Supporting Information: Molecular mechanism of rRNA maturation by M5: interplay between conformational flexibility and reactivity

Julie Puyo-Fourtine,[*,†] Laetitia Kantin,[†] Milorad Andelkovic,[‡] Carine Tisné,[¶] Elisa Frezza,[§] Iñaki Tuñón,[‡] and Elise Duboué-Dijon[*,†]

†*Université Paris Cité, CNRS, Laboratoire de Biochimie Théorique, 13 rue Pierre et Marie Curie, 75005, Paris, France*

‡*Departament de Química Física, Universitat de València, 46100 Burjassot, Spain*

¶*Université Paris Cité, CNRS, Institut de Biologie Physico-Chimique, IBPC, Expression Génétique Microbienne, Paris 75005, France*

§*Université Paris Cité, CiTCoM, Paris, France*

E-mail: puyo@ibpc.fr; elise.duboue-dijon@cnrs.fr

Additional computational details

Force fields

Table S1 provides the details of the force field combinations tested to study the structural stability of the active site, and Table S2 recalls the ionic parameters associated with the employed full and scaled charge Mg^{2+} force fields.

Table S1: Detailed force field combinations tested to characterize the active site structural stability.

Simulation	Protein	RNA	Ions	Water
Standard	ff14sb	amberff99_bsc0-χ_{OL3}	Mg^{2+} [S1]	TIP3P
Pair-specific Lennard-Jones	ff14sb	amberff99_bsc0-χ_{OL3}	mMg [S2]	TIP3P
ECC0.8	ff14sb	amberff99_bsc0-χ_{OL3}	Mg ECC0.8 [S3]	SPCE

Table S2: Parameters of the employed full charge [S1] and ECC 0.8 force fields for Mg^{2+}

Type	Charge	Mass	At.num	σ (nm)	ϵ (kJ/mol)
Standard	2.0	24.305	12	0.242324	0.04268672
ECC 0.8	1.6	24.305	12	0.13600	3.661000

Tables S6, S5, S4 and S3 provide the list of charges used for the Glu, Asp, RNA A and C residues in our ECC0.8 force field. The corresponding parameter files in Gromacs format are provided as supplementary files deposites on a public Zenodo folder. [S4] The choice was made to limit charge scaling to a few atoms, similarly to what was done previously in proteins or phospholipids. [S5,S6]

Table S3: Charges of the glutamate residue within our ECC0.8 force field.

GLU	OL15 14sb	ECC 0.8
N	-0.5163	-0.5163
H	0.2936	0.2936
CA	0.0397	0.0397
HA	0.1105	0.1105
CB	0.056	0.056
HB1	-0.0173	-0.0173
HB2	-0.0173	-0.0173
CG	0.0136	0.0136
HG1	-0.0425	-0.034
HG2	-0.0425	-0.034
CD	0.8054	0.66088
OE1	-0.8188	-0.65504
OE2	-0.8188	-0.65504
C	0.5366	0.5366
O	-0.5819	-0.5819
Residue charge	**-1**	**-0.8**

Table S4: Charges of the aspartate residue (Asp 56) with ECC 0.8 for the active site stability study, with the rescaling applied on the magnesium ions and the residues on their coordination shell

ASP	OL15 14sb	ECC 0.8
N	-0.5163	-0.5163
H	0.2936	0.2936
CT	0.0381	0.0381
H1	0.088	0.088
CT	-0.0303	-0.0303
HC	-0.0122	-0.0122
HC	-0.0122	-0.0122
C	0.7994	0.67884
O2	-0.8014	-0.64112
O2	-0.8014	-0.64112
C	0.5366	0.5366
O	-0.5819	-0.5819
Residue charge	**-1**	**-0.8**

Table S5: Charges of the cytosine residue (C 116) with ECC 0.8 for the active site stability study, with the rescaling applied on the magnesium ions and the residues on their coordination shell

RC		**OL15**	**ECC 0.8**
P	P	1.1662	0.93296
O1P	O2	-0.776	-0.6208
O2P	O2	-0.776	-0.6208
O5'	OS	-0.4989	-0.44005
C5'	CT	0.0558	0.0558
H5'1	H1	0.0679	0.0679
H5'2	H1	0.0679	0.0679
C4'	CT	0.1065	0.1065
H4'	H1	0.1174	0.1174
O4'	OS	-0.3548	-0.3548
C1'	CT	0.0066	0.0066
H1'	H1	0.2029	0.2029
N1	N*	-0.0484	-0.0484
C6	CM	0.0053	0.0053
H6	H	0.1958	0.1958
C5	CM	-0.5215	-0.5215
H5	H1	0.1928	0.1928
C4	CA	0.8185	0.8185
N4	N2	-0.953	-0.953
H41	H	0.4234	0.4234
H42	H	0.4234	0.4234
N3	NC	-0.7584	-0.7584
C2	C	0.7538	0.7538
O2	O	-0.6208	-0.6208
C3'	CT	0.0679	0.0679
H3'	H1	0.0615	0.0615
C2'	CT	0.0972	0.0972
H2'1	H1	0.0972	0.0972
O2'	OH	-0.6139	-0.6139
HO2'	HO	0.4186	0.4186
O3'	OS	-0.5246	-0.46061
Charge residue		**-1**	**-0.8**

Table S6: Charges of the adenine residue (A 117) with ECC 0.8 for the active site stability study, with the rescaling applied on the magnesium ions and the residues on their coordination shell

RA		**OL15**	**ECC 0.8**
P	P	1.1662	0.93296
O1P	O2	-0.776	-0.6208
O2P	O2	-0.776	-0.6208
O5'	OS	-0.4989	-0.44005
C5'	CT	0.0558	0.0558
H5'1	H1	0.0679	0.0679
H5'2	H1	0.0679	0.0679
C4'	CT	0.1065	0.1065
H4'	H1	0.1174	0.1174
O4'	OS	-0.3548	-0.3548
C1'	CT	0.0394	0.0394
H1'	H2	0.2007	0.2007
N9	N*	-0.0251	-0.0251
C8	CK	0.2006	0.2006
H8	H5	0.1553	0.1553
N7	NB	-0.6073	-0.6073
C5	C5	0.0515	0.0515
C6	CA	0.7009	0.7009
N6	N2	-0.9019	-0.9019
H61	H	0.4115	0.4115
H62	H	0.4115	0.4115
N1	NC	-0.7615	-0.7615
C2	CQ	0.5875	0.5875
H2	H5	0.2022	0.2022
N3	NC	-0.6997	-0.6997
C4	CB	0.3053	0.3053
C3'	CT	0.0679	0.0679
H3'	H1	0.0615	0.0615
C2'	CT	0.067	0.067
H2'1	H1	0.0972	0.0972
O2'	OH	-0.6139	-0.6139
HO2'	HO	0.4186	0.4186
O3'	OS	-0.5246	-0.46061
Charge residue		**-1**	**-0.8**

Equilibration protocol

Table S7 provides the details of the restraints applied at each step of the equilibration procedure. The idea was to gradually release positional restreints applied on different parts of the complex, releasing first the parts further away from the active site.

