Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Electronic Supplementary Information for

Unraveling cGAS catalytic mechanism upon DNA activation through molecular dynamics simulations

Jordi Soler[1], Pedro Paiva[2], Maria J. Ramos[2], Pedro A. Fernandes[2], and Marie Brut[3]

[1] Institute of Computational Chemistry and Catalysis, and Department of Chemistry, University of Girona, Girona, Spain.

[2] LAQV@REQUIMTE, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Porto, Portugal.

[3] LAAS-CNRS, Université de Toulouse, CNRS, UPS, Toulouse, France.

Table of contents.

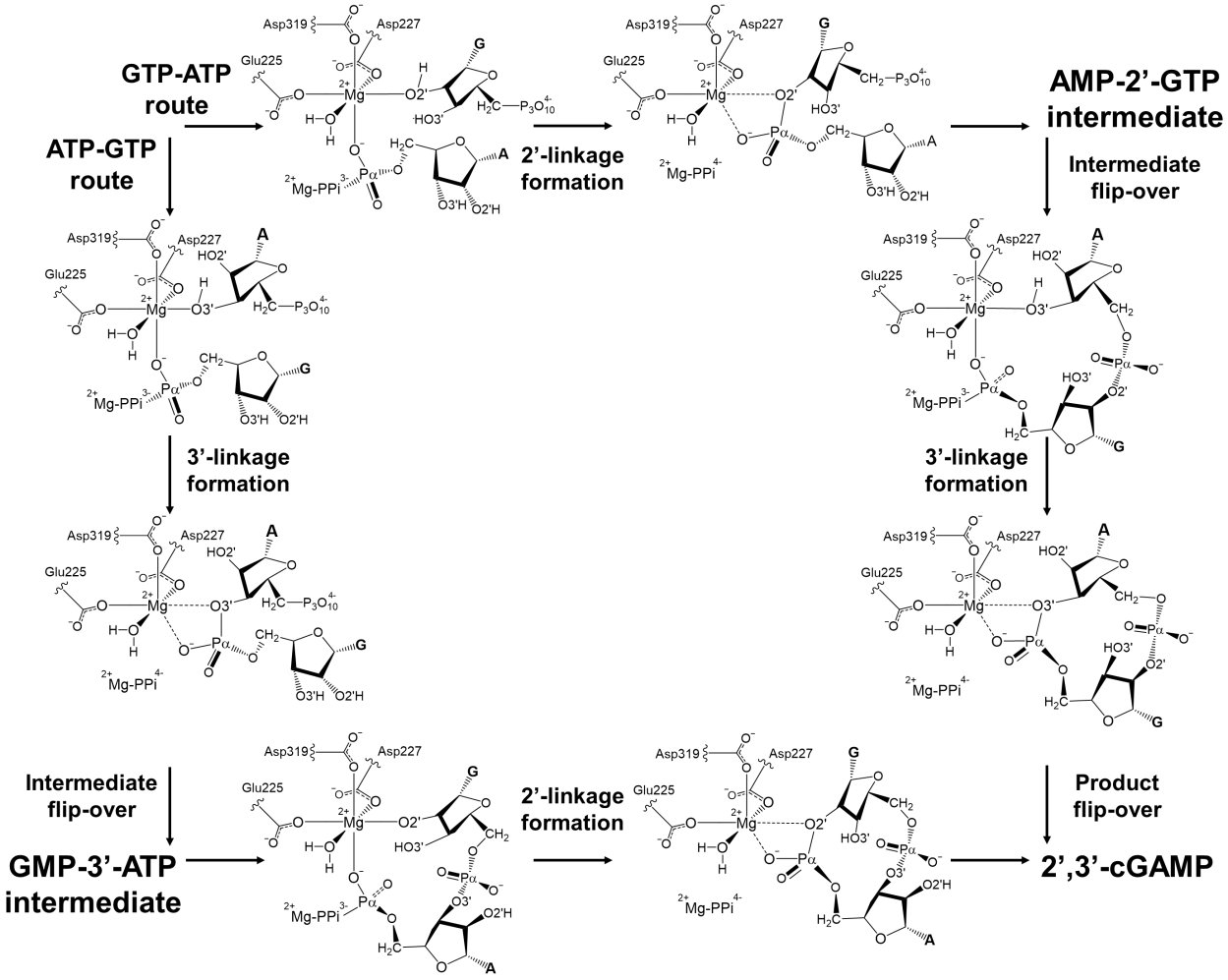


Figure S1. Extended scheme for the GTP-ATP and ATP-GTP catalytic routes.

