Supporting information for *A-type helicity controls efficient nonenzymatic template copying of nucleic acids*

Barbara K. Lech, Marie Zgarbová, Petr Jurečka, Rafał Szabla

1 Methodology

Each system was build with use of NAB [1] and the sequence of the template is "AUGGCACGGGCAGUC". The magnesium ion was placed in close proximity to the 3′O and P atoms. The deprotonated 3′OH group was parametrized with the Antechamber program and the parameters for magnesium ions are taken from `https://doi.org/10.1021/ct3000734`. The parameters for the imidazolium bridge have been published in the following study: `https://doi.org/10.1002/syst.202400086`.

All systems were immersed in a truncated octahedral box of SPC/E water molecules [2], ensuring at least 12 Å separation between the RNA and the box edges. The simulations employed the Amber force fields OL3 (RNA) and OL24 (DNA). To achieve electroneutrality, three potassium (K^+) counterions were introduced, along with an additional 0.15 M concentration of KCl, following the parameters described by Joung and Cheatham [3].

The equilibration began with energy minimization, consisting of 500 steps using the steepest descent algorithm and 500 steps with the conjugate gradient method, while heavy atoms were restrained with a force constant of 25 kcal·mol^{-1}. Subsequently, the system was gradually heated from 100 K to 300 K over a 100 ps simulation at constant volume, maintaining the same restraints. In the next phase, minimization was repeated with reduced restraints of 5 kcal·mol^{-1}, again using both steepest descent and conjugate gradient (500 steps each). This was followed by 50 ps of equilibration at constant temperature (300 K) and pressure (1 bar). The minimization and equilibration steps were repeated with stepwise reductions in restraint strength (4, 3, 2, and 1 kcal·mol^{-1}). In the final stage, restraints were lowered to 0.5 kcal·mol^{-1}, after which a 50 ps unrestrained MD simulation was performed under constant temperature and pressure.

For the production phase, six independent 0.5 μs simulations were performed for each system, starting from distinct initial conformations, yielding a total simulation time of 3.0 μs per system. Throughout these simulations, temperature and pressure were controlled using the Langevin thermostat [4] and the Monte Carlo barostat [5], respectively. A 9 Å cutoff was applied for nonbonded interactions, and the integration step was set to 4 fs. This was made possible by employing the SHAKE [6], SETTLE [7], and hydrogen mass repartitioning techniques [8]. All production runs were carried out using the GPU-accelerated pmemd engine from AMBER22 [1].

For trajectory analysis, every 10th frame was selected, excluding water and ions. Clustering was performed using two RMSD-based approaches: DBSCAN [9] and K-means [10]. For DBSCAN, a minimum of 25 points per cluster was required, and the ϵ parameter was optimized via k-distance analysis. In contrast, the K-means algorithm, which partitions the data into a predefined number of clusters, was executed iteratively with cluster counts ranging from 3 to 13. The optimal number was chosen based on the ratio of SSR (sum of squares regression) to SST (total sum of squares). The analysis of helix type, distances, and angles was performed using the Cpptraj module from AmberTools and the MDAnalysis package.

2 Clustering results

In the figures, the primer is shown in purple, the helper in orange, and the template strand is displayed as semi-transparent. The imidazolium-bridged dinucleotide is located between the helper and the primer.