Table S7: Force constant (in kJ mol^{-1} nm^{-2}) of the positional restraints applied on either all heavy atoms or only backbone atoms of each component of the system during the different stages of the equilibration procedure.

	NTD (Toprim)	CTD	RNA	UL18	Mg^{2+}	duration
EQ1 (heavy atoms)	10000	10000	10000	10000	10000	5 ns
EQ2 (heavy atoms)	8000	8000	8000	8000	8000	5 ns
EQ3 (heavy atoms)	5000	5000	5000	5000	5000	5 ns
EQ4 (heavy atoms)	3000	3000	3000	3000	3000	5 ns
EQ5 (heavy atoms)	1000	1000	1000	1000	1000	5 ns
EQ6 (heavy atoms)	1000	700	1000	700	1000	5 ns
EQ7 (heavy atoms)	1000	500	900	500	1000	5 ns
EQ8 (backbone)	1000	300	800	300	1000	5 ns
EQ9 (backbone)	800	150	500	150	1000	5 ns
EQ10 (backbone)	500	70	500	70	1000	5 ns
EQ11 (backbone)	300	30	300	30	1000	5 ns
EQ12 (backbone)	150	0	150	0	1000	5 ns
EQ13 (backbone)	70	0	70	0	1000	5 ns
EQ14 (backbone)	30	0	30	0	1000	5 ns
EQ15 (backbone)	0	0	0	0	1000	5 ns
EQ16 (backbone)	0	0	0	0	1000	20 ns

QM/MM simulations

The QM region comprises the ribose phosphate backbone of A117 (cleavage between C1'-N9) and C116 (cleavage between C1' and N1), the two active site Mg^{2+} ions, the water molecules in their first hydration shell as well as protein residue side chains involved in Mg2+ inner-sphere coordination (Glu96, Glu10, Asp56), for a total of 94 atoms, as represented in Figure S1. G115 was cleaved between C3'-C4' and C3'-C2' , A118 has been cleaved between C4'-C5'. On the protein side, Glu96 and Glu10 were cleaved between C_β-C_γ, and Asp 56 between C_α-C_β.

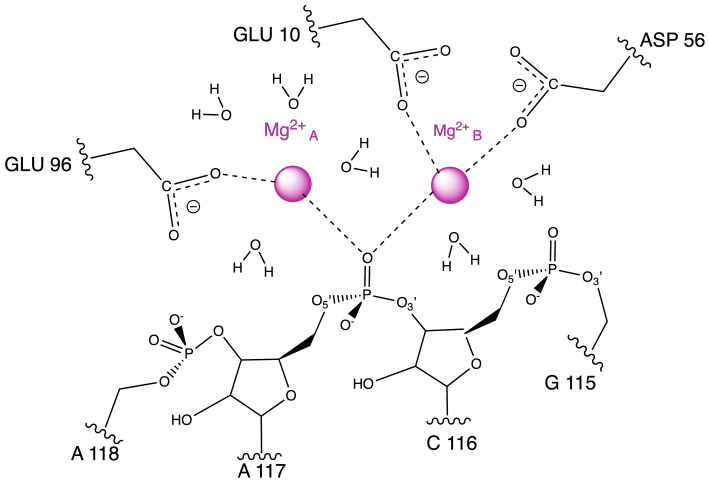

Figure S1: Representation of the QM zone.

Tables S8, S9 and S10 provide the list and definition of collective variables (CVs) used to explore the reaction free energy surface with the ASM methodology, for each of the mechanisms detailed in the main text (see Figure 6).

Table S8: Set of CVs used in the Adaptative String Method for the mechanism with water deprotonation by Glu96.

CV1	HW (from 1st water mol) - OE1 (GLU96)
CV2	OW (from 1st water mol) - P (A117)
CV3	OW - HW (1st water mol)
CV4	P (A117) - O3' (C116)
CV5	Mg_A^{2+} - OW (from 1st water mol)
CV6	HW (from 2nd water mol)- O3' (C116)
CV7	HW - OW (2nd water mol)

Table S9: Set of CVs used in the Adaptative String Method for the mechanism with proton shuttling through the scissile phosphate.

CV1	OW - P (A117)
CV2	O2P (A117) - H
CV3	HW - OW
CV4	P (A117) - O3' (C116)
CV5	HW - O3' (C116)

Table S10: Set of CVs used in the Adaptative String Method for the mechanism with water deprotonation by A118 phosphate.

CV1	O2P (A118) - HW (from 1st water mol)
CV2	OW - P (A117)
CV3	OW - HW (1st water mol)
CV4	P (A117) - O3' (C116)
CV5	Mg_A^{2+} - OW (from 1st water mol)
CV6	OW - HW (2nd water mol)
CV7	HW (from 2nd water mol) - O3' (C116)

Additional analyses

Active site structure

Using as a starting point the structure of the complex obtained by direct superimposition of the cryo-EM and X-ray structure of the complex and M5 N-terminal domain respectively— with the scissile phosphate coordinated to a single Mg^{2+} ion (see Figure 2a)— the active site structure very quickly breaks apart, with the inter-magnesium distance increasing after only a few ns (Figure S2).