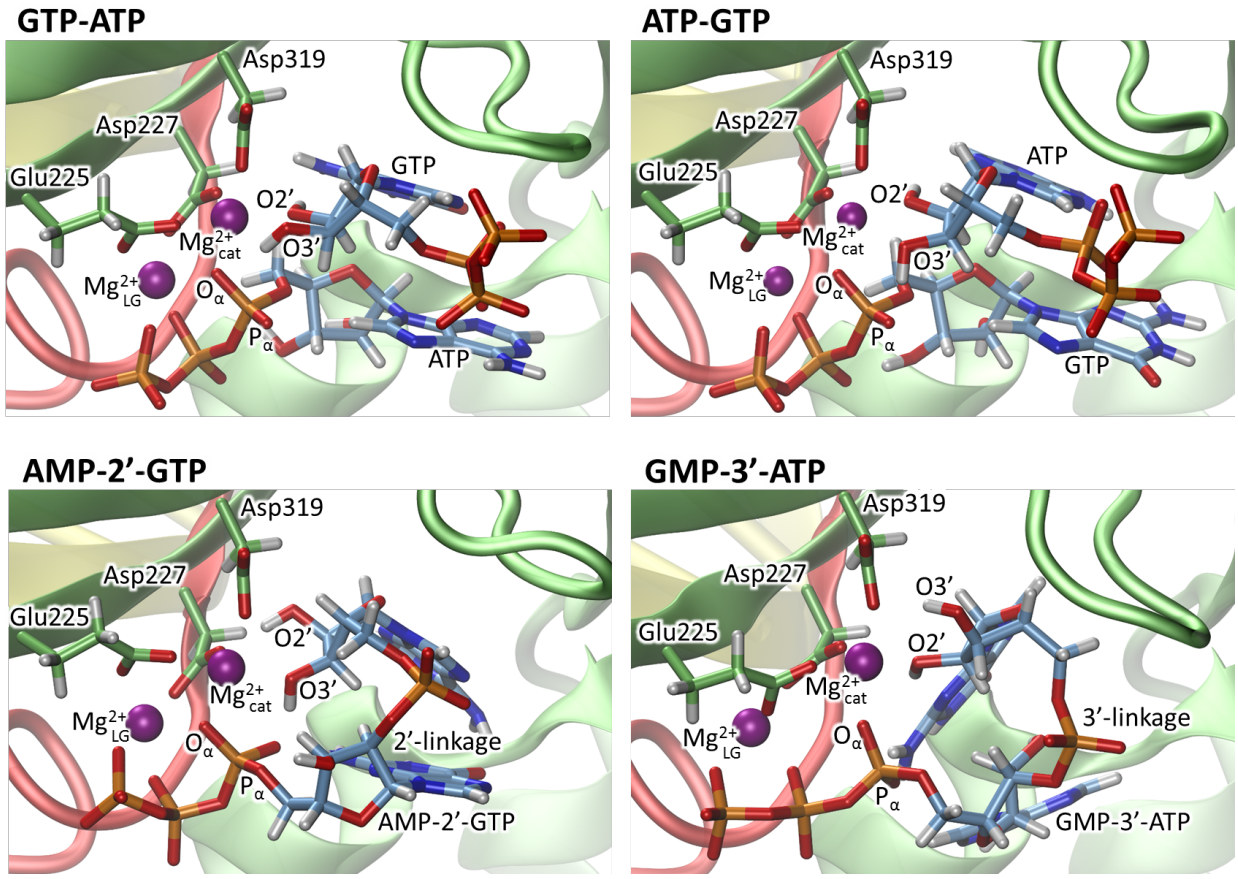


Figure S2. Relative orientation of the reactants and intermediates of both routes in cGAS active site.

cGAS, its activation loop, and DNA are shown in green, red, and yellow cartoons while the substrates are shown in blue sticks and the Mg cations as purple spheres. These structures were used as starting points for the MD simulations carried out along this study.

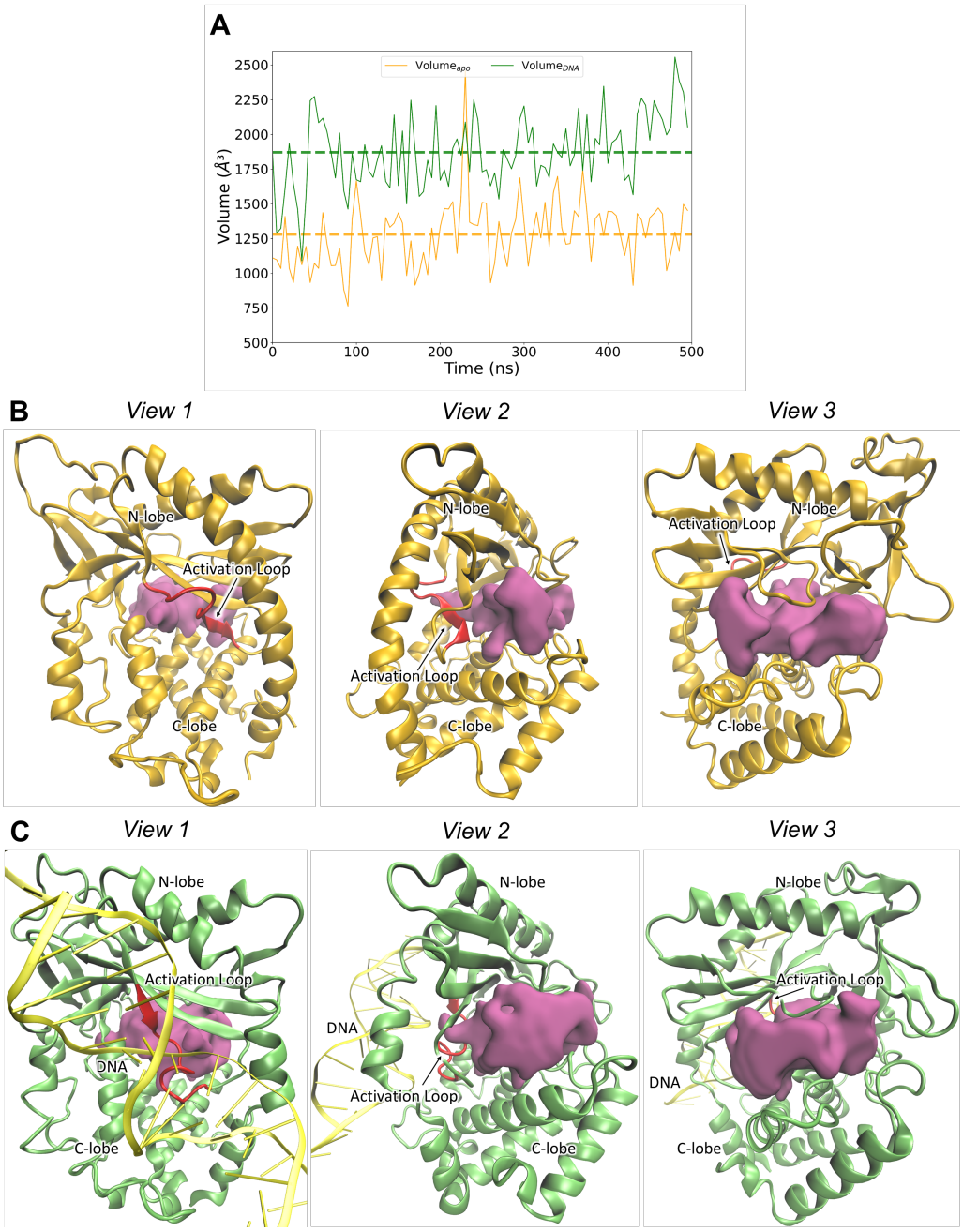


Figure S3. Active site volume analysis along time (**A**) and representative active site volumes for the apoenzyme and the DNA-bound apoenzyme (**B** and **C**).