2.1 Protonated systems

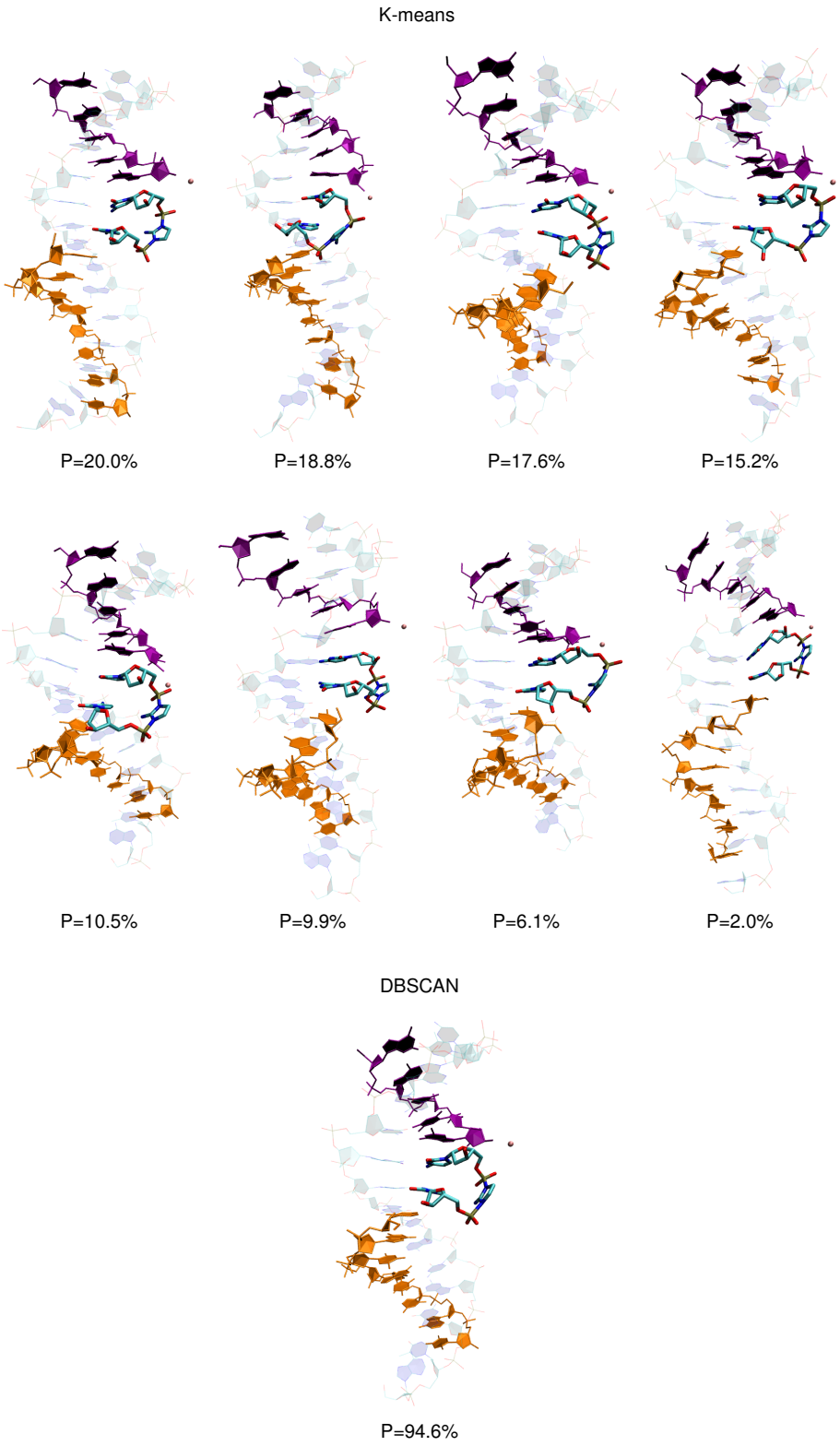

Fig. S 1: DNA structure

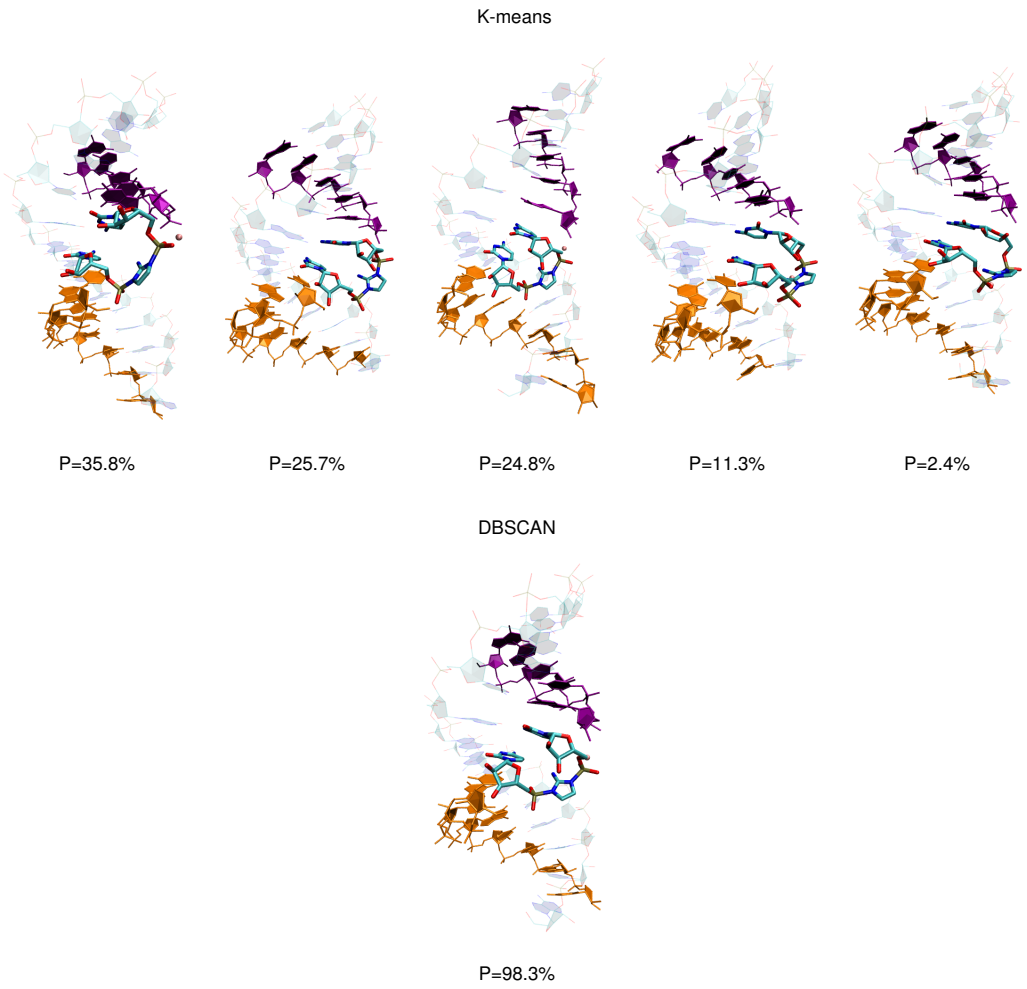

Fig. S 2: RNA structure

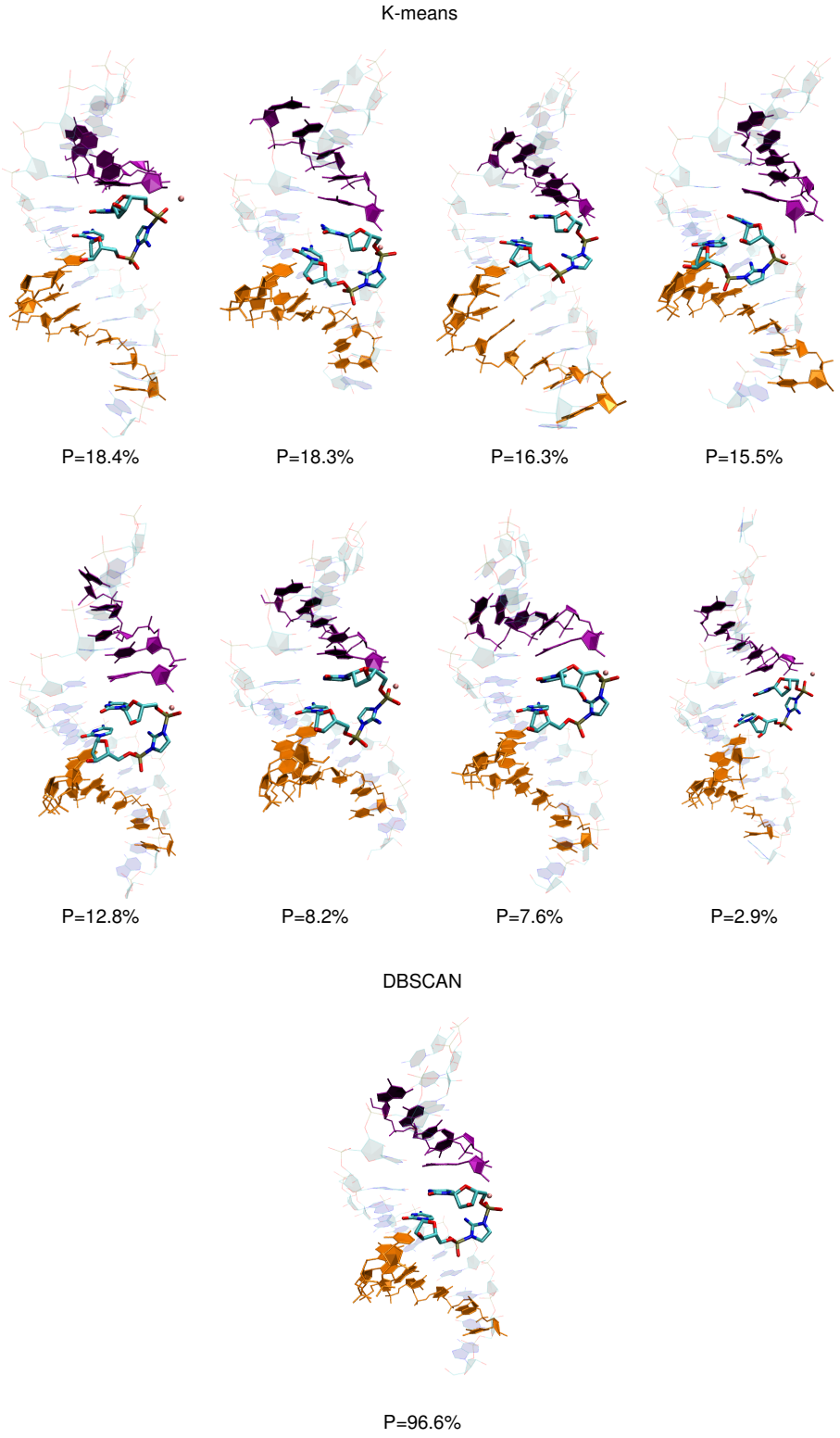

Fig. S 3: D/RNA A structure

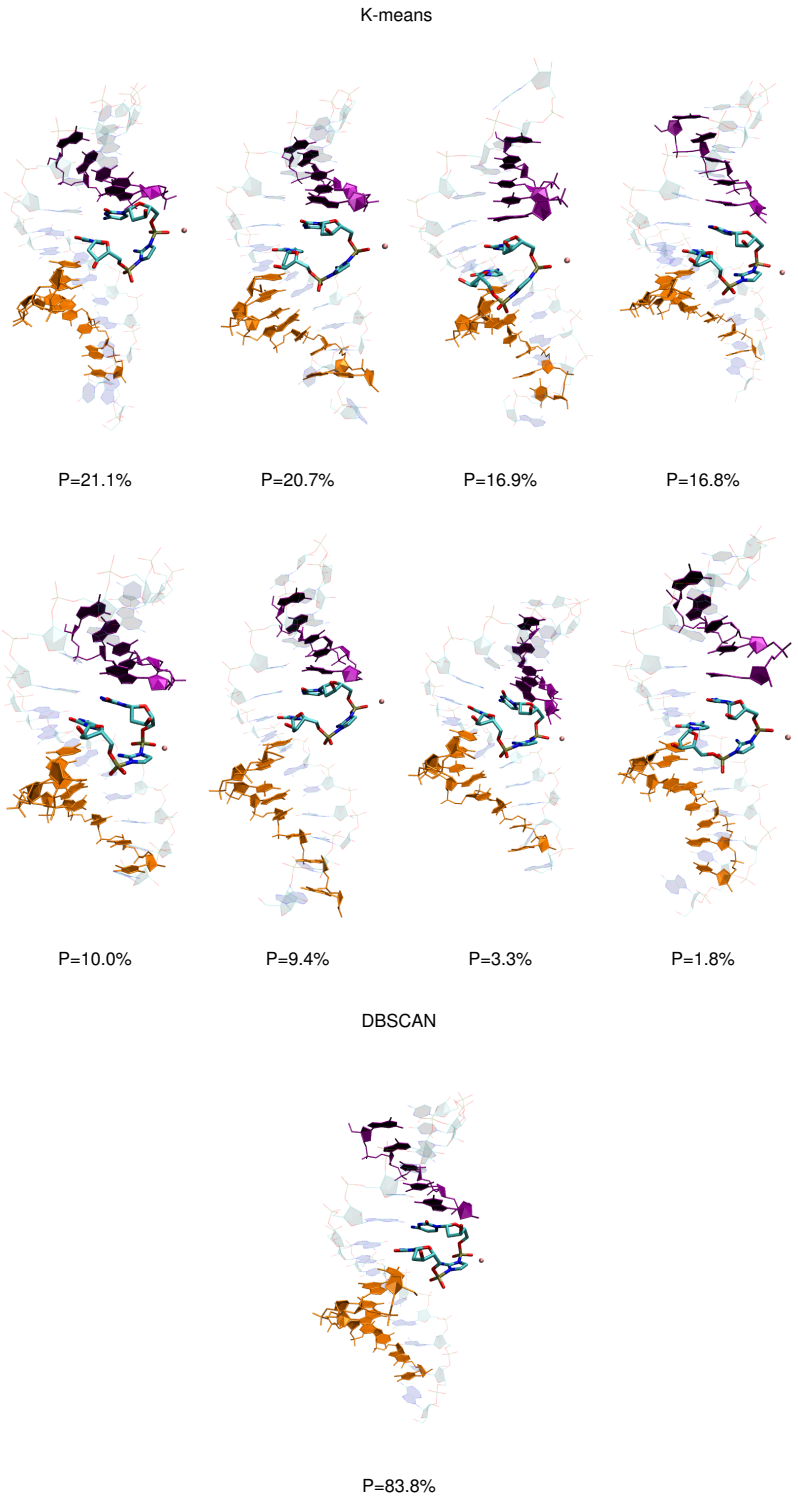

Fig. S 4: D/RNA B structure