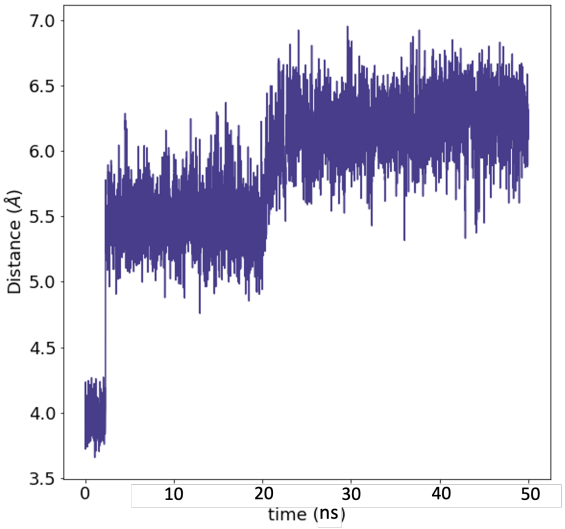

Figure S2: Evolution of the distance between the two magnesiums (modelled with standard force field) of the initial active site, without having moved the phosphate group between the two ions

Interface analysis

Figures S3 and S4 provide respectively the RMSD and RMSF calculated for each member of the protein:RNA complex, on each of the four 500 ns replicas. The residues that exhibit the largest fluctuations are without surprise the ends of the chains, or unstructured loops.

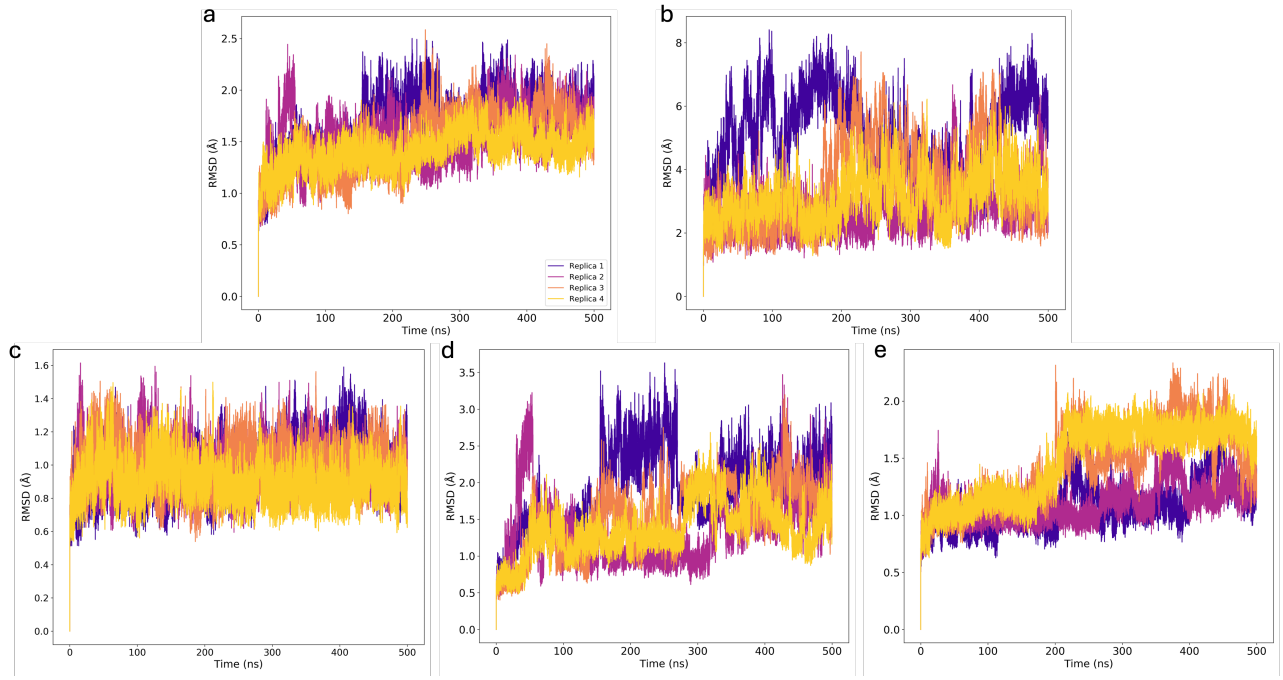

Figure S3: Root Mean Square Displacement (RMSD) for the four replicas of 500 ns for : a) M5 b) RNA c) M5 N-terminal domain d) M5 C-terminal domain and e) UL18

Figure S5 shows the evolution of the fraction of native contacts along our 4 500 ns replicas. Despite fluctuations, that are usual in such systems, and distinct behaviors in the different replicas, the fraction of native contacts remains above 70%, which is a sign of a stable interface.