A) Volume variations along the simulations of the apoenzyme (orange line) and the DNA-bound apoenzyme (green line). The average value for each system is shown with a dashed line (1280 ± 235 Å^3 for the apoenzyme and 1871 ± 260 Å^3 for the cGAS:DNA complex). 100 snapshots uniformly distributed along the MDs were used for volume analysis. **B)** Three different views of two representative structures from the apoenzyme MD simulation (snapshot 85 at 425ns and volume of 1276 Å^3) and **C)** the DNA-bound apoenzyme MD simulation (snapshot 68 at 340ns and volume of 1870 Å^3). Activation loop is shown in red cartoon, cGAS enzyme in orange or green cartoon (for apoenzyme and DNA-bound apoenzyme, respectively) and DNA is shown in yellow cartoon. The cavity volumes are shown as purple surfaces.

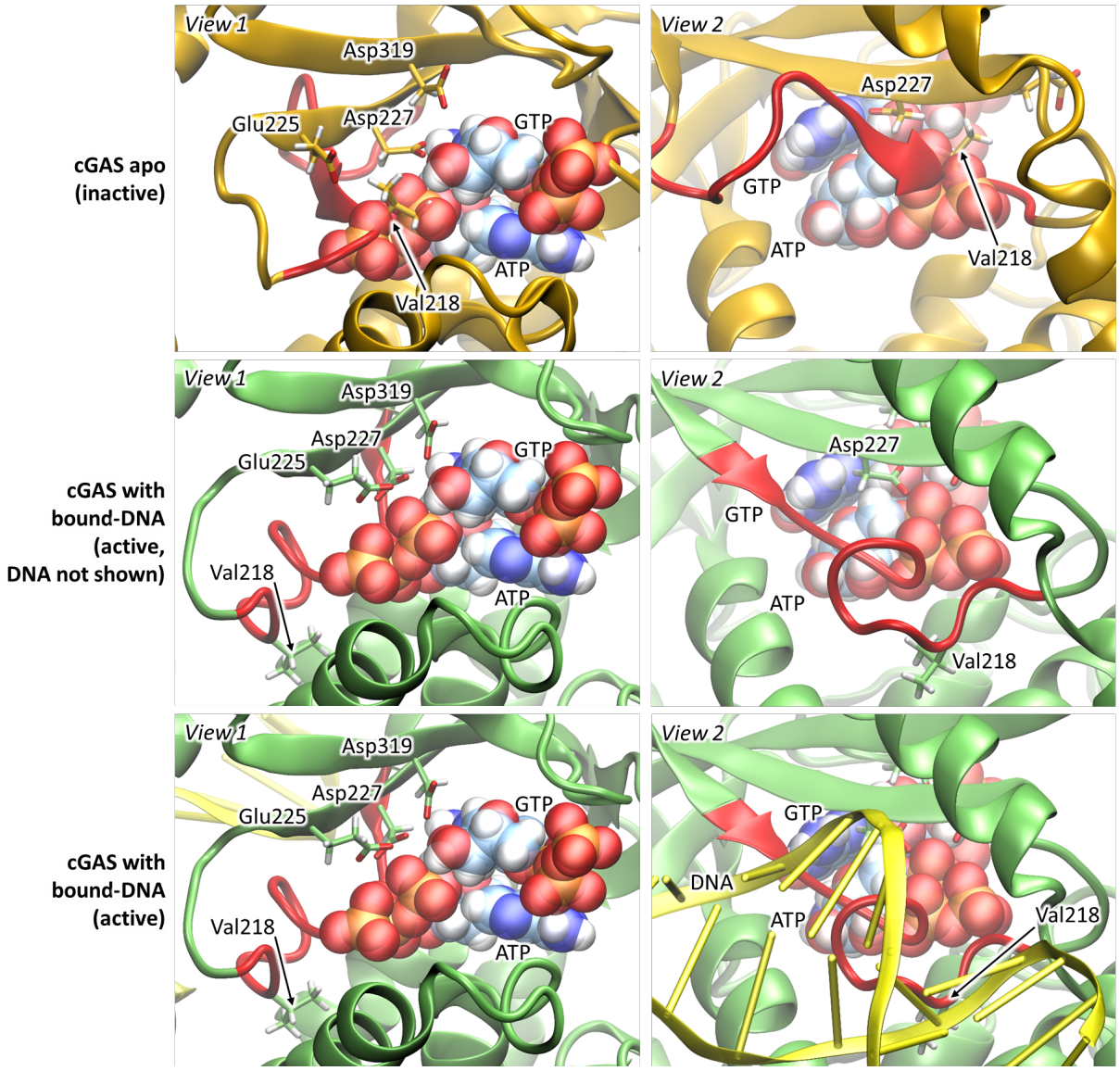


Figure S4. Importance of the activation loop conformation for catalysis.

A superposition between the apoenzyme cGAS and the substrates of the GTP-ATP catalytic route (shown as spheres) was performed to show how the inactive conformation of the activation loop (in red) hinders the catalytic site. As can be seen, in the orange inactive conformation, Val218 overlaps with the leaving group charge stabilizer magnesium cation (Mg^{2+}_{LG}) and the triphosphate chain of the ATP substrate. The latter is ATP in this case, but depending on the different routes, it can be GTP, AMP-2'-GTP or GMP-3'-ATP. In the green active conformation, Val218 is far from the PPi group and catalysis is permitted. As a consequence, the activation loop has to be in the active conformation to have a competent catalytic active site that can accommodate the Mg^{2+} ions, as well as the bottom substrate.