2.2 Deprotonated systems

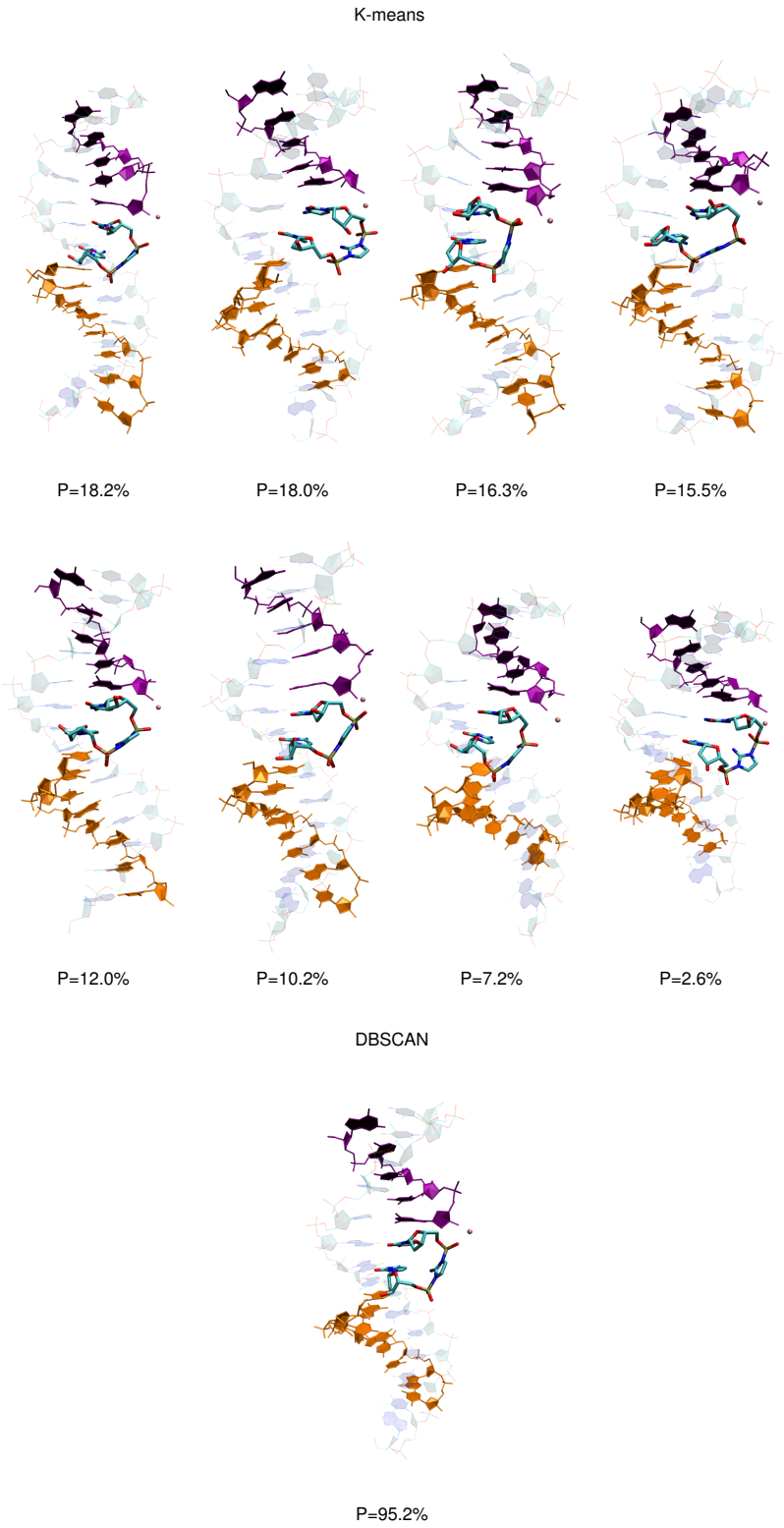

Fig. S 5: DNA structure

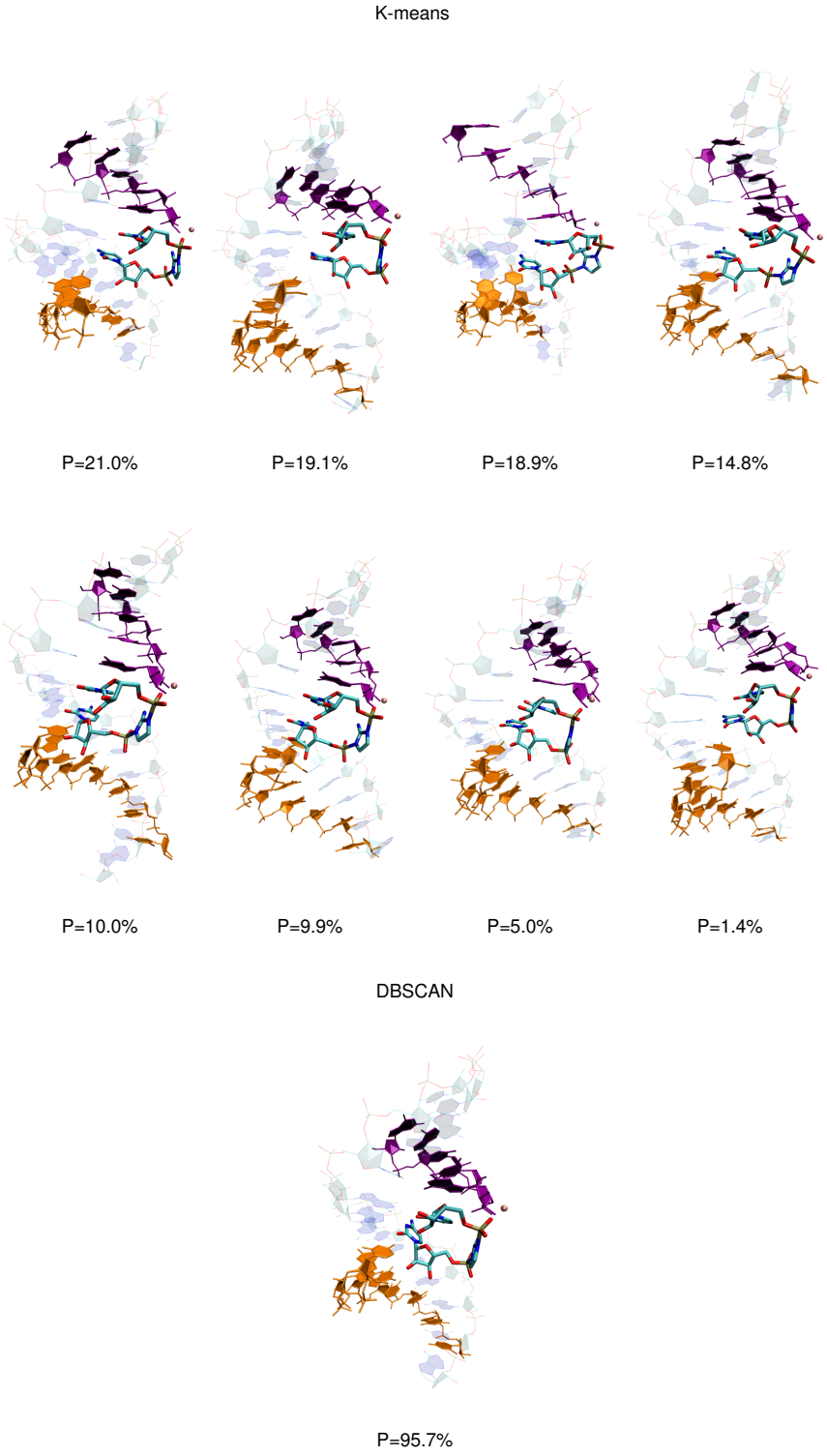

Fig. S 6: RNA structure

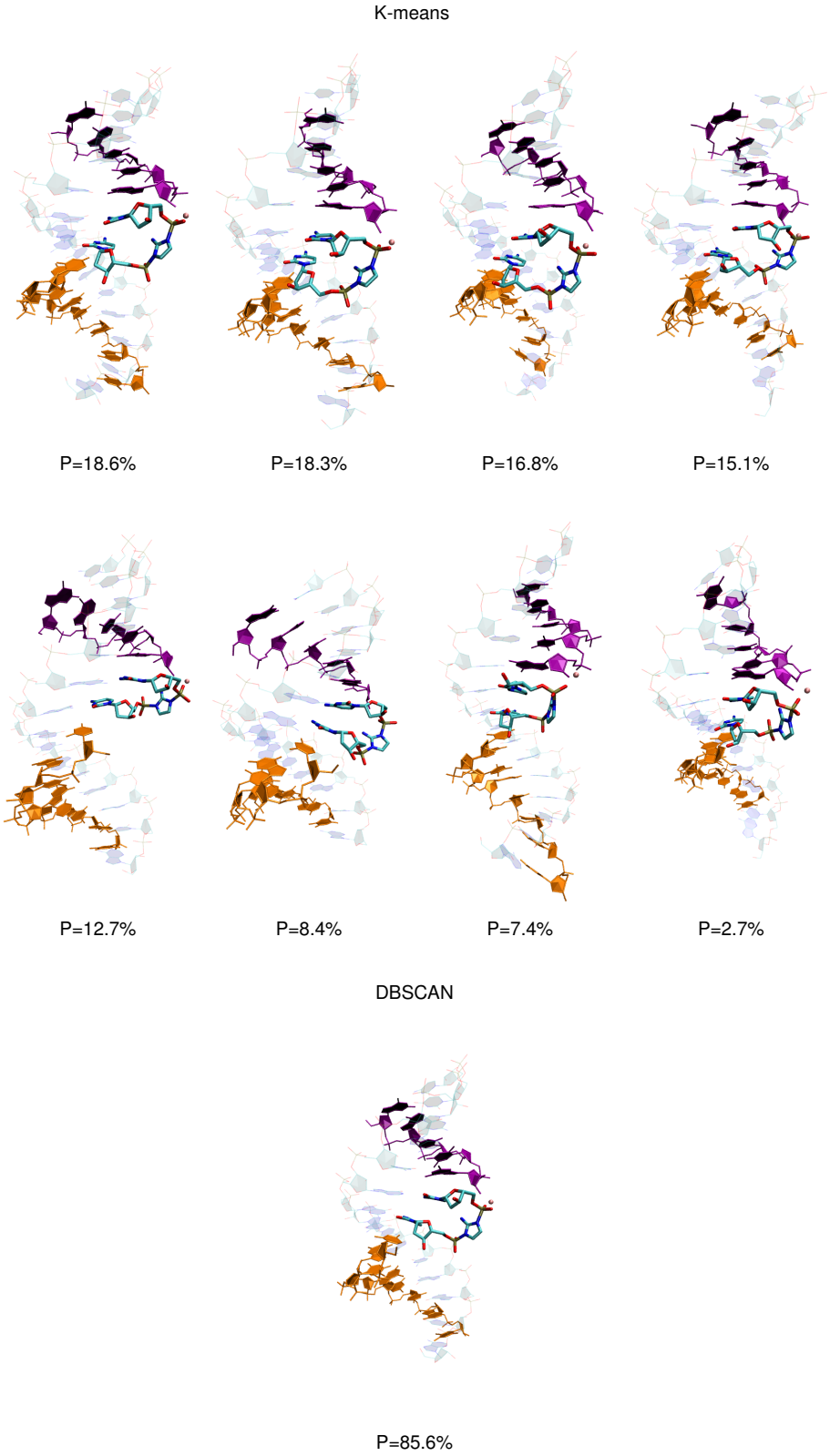

Fig. S 7: D/RNA A structure

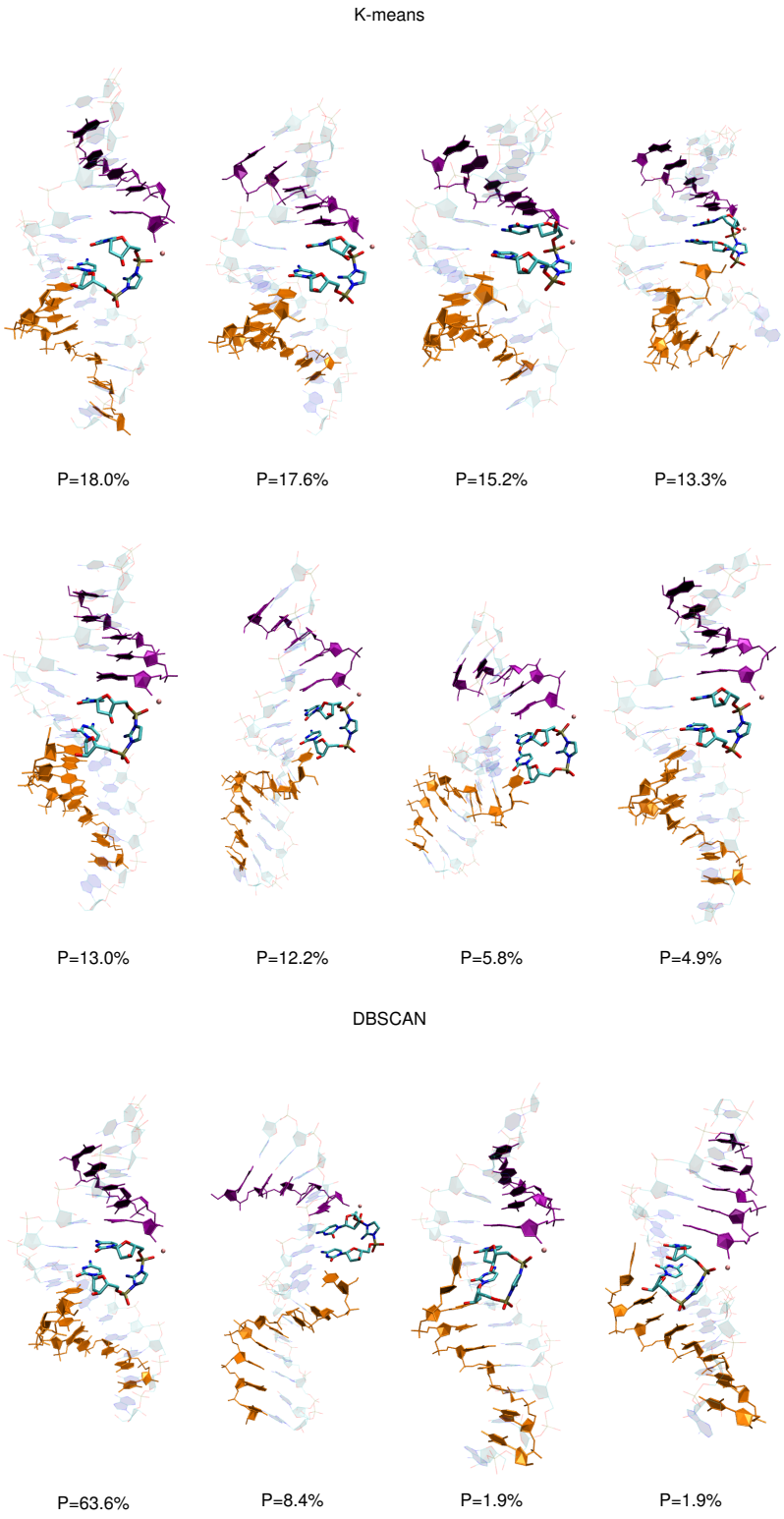

Fig. S 8: D/RNA B structure