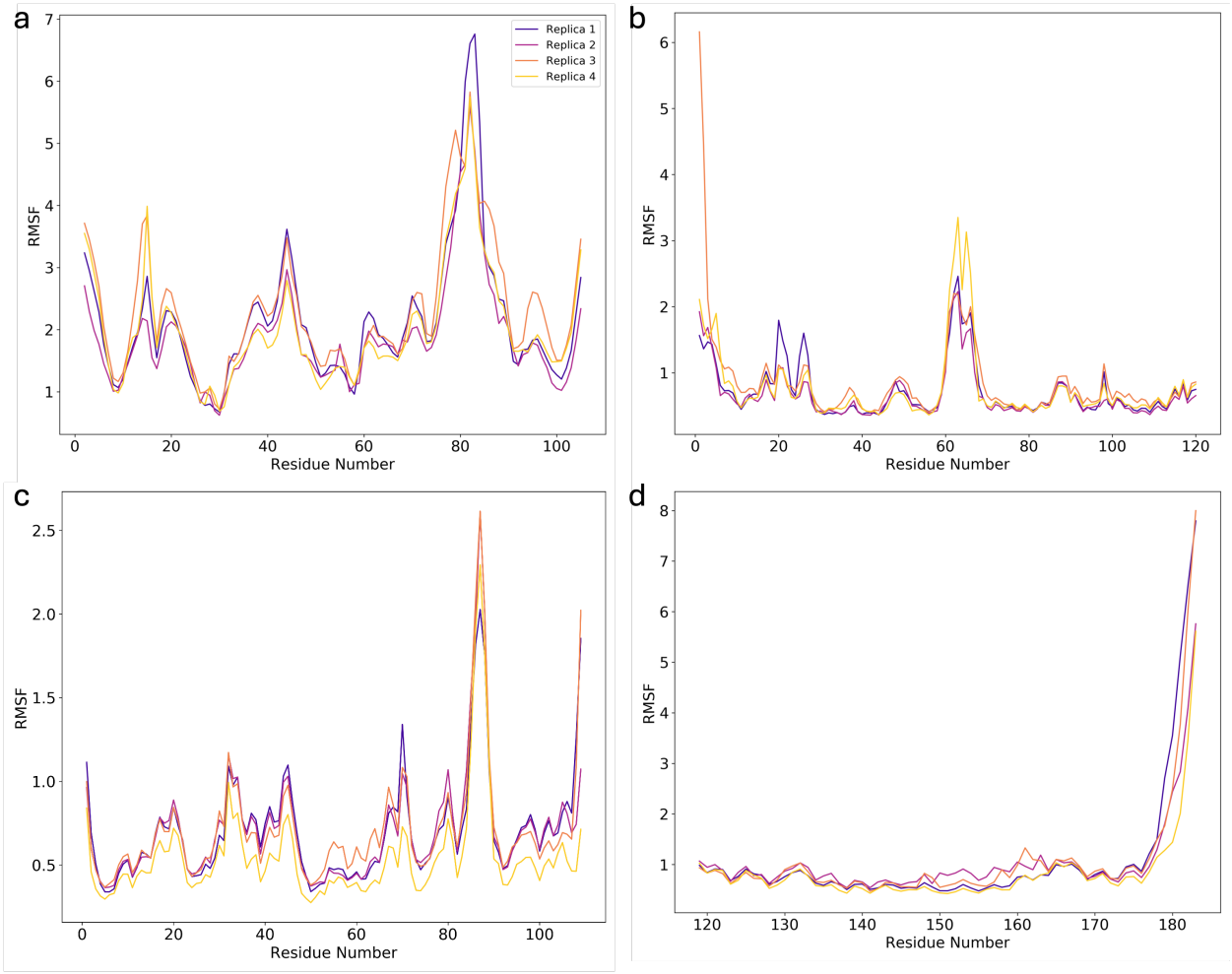

Figure S4: Root Mean Square Fluctuation (RMSF) for the four replicas of 500 ns for : a) M5 b) RNA c) M5 N-terminal domain d) M5 C-terminal domain and e) UL18

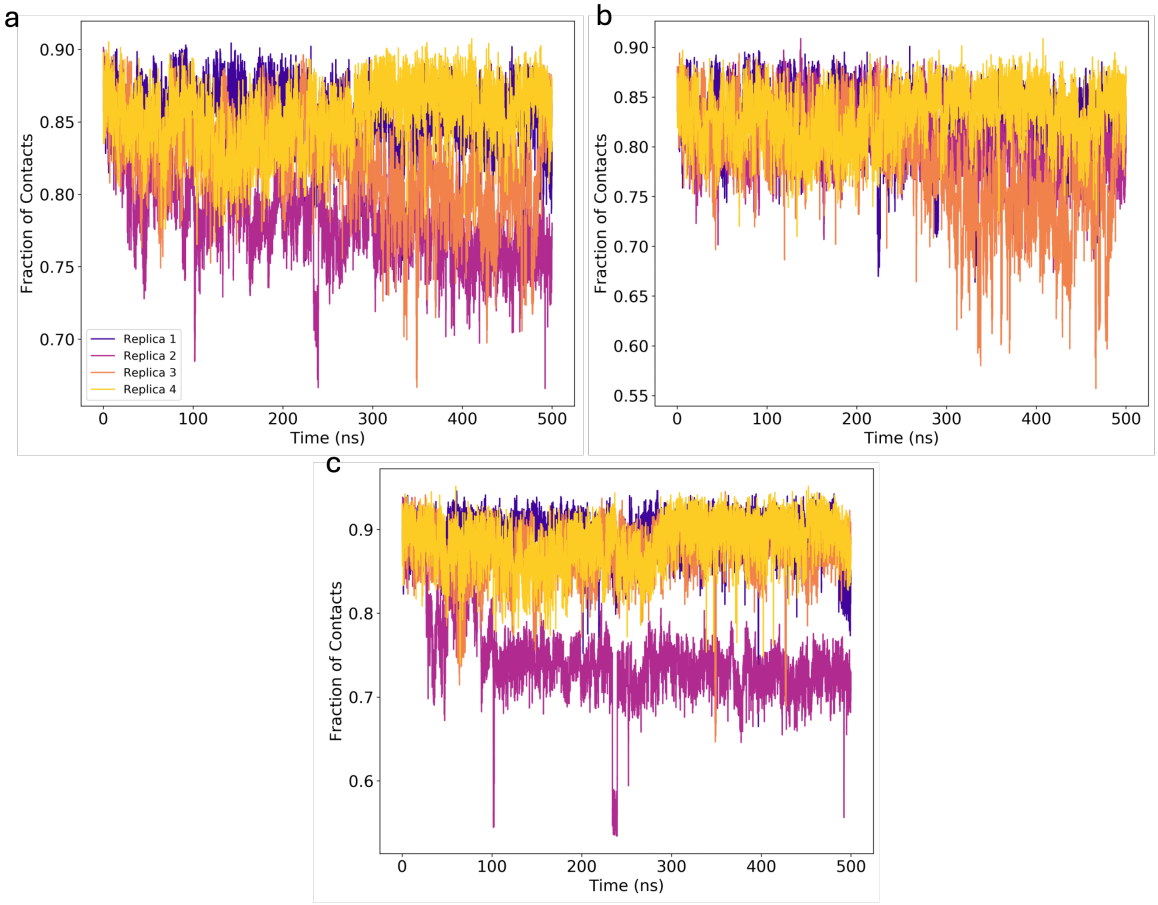

Figure S5: Native Contact Fraction (NCF) for a) M5 b) CTD c) NTD using the radius cut method

Clustering was performed as described in the Methods section based on the matric of interfacial contacts. Figures S6 and S7 show different analyses of this clustering, that justify the final choice of the number of clusters for both standard MD and REST2 simulations.