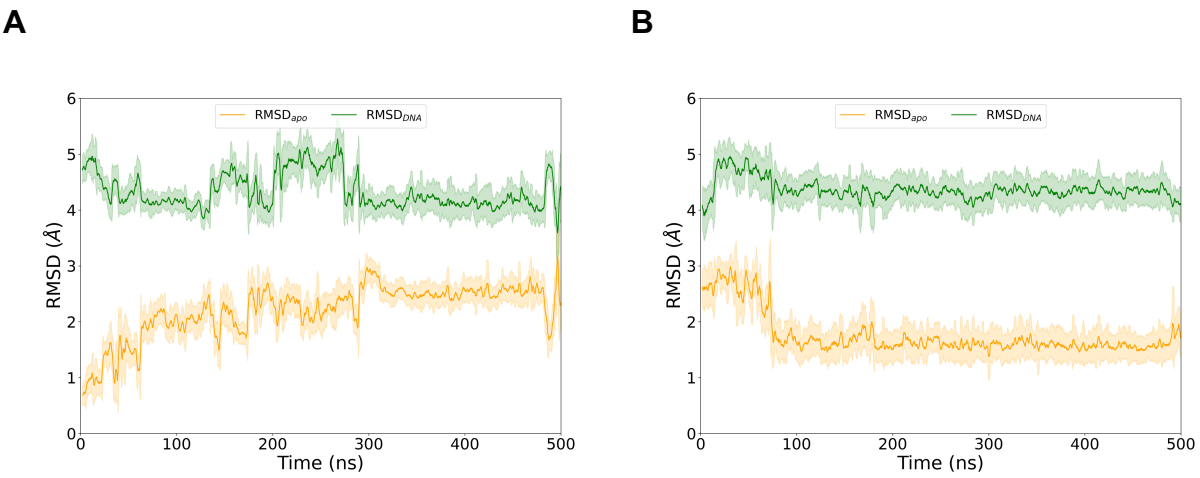


Figure S5. RMSD evolution of the apo cGAS along cMD (**A**) **and** aMD (**B**) simulations.

The orange curves represent the RMSD evolution of the backbone atoms of the activation loop (res. 210 to 220) taking as a reference the X-ray crystal inactive state of cGAS ($RMSD_{apo}$). The green curves plot the same RMSD but taking the X-ray DNA-bound cGAS apoenzyme as a reference ($RMSD_{DNA}$). **A)** corresponds to cMD and **B)** to aMD simulations.

The average $RMSD_{apo}$ and $RMSD_{DNA}$ for cMD are 2.20 ± 0.49 Å and 4.33 ± 0.35 Å, respectively. The average $RMSD_{apo}$ and $RMSD_{DNA}$ for aMD are 1.76 ± 0.41 Å and 4.37 ± 0.23 Å, respectively. Thus the activation loop resembles more the inactive conformation than the active one. As can be inferred from these values, the activation loop remains in the same conformation as during the conventional MD, and even after 500 ns of accelerated MD, no significant change is observed.

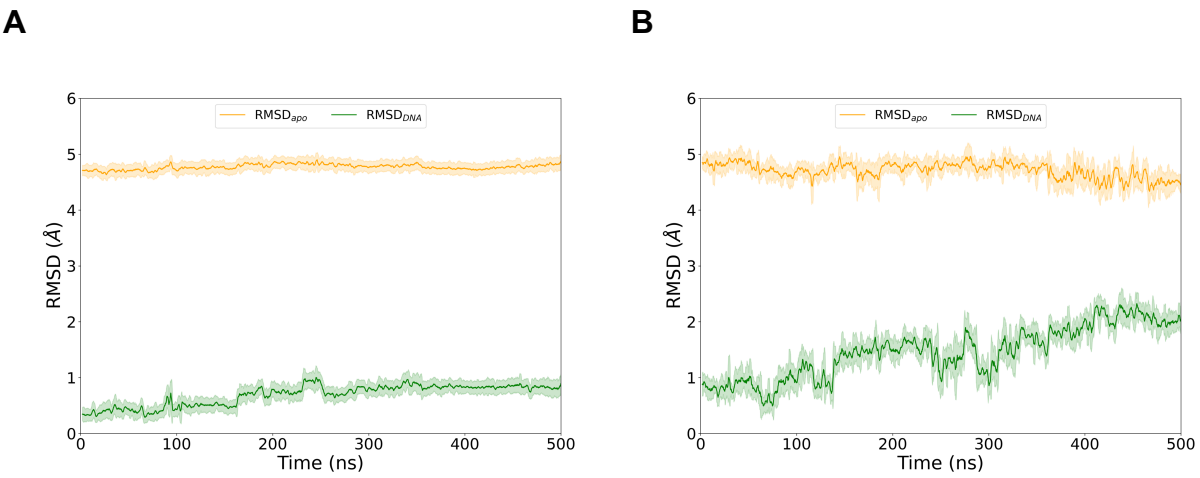


Figure S6. RMSD evolution of the DNA-bound apoenzyme along the cMD (**A**) and aMD simulations (**B**).

The orange curves represent the RMSD evolution of the backbone atoms of the activation loop (res. 210 to 220) taking as a reference the X-ray crystal inactive state of cGAS ($RMSD_{apo}$). The green curves represent the same RMSD
but taking the X-ray DNA-bound cGAS apoenzyme as a reference ($RMSD_{DNA}$). **A)** corresponds to cMD and **B)** to aMD simulations.

The average $RMSD_{apo}$ and $RMSD_{DNA}$ for cMD are 4.772 ± 0.080 Å and 0.69 ± 0.20 Å, respectively. The average $RMSD_{apo}$ and $RMSD_{DNA}$ for aMD are 4.71 ± 0.16 Å and = 1.46 ± 0.47 Å, respectively. RMSD values then suggest that the activation loop remains in an active conformation, and even after 500 ns of aMD, no significant change is observed.