3 RMSD analysis of the cavity for the binding of the activated dinucleotide

We performed RMSD analysis of all frames relative to the structures identified by clustering. In the case of DNA, two analyses were carried out: one with respect to the frame where the activated dinucleotide was located near the 3′ end of the primer, and another with respect to the structure where it was positioned near the 5′ end of the helper. For RNA, the analysis was performed relative to the average structure identified by the DBSCAN algorithm. For the analysis, we selected all carbon atoms from four nucleotides: the 3′ terminal nucleotide of the primer, the activated dinucleotide, and the 3′ terminal nucleotide of the helper strand. On the graphs is showed moving average for better visualization.

3.1 Protonated systems

3.1.1 RNA

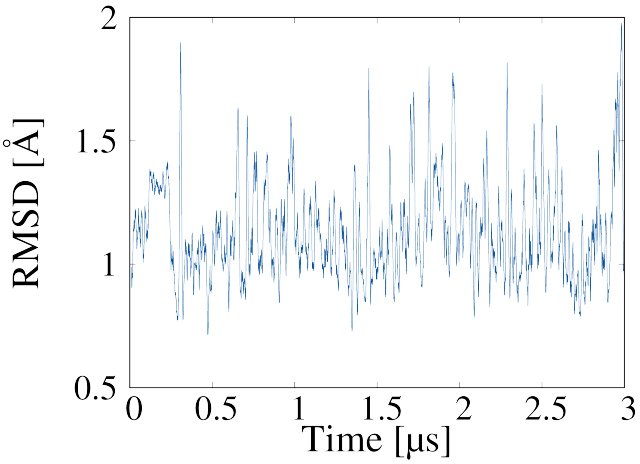

Fig. S 9: RMSD anaylsis with respect to the average structure identified by the DBSCAN algorithm.

3.1.2 DNA

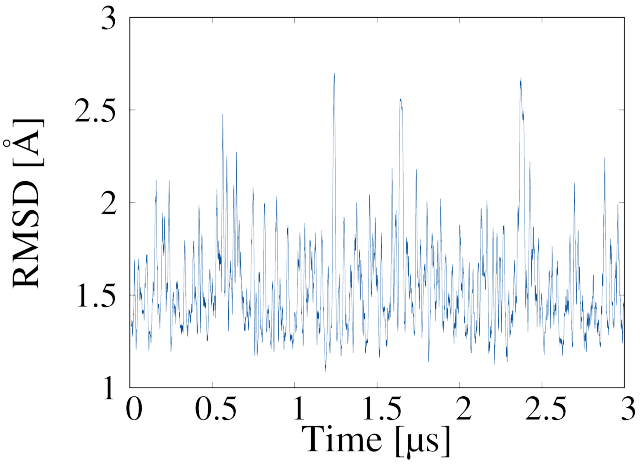

Fig. S 10: RMSD anaylsis with respect to the frame where the activated dinucleotide was located near the 3′ end of the primer (K-means analysis cluster 0).