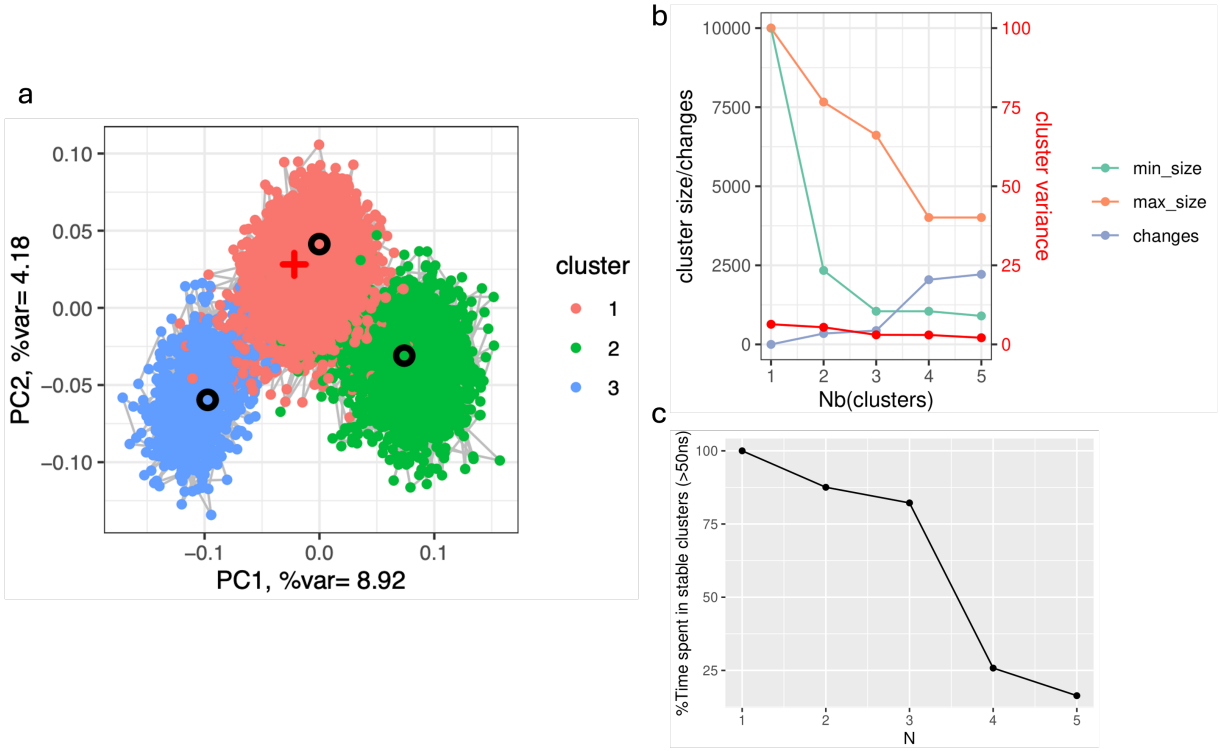

Figure S6: Clustering analysis for standard simulations (concatenation of 4 x 500 ns): a) Projection of principal component analysis, clusters are identified in red, green and blue. The circles represent the centroids, and the red cross, the first frame of the simulation. b) Changes as a function of the number of clusters considered; in green for a minimum size considered, in red for a maximum size and in blue the number of changes. c) tTime spent in the various clusters as a function of the number of clusters considered.

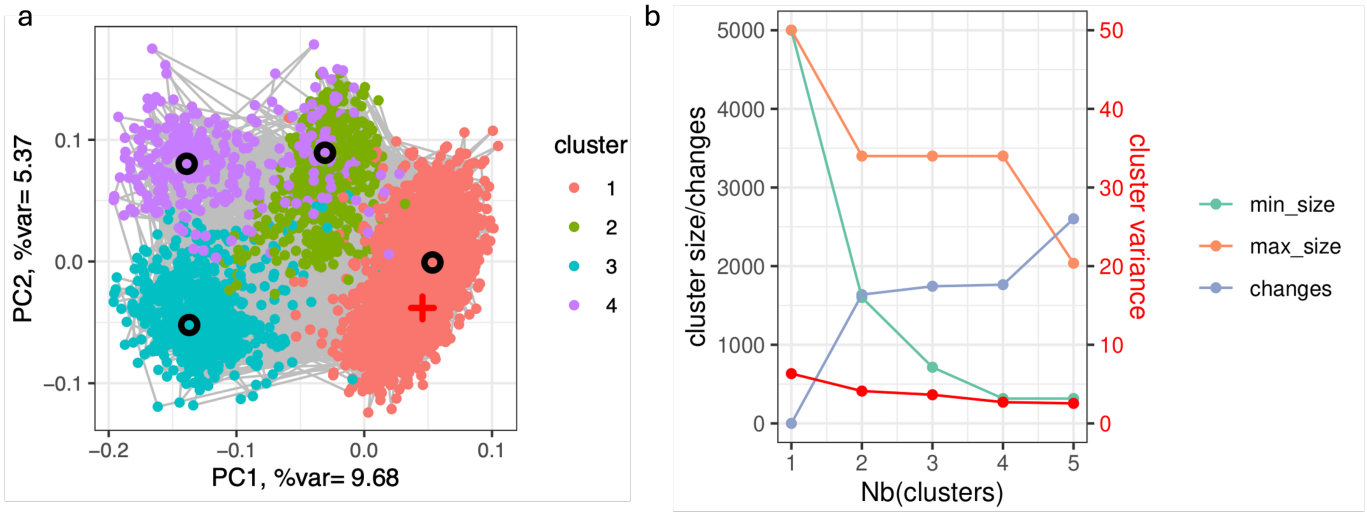

Figure S7: Clustering analysis for REST2 simulations (concatenation of 2 x 500 ns): a) Projection of principal component analysis, clusters are identified in red, green, blue and purple. The circles represent the centroids, and the red cross, the first frame of the simulation. b) Changes as a function of the number of clusters considered; in green for a minimum size considered, in red for a maximum size and in blue the number of changes.

Active site conformations

While our plain MD simulations show only a very small population with a water molecule directly bridging the phosphate A118 and Mg_A^{2+}, the population of this conformation with a short O2P-Mg^{2+} distance strongly increased in REST2 simulations, with our 2 replicas showing a qualitatively consistent behavior (Figure S8). To further assess the convergence of our REST2 simulations, we estimated (directly from the observed population distributions) the free energy profile along the Mg_A^{2+}-O2P(A118) distance on 10 simulation blocks of 50 ns (Figure S9). After a $\sim$ 200ns equilibration period, the free energy profile (and thus populations of the two conformations) appears reasonably converged.