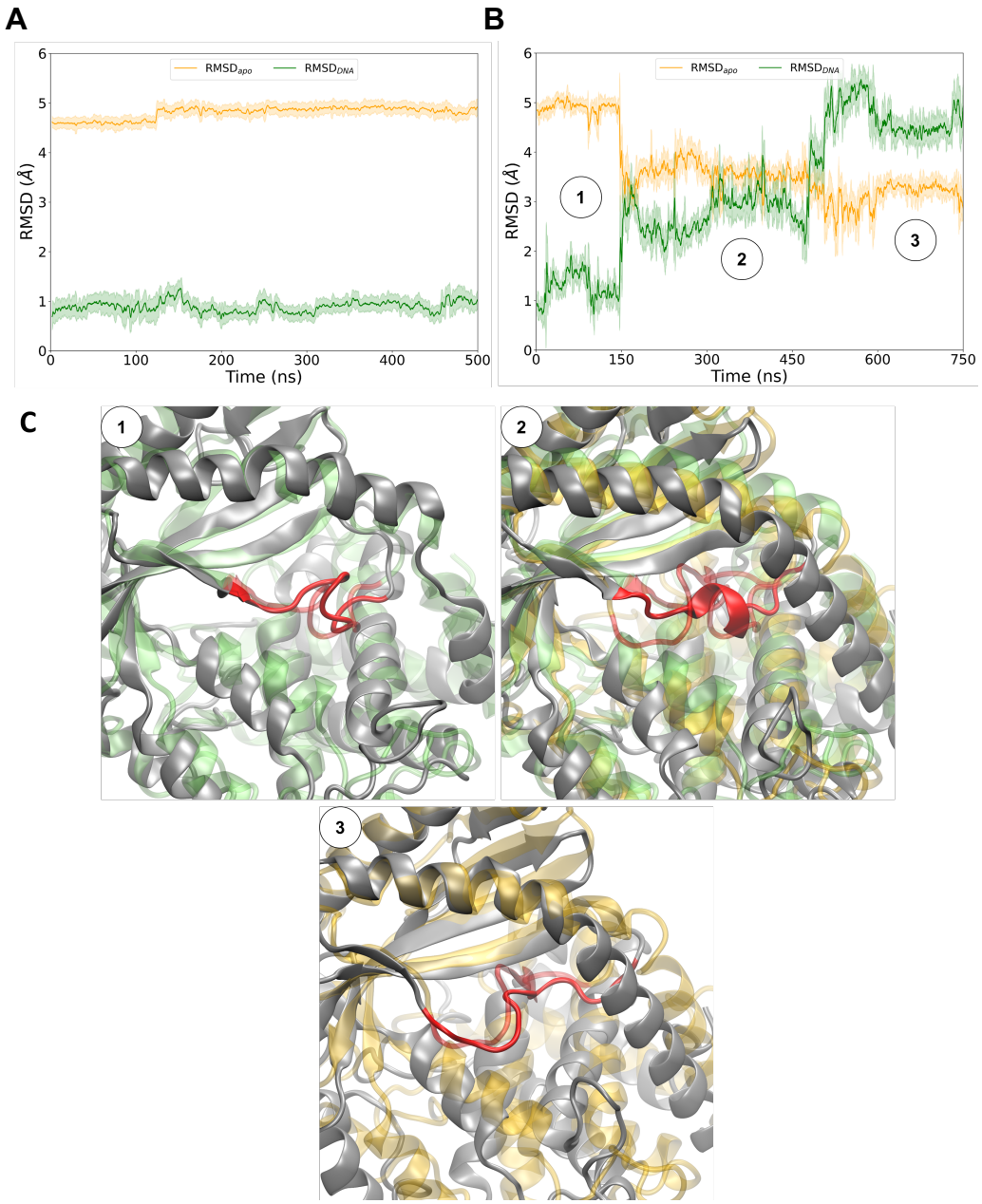


Figure S7. RMSD evolution of the apoenzyme starting from an active cGAS state and without DNA along cMD and aMD simulations.

The orange curves represent the RMSD evolution of the backbone atoms of the activation loop (res. 210 to 220) taking as a reference the X-ray crystal inactive state of cGAS ($RMSD_{apo}$). The green curves represent the same RMSD but taking the X-ray DNA-bound cGAS apoenzyme as a reference ($RMSD_{DNA}$). **A** corresponds to cMD and **B** to aMD simulation. **C** presents three different conformational states of the activation loop explored during the aMD simulation. cGAS along the aMD is show with silver cartoon while the activation loop is shown in red. The active and inactive states are also shown in green and orange transparent cartoon, while the activation loop is shown in a transparent red cartoon. For the sake of clarity, the inactive and active states are not shown in C.1 and C.3, respectively.

As can be seen, the conformational change of the activation loop can occur in 750ns of aMD. This loop explores a dynamic intermediate conformation shown in C.2 that connects the active and inactive conformations.