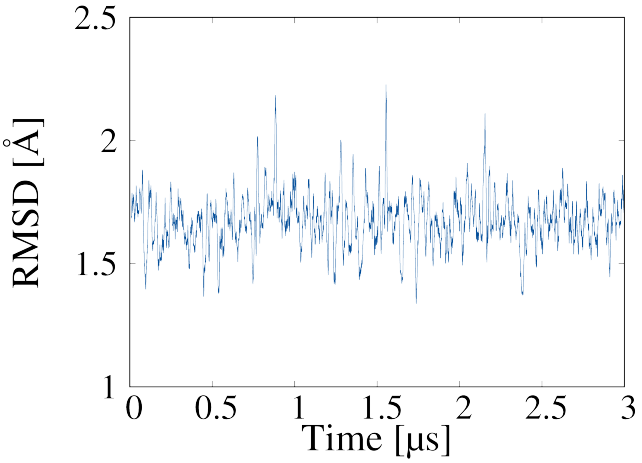

Fig. S 11: RMSD anaylsis with respect to the frame where the activated dinucleotide was located near the 5′ end of the helper (K-means analysis cluster 1).

3.2 Deprotonated systems

3.2.1 RNA

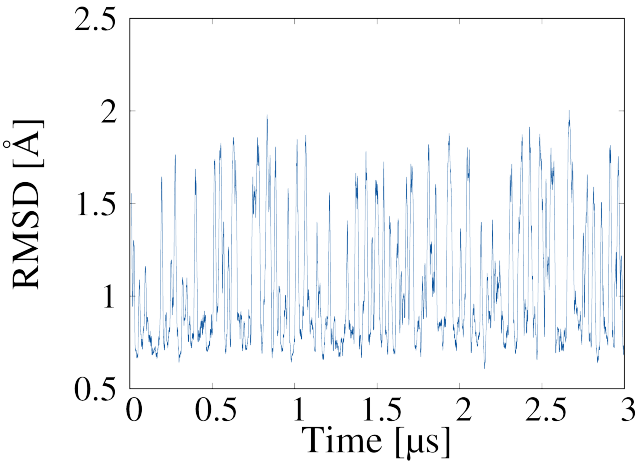

Fig. S 12: RMSD anaylsis with respect to the average structure identified by the DBSCAN algorithm.

3.2.2 DNA

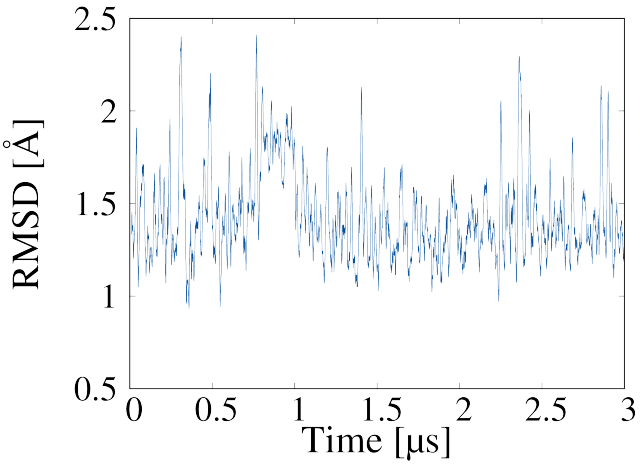

Fig. S 13: RMSD anaylsis with respect to the frame where the activated dinucleotide was located near the 3′ end of the primer (K-means analysis cluster 0).

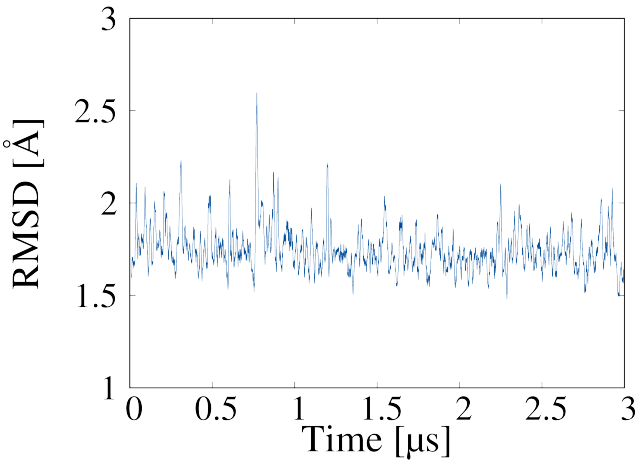

Fig. S 14: RMSD anaylsis with respect to the frame where the activated dinucleotide was located near the 5′ end of the helper (K-means analysis cluster 2).

4 Angle of the attack - deprotonated systems

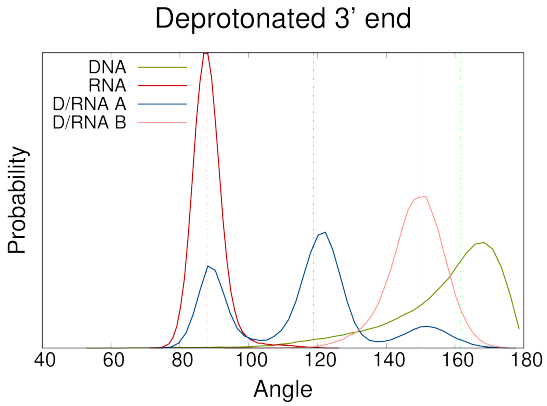

Fig. S 15: Probability distribution of O–P–N angle (attack angle). Dashed lines represents the median values.

5 Distances analysis