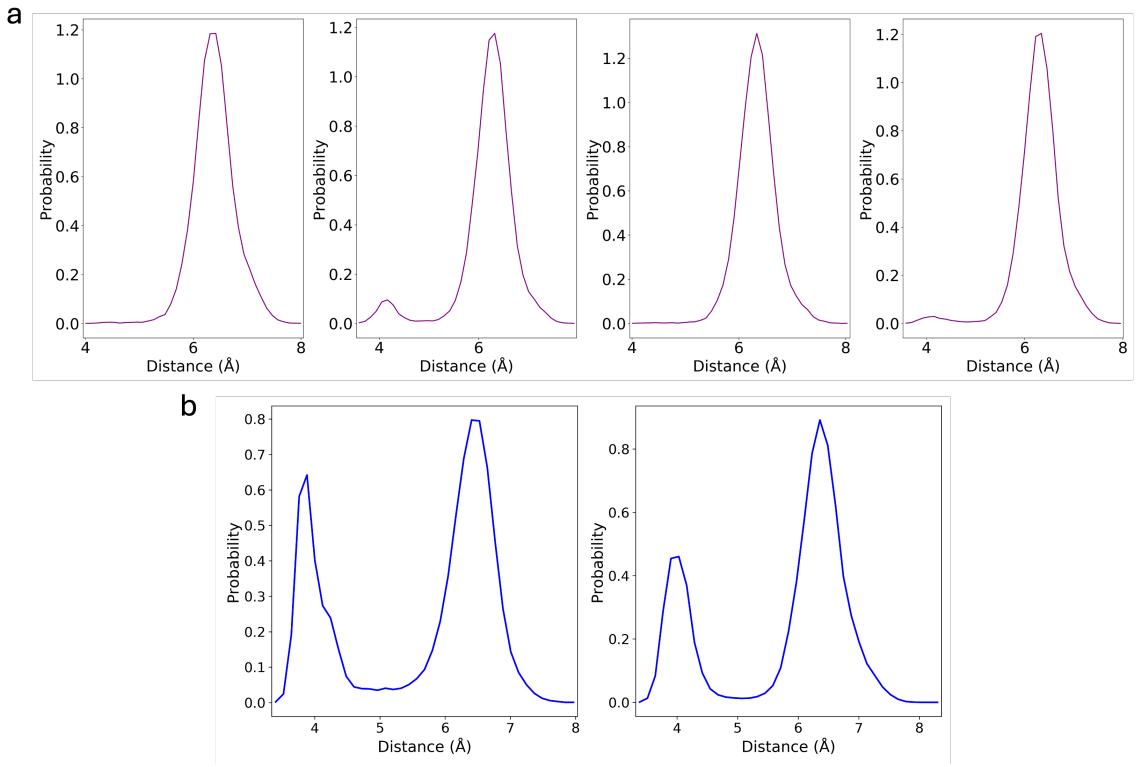

Figure S8: Histograms of the distance between oxygen O2P (from residue A118) and magnesium (Mg^{2+} A) a) in purple for the 4 standard 500 ns simulations and b) un blue for the 2 REST2 500 ns simulations

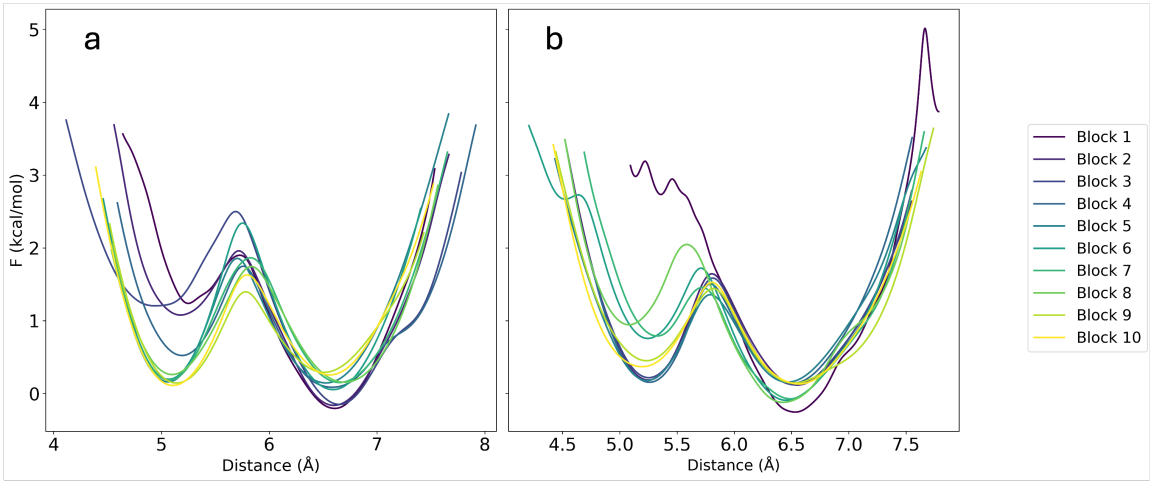

Figure S9: Evolution - of the Gibbs free energy of the distance between the non-bridging oxygen O2P of residue A118 and the phosphorus of the residue A117 involved on the mechanism - in blocks of 50 ns , for the two 500 ns simulations in REST2 a) is the first simulation of 500ns, b) the second simulation

To investigate the correlation between active site conformations and interface fluctuations, we decomposed the probability distribution of the distance between neighboring phosphate O2P(A118) and the scissile phosphate P(A117)—that separated well the two conformations—based on the clustering performed on the matrix of interfacial contacts (see Figure S10).

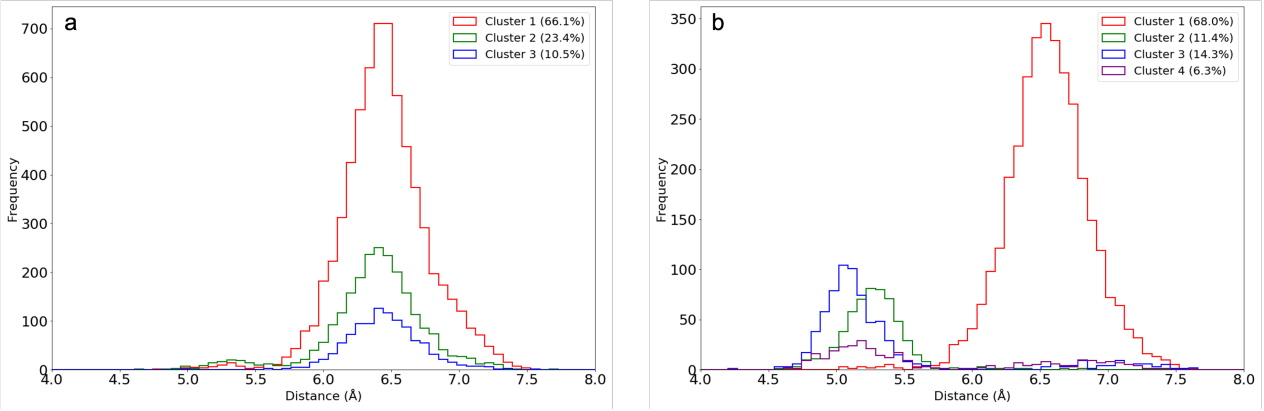

Figure S10: Decomposition of the overall distribution of O2P(A118)–P(A117) distance by clusters a) in the 4 standard 500 ns simulations and b) in the 2 REST2 500 ns simulations. The overall fraction of population represented by each cluster is indicated in parenthesis.