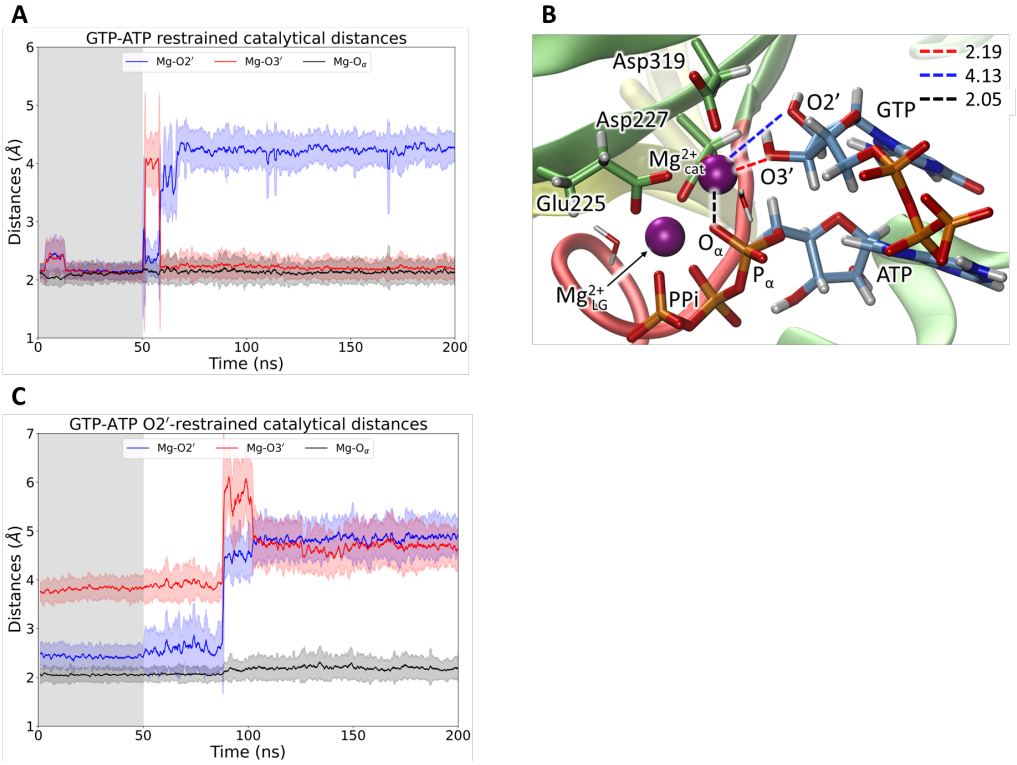


Figure S8. Restrained MD with the substrates of the GTP-ATP route.

A) Conventional MDs were run applying a harmonic force constant of 10 kcal mol^{-1} $Å^{-2}$ to the substrates during the equilibration and the first 50 ns of production (see shadowed region). 150ns of unrestrained MD simulation were then carried out. **B)** Last snapshot at the end of the GTP-ATP restrained MD from the previous cMD. cGAS, its activation loop, and DNA are shown in green, red, and yellow cartoons while the substrates are shown in blue sticks and the Mg cations as purple spheres. The legend includes the distances in Angstrom (Å). **C)** cMD simulation applying a harmonic force constant of 10 kcal mol^{-1} $Å^{-2}$ to the Mg-O2' bond during the equilibration and the first 50 ns of production (see shadowed region), followed by 150ns of unrestrained MD. Figures **A** and **C** show the Mg-O2' distances in blue and the Mg-O3' distances, which will not produce 2',3'-cGAMP, in red.

During the minimization prior to the first restrained MD, the GTP substrate coordinates the Mg^{2+} ion with the two hydroxyl groups, instead of the water molecule. This conformation is stable due to the restraint. Indeed, when the restraint is released, a water molecule interacts with the catalytic Mg^{2+} and only one GTP hydroxyl group is involved (**A**). During 8 ns after removing the restraints, the 2'-hydroxyl coordinates the Mg^{2+} ion but is quickly substituted by the 3'-hydroxyl and then becomes stable during the simulation. In this conformation (**B**), the 2'-hydroxyl group forms a H-bond with Thr321, while in the unrestrained MD, it forms an intramolecular H-bond. This may explain why the 3' conformation is stable along the restrained MD.

Finally, to check that the results are meaningful, we performed another MD, but applying a single restraint between Mg^{2+} and the O2' atom. Here, we expect that the enzyme and the substrate recognize each other and stabilize the conformation that will lead to the correct intermediate. As can be seen in **C**, the first 50 ns of restrained MD present acceptable structures. However, just after removing the restraint, the catalytic distance between O2' and Mg^{2+} (in blue) starts to shake until the 2'-hydroxyl is expelled and a water molecule enters and occupies its position. At the end, the conformation that is sampled during this MD is the same as the stable conformation shown in Figure 3c. Therefore, these results suggest that the GTP-ATP catalytic route does not provide a reasonable reactive conformation to produce the intermediate required to generate the 2',3'-cGAMP product.

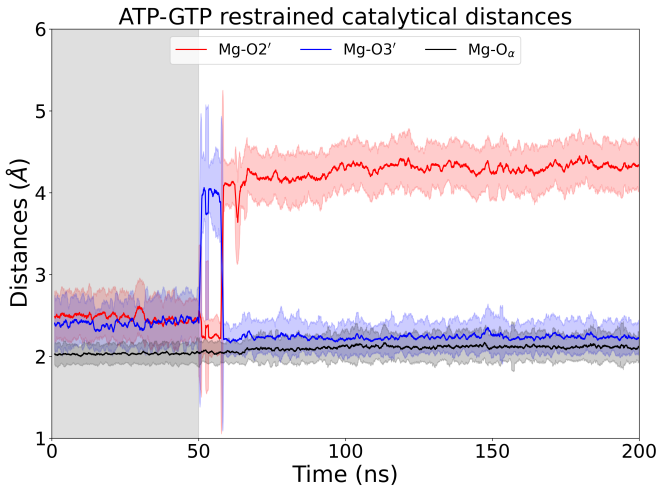
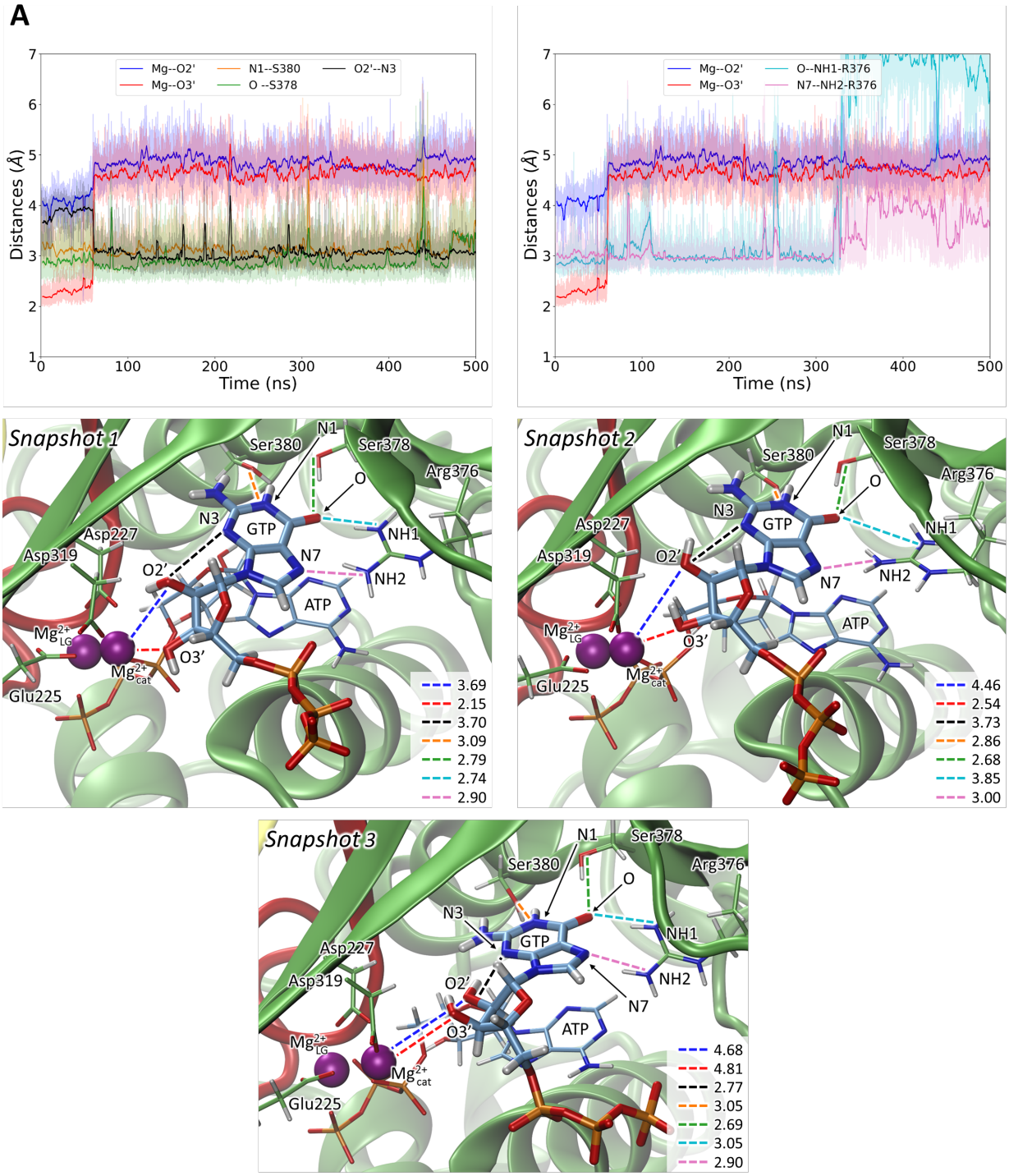


