water molecule coordinating with the non-bridging phosphate oxygen of the cleaved phosphate was replaced by the third Mg^{2+}. The modelled structure was then subjected to MM energy minimisation (AMBER12)[11] with all QM atoms frozen, followed by QM/MM optimisation (QChem/Amber)[12] without restraints. Subsequent QM/MM calculations indicate that Mg_C can stabilise the overall energy profile. Specifically, unrestrained QM/MM energy minimisation shows that introduction of Mg_C lowers the potential energy change from reactant to product state by 3.4 kcal/mol. For (d), our product state structure was overlaid with a X-ray structure (PDB ID: 6DP7)[20] by the alignment of the U11-G12 backbone atoms to generate three-Mg^{2+} model. The modelled structure was then subjected to MM energy minimisation (AMBER12)[11]. All the QM atoms, the modelled third Mg^{2+} ion was frozen in the first step, while only the QM atoms remain frozen in the second step. The system was then subjected to QM/MM optimisation (QChem/Amber)[12].



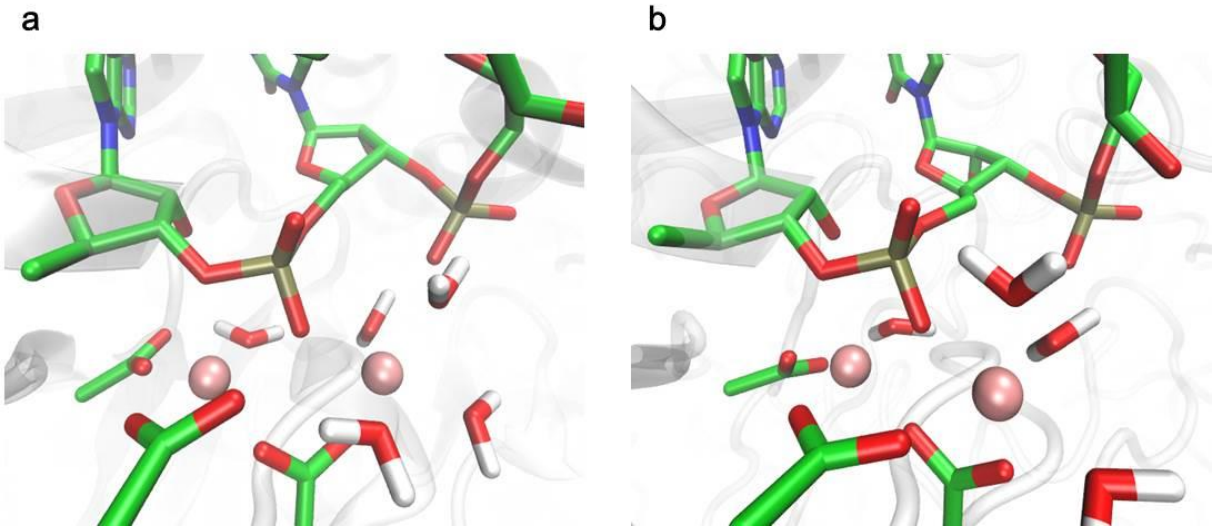
Supplementary Figure 14. Three set of MD simulations modelled with protonated D485: (a) Production run with 6-coordinated Mg_A; (b) Position-restrained simulation with 5-coordinated Mg_A; and (c) Position-restrained simulation with 5-coordinated Mg_A and additional position-restraint on the proton. The black arrows indicate the position of the flipped proton. In (c) the Mg-Mg distance increase to 5 Å.



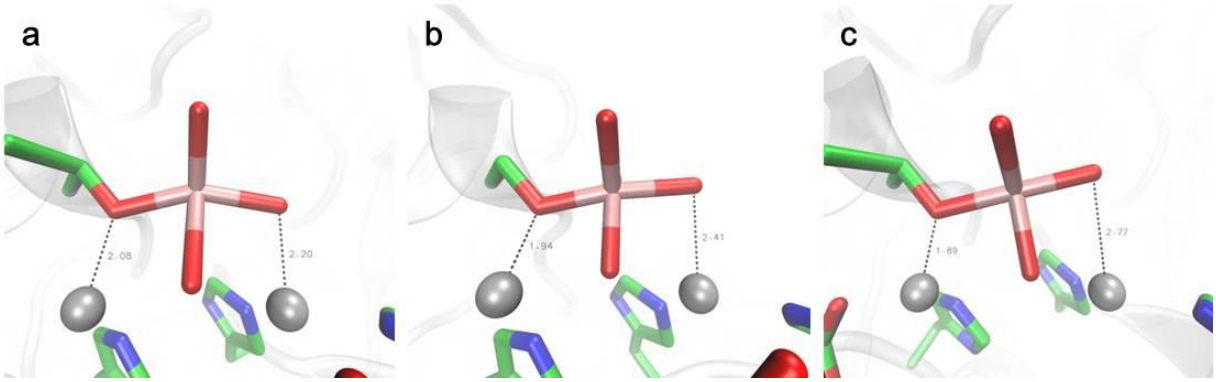
Supplementary Figure 15. Modelling of Mg_B by replacing an active site water molecule identified in the simulation box. (a) The bridging water molecule identified after energy minimization. (b) Active site configuration after equilibration.