Mechanistic investigations

Several mechanisms were investigated from distinct conformations of the active (see main text, Fig. 5a). The initial deprotonation of the nucleophilic water molecule can be performed directly by Glu96 (most favored mechanism "Glu_1wat", described in detailed in the main text) or through another water molecule ("Glu_2wat", see Figure S11a). This deprotonation can also be done by A118 pro-Rp (O_R) phosphate directly (see Figure 6 in main text) or through another water molecule ("A118_2wat", Figure S11a). Similarly, in addition to the direct proton shuttling described in the main text and inspired by previous work on similar systems,[S7] we also considered the possibility of proton shuttling through a chain of water molecules ("shuttling_2wat", Figure S11b).

Figures S13, S12 and S14 gather the free energy profiles and evolution of the CVs com-

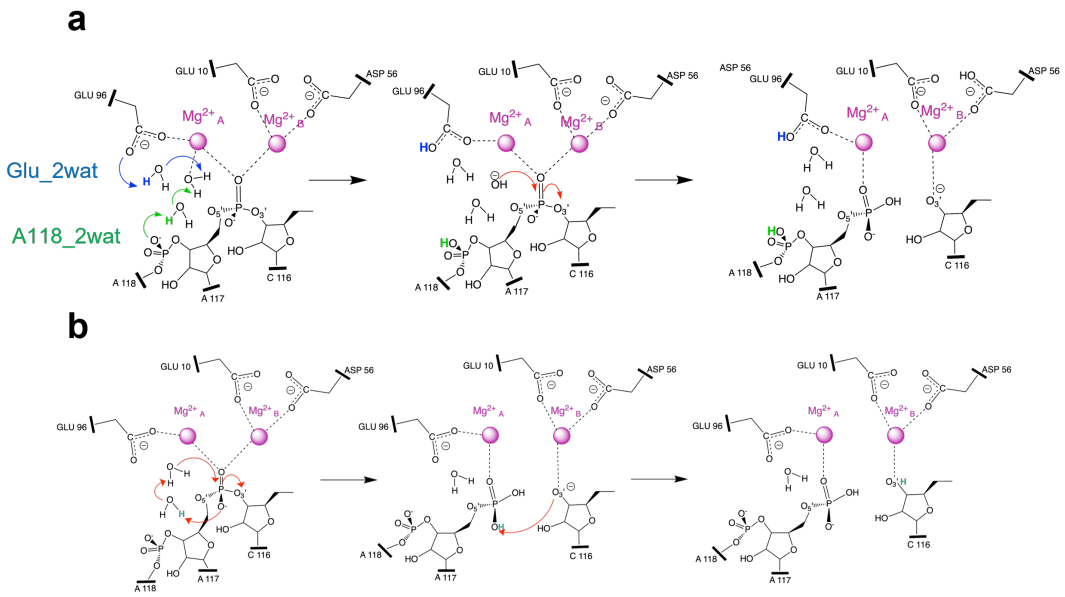

Figure S11: Schematic schemes for additional reaction mechanisms, a) "Glu_2wat", "A118_2wat" or b) proton shuttling through the scissile phosphate and a second water molecule.

puted for all the additional examined mechanisms at the DFTB3/MM-MD level.

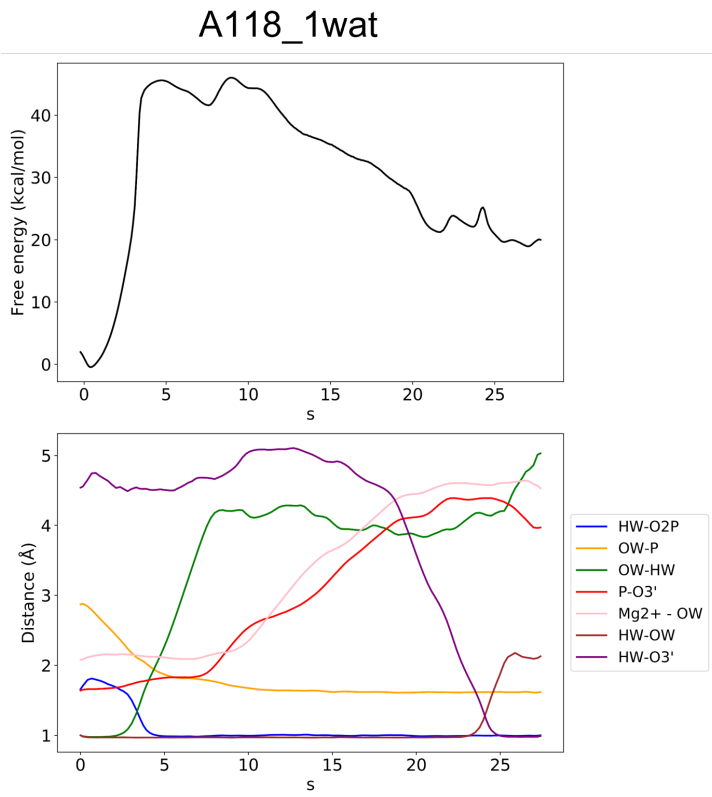

Figure S12: Free energy profile and evotuion of the CVs along the path reaction coordinate for the "A118_1wat" mechanism.