Figure S9. Restrained MD with the substrates of the ATP-GTP route.

Conventional MD simulations were run, applying a harmonic force constant of 10 kcal mol^{-1} $Å^{-2}$ to the substrates during the equilibration and the first 50 ns of production (see shadowed region). 150ns of unrestrained MD simulation were then carried out. This figure shows the evolution of the Mg-O3' distances in blue and the Mg-O2' distances, which will not produce 2',3'-cGAMP, in red.

As can be seen, after releasing the restraint, the 2'-hydroxyl group interacts with the Mg^{2+} ion but is quickly substituted by the 3'-hydroxyl group. This stable conformation resembles the 3' conformation explored during the ATP-GTP unrestrained MD (see Figure 3d).


A
Mg--O2'
Mg--O3'
N1--S380
O --S378
O2'--N3
Distances (Å)
Time (ns)
Mg--O2'
Mg--O3'
O--NH1-R376
N7--NH2-R376
Distances (Å)
Time (ns)
Snapshot 1
Ser380
N1
Ser378
Arg376
O
N3
GTP
NH1
Asp227
Asp319
O2'
N7
NH2
ATP
Mg2+LG
O3'
Mg2+cat
Glu225
3.69
2.15
3.70
3.09
2.79
2.74
2.90
Snapshot 2
Ser380
N1
Ser378
Arg376
O
N3
GTP
NH1
Asp227
Asp319
O2'
N7
NH2
ATP
Mg2+LG
O3'
Mg2+cat
Glu225
4.46
2.54
3.73
2.86
2.68
3.85
3.00
Snapshot 3
N1
Ser378
Ser380
Arg376
O
N3
GTP
NH1
Asp227
Asp319
O2'
NH2
N7
ATP
O3'
Mg2+LG
Mg2+cat
Glu225
4.68
4.81
2.77
3.05
2.69
3.05
2.90