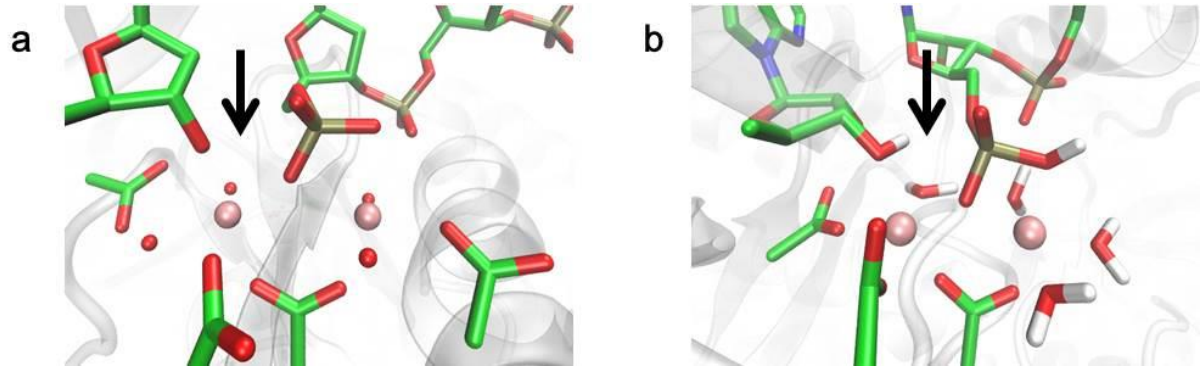
Supplementary Figure 16. (a) Three distances chosen as a measurement of coordination stability. (b) Plots of distances in (a) are displayed as a function of simulations time for 30 independent MD trajectories (Supplementary Note 1). (c) Similar plot as (b) for 20 independent MD trajectories using an alternative model from a previous study[13] (Supplementary Note 3).



Supplementary Figure 17. Comparison between the equilibrated structure of (a) our model and (b) model from a previous study[13].



Supplementary Figure 18. (a) VO_5 structure from *Xac* nucleotide pyrophosphatase/phosphodiesterase (PDBID: 2GSO)[14]; (b) *E. coli* alkaline phosphatase (PDBID: 1B8J)[15] and (c) Autotaxin (PDBID: 5IJS)[21]. Only the first molecule in the unit cell is shown.



Supplementary Figure 19. Comparison between cleavage product resolved in (a) DNA/DNA *T. thermophilus* Argonaute protein (PDBID: 4NCB)[22] and (b) relaxed product structure. Black arrows indicate the location of cleaved phosphodiester bond.

Supplementary References

1. Pronk, S. *et al.* GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
2. Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012 (2003).
3. Silva, D.-A. *et al.* Millisecond dynamics of RNA polymerase II translocation at atomic resolution. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7665–7670 (2014).
4. Wang, D. *et al.* Structural basis of transcription: Backtracked RNA polymerase II at 3.4 Ångstrom resolution. *Science* **324**, 1203–1206 (2009).
5. Pérez, A. *et al.* Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys. J.* **92**, 3817–3829 (2007).
6. Allnér, O., Nilsson, L. & Villa, A. Magnesium ion–water coordination and exchange in biomolecular simulations. *J. Chem. Theory Comput.* **8**, 1493–1502 (2012).
7. Molina, R. *et al.* Visualizing phosphodiester-bond hydrolysis by an endonuclease. *Nat. Struct. Mol. Biol.* **22**, 65–72 (2014).
8. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
9. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
10. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. . Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
11. Case, D. A. *et al.* AMBER 12. (2012).
12. Zhou, Y., Wang, S., Li, Y. & Zhang, Y. Born-Oppenheimer ab initio QM/MM molecular dynamics simulations of enzyme reactions. *Methods Enzymol.* **577**, 105–118 (2016).
13. Sosunova, E., Sosunov, V., Epshtein, V., Nikiforov, V. & Mustaev, A. Control of transcriptional fidelity by active center tuning as derived from RNA polymerase endonuclease reaction. *J. Biol. Chem.* **288**, 6688–6703 (2013).
14. Zalatan, J. G., Fenn, T. D., Brunger, A. T. & Herschlag, D. Structural and functional comparisons of nucleotide pyrophosphatase/phosphodiesterase and alkaline phosphatase: Implications for mechanism and evolution. *Biochemistry* **45**, 9788–9803 (2006).
15. Holtz, K. M., Stec, B. & Kantrowitz, E. R. A model of the transition state in the alkaline phosphatase reaction. *J. Biol. Chem.* **274**, 8351–8354 (1999).
16. Hausmann, J. *et al.* Structural snapshots of the catalytic cycle of the phosphodiesterase Autotaxin. *J. Struct. Biol.* **195**, 199–206 (2016).
17. Davies, D. R. & Hol, W. G. J. The power of vanadate in crystallographic investigations of phosphoryl transfer enzymes. *FEBS Lett.* **577**, 315–321 (2004).
18. Hammond, G. S. A correlation of reaction rates. *J. Am. Chem. Soc.* **77**, 334–338 (1955).

19. Sekine, S., Murayama, Y., Svetlov, V., Nudler, E. & Yokoyama, S. The ratcheted and ratchetable structural states of RNA polymerase underlie multiple transcriptional functions. *Mol. Cell* **57**, 408–421 (2015).
20. Samara, N. L. & Yang, W. Cation trafficking propels RNA hydrolysis. *Nat. Struct. Mol. Biol.* **25**, 715–721 (2018).
21. Hausmann, J. *et al.* Structural snapshots of the catalytic cycle of the phosphodiesterase Autotaxin. *J. Struct. Biol.* **195**, 199–206 (2016).
22. Sheng, G. *et al.* Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 652–657 (2014).