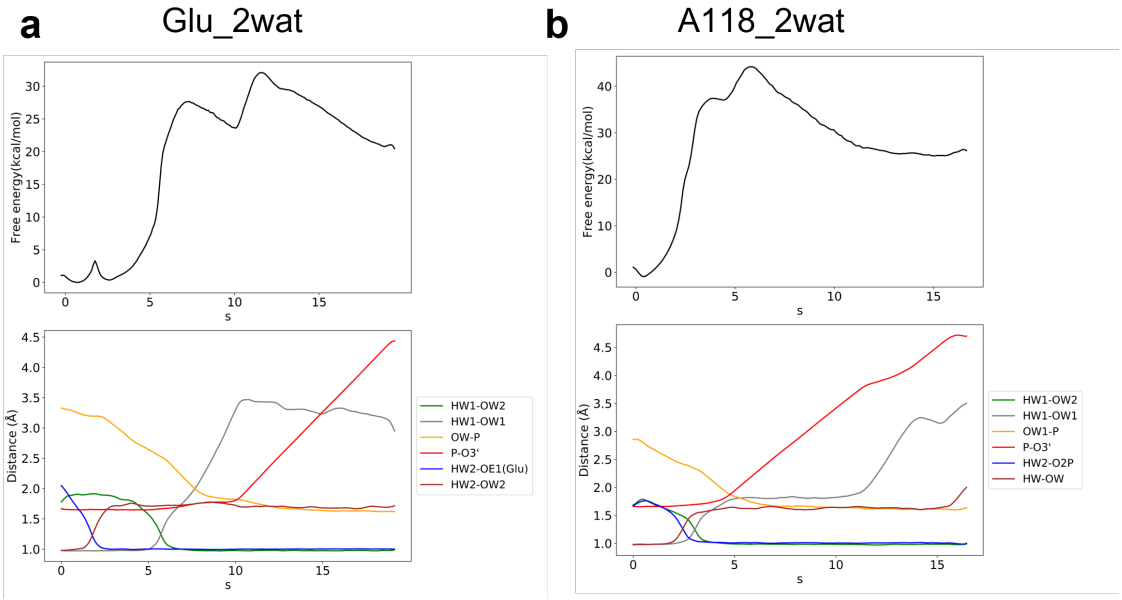

Figure S13: Free energy profile and evotuion of the CVs along the path reaction coordinate for a) "Glu_2wat" and b) "A118_2wat" mechanisms.

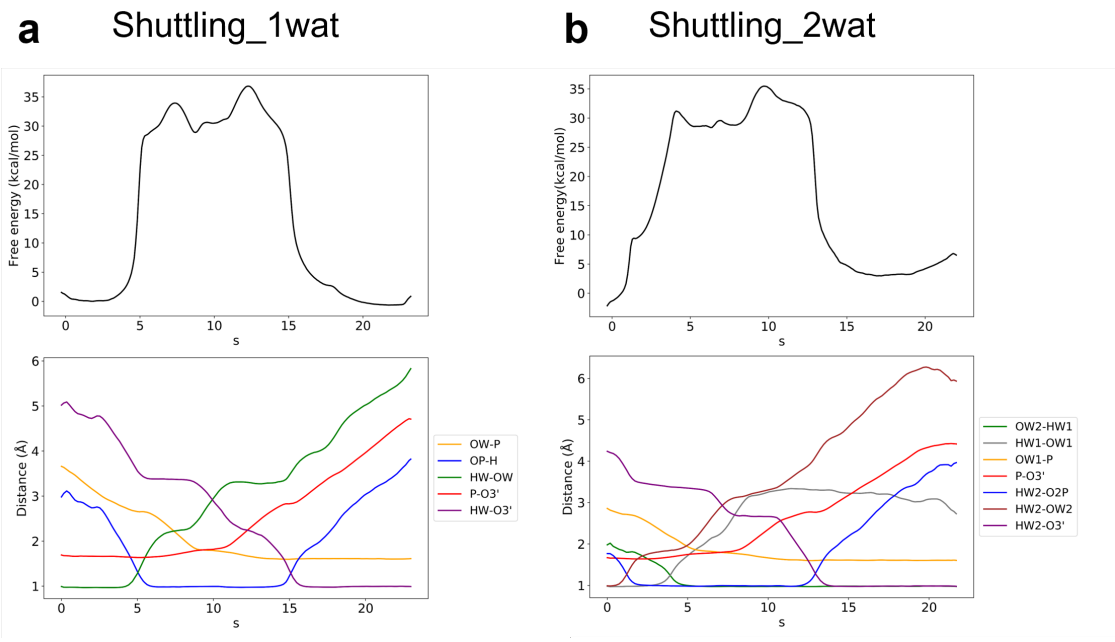

Figure S14: Free energy profile and evotuion of the CVs along the path reaction coordinate for the "Shuttling_1wat" and "Shuttling_2wat" mechanism.

References

(S1) Li, P.; Song, L. F.; Merz, K. M. Parameterization of Highly Charged Metal Ions Using the 12-6-4 LJ-type Nonbonded Model in Explicit Water. *J. Phys. Chem. B* **2015**, *119*, 883–895.

(S2) Grotz, K. K.; Cruz-León, S.; Schwierz, N. Optimized Magnesium Force Field Parameters for Biomolecular Simulations with Accurate Solvation, Ion-Binding, and Water-Exchange Properties. *J. Chem. Theory Comput.* **2021**, *17*, 2530–2540.

(S3) Puyo-Fourtine, J.; Juillé, M.; Hénin, J.; Clavaguéra, C.; Duboué-Dijon, E. A Consistent Picture of Phosphate-Divalent Cation Binding from Models with Implicit and Explicit Electronic Polarization. *J. Phys. Chem. B* **2022**, *126*, 4022–4034.

(S4) Zenodo repository.

(S5) Duboué-Dijon, E.; Delcroix, P.; Martinez-Seara, H.; Hladílková, J.; Coufal, P.; Křížek, T.; Jungwirth, P. Binding of divalent cations to insulin: capillary electrophoresis and molecular simulations. *J. Phys. Chem. B* **2018**, *122*, 5640–5648.

(S6) Nencini, R.; Tempra, C.; Biriukov, D.; Polák, J.; Ondo, D.; Heyda, J.; Ollila, S. O.; Javanainen, M.; Martinez-Seara, H. Prosecco: polarization reintroduced by optimal scaling of electronic continuum correction origin in MD simulations. *Biophys. J.* **2022**, *121*, 157a.

(S7) De Vivo, M.; Dal Peraro, M.; Klein, M. L. Phosphodiester Cleavage in Ribonuclease H Occurs via an Associative Two-Metal-Aided Catalytic Mechanism. *J. Am. Chem. Soc.* **2008**, *130*, 10955–10962.