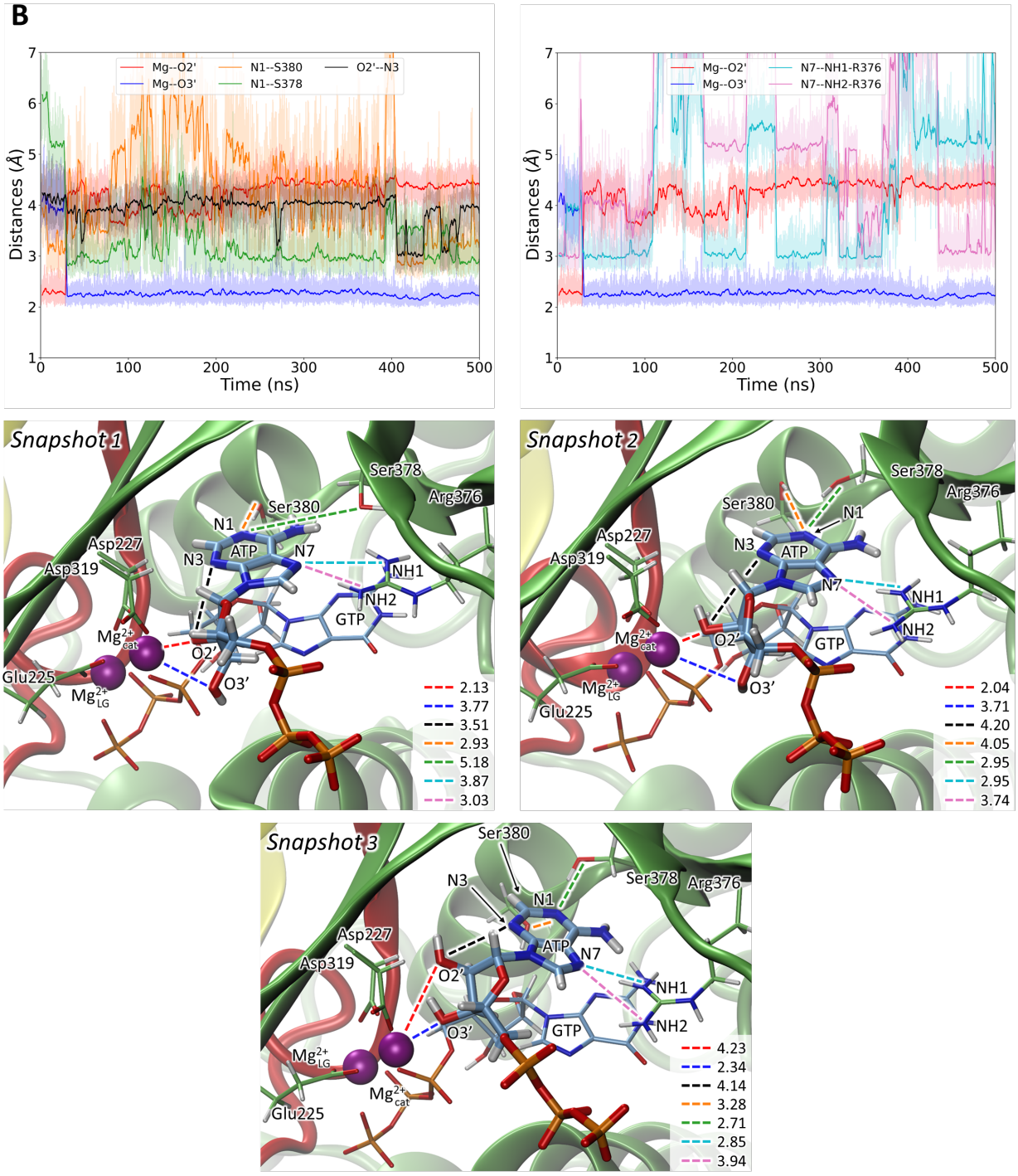


Figure S10. Analysis of the base-specific interactions of the nucleotide bases with active site residues during the first steps of the two proposed catalytic routes.

A) cGAS base-specific interactions between the guanine base of GTP in the GTP-ATP route are monitored during the unrestrained cMD simulation shown in Figure 3a. The first figure shows the hydrogen bonds that are formed between guanine-N1 atom and the hydroxyl group of Ser380 (in orange), the guanine-O atom and the hydroxyl group of Ser378 (in green) and the intramolecular H-bond formed by GTP 2'-hydroxyl group and guanine-N3 atom (in black). The second figure shows the H-bonds formed between guanine-O and

Arg376 (in cyan) and guanine-N7 and Arg376 (in pink). Both figures also show the time evolution of the Mg-O3' distance (in red) and the Mg-O2' distance (in blue). To facilitate comprehension of the interactions along the MD simulation, three snapshots along the MD simulation are provided with the analyzed distances showed as dashed lines. These three snapshots correspond to: 1) a representative conformation before the interaction between Mg^{2+}_{cat} and O3' vanishes, 2) the snapshot just before the Mg-O3' distance becomes larger, and 3) a representative snapshot after the interaction between Mg^{2+}_{cat} and O3' vanishes. **B**) cGAS base-specific interactions between the adenine base of ATP in the ATP-GTP route are monitored during the unrestrained cMD simulation shown in Figure 3b. The first figure shows the hydrogen bonds that are formed between guanine-N1 atom and the hydroxyl group of Ser380 (in orange), the guanine-N1 atom and the hydroxyl group of Ser378 (in green) and the intramolecular H-bond formed by ATP 2'-hydroxyl group and adenine-N3 atom (in black). The second figure shows the H-bonds formed between adenine-N7 and Arg376 (in cyan) and adenosine-N7 and Arg376 (in pink). Both figures also show the time evolution of the Mg-O2' distance (in red) and the Mg-O3' distance (in blue). To facilitate comprehension of the interactions along the MD simulation, three snapshots along the MD simulation are provided with the analyzed distances shown as dashed lines. These three snapshots correspond to: 1) a representative conformation before the 2' and 3'-hydroxyl groups of ATP exchange their positions in the coordination sphere of the Mg^{2+}_{cat} ion, 2) the snapshot just before this exchange, and 3) a representative snapshot after the exchanging. cGAS and relevant active site residues are shown in green cartoon and sticks, respectively. The activation loop and DNA are shown in red and yellow cartoon. The legends in the three snapshots include the distances in Angstrom (Å).

At the beginning of the GTP-ATP MD simulation (**A**, snapshot 1), GTP is coordinated to the Mg^{2+} ion through the 3'-hydroxyl group, being the distance close to 2.2 Å. When the H-bond between Arg376 and the guanine-O atom vanishes in snapshot 2 (colored in cyan), it permits GTP to undergo a conformation change and the Mg-O3' interaction vanishes. This new conformation is stabilized by the intramolecular hydrogen bond formed between the 2'-hydroxyl group and the guanine-N3 atom (shown in black), that persists during the remaining MD simulation. Ser380 and Ser378 are H-bonded during the entire MD with guanine-N1 and guanine-O (shown in orange and green), respectively.

In the ATP-GTP MD simulation (**B**), Ser378 is not performing any interaction since the distance is bigger than 5 Å with the ATP substrate at the beginning and Ser380 performs an H-bond with guanine-N1 atom. However, just before the 3'-hydroxyl substitutes the 2'-hydroxyl group in the Mg^{2+} coordination sphere in Snapshot 2, this serine interacts with the guanine base through the N1 atom thus substituting Ser380. When the H-bond with Ser378 is formed, ATP orientates itself in a near-attack conformation by coordinating the 3'-hydroxyl group with the Mg^{2+} ion with a distance of ca. 2.2 Å. This reactive conformation is only obtained when Ser378 is H-bonded to the guanine-N1 atom.

As can be seen, Ser378 performs a base-specific interaction with the adenine base of ATP. The same interaction is not possible to be formed in the GTP-ATP route because the guanine base has the N1 atom saturated with a hydrogen and it is interacting with Ser380, while Ser378 is H-bonded with the guanine-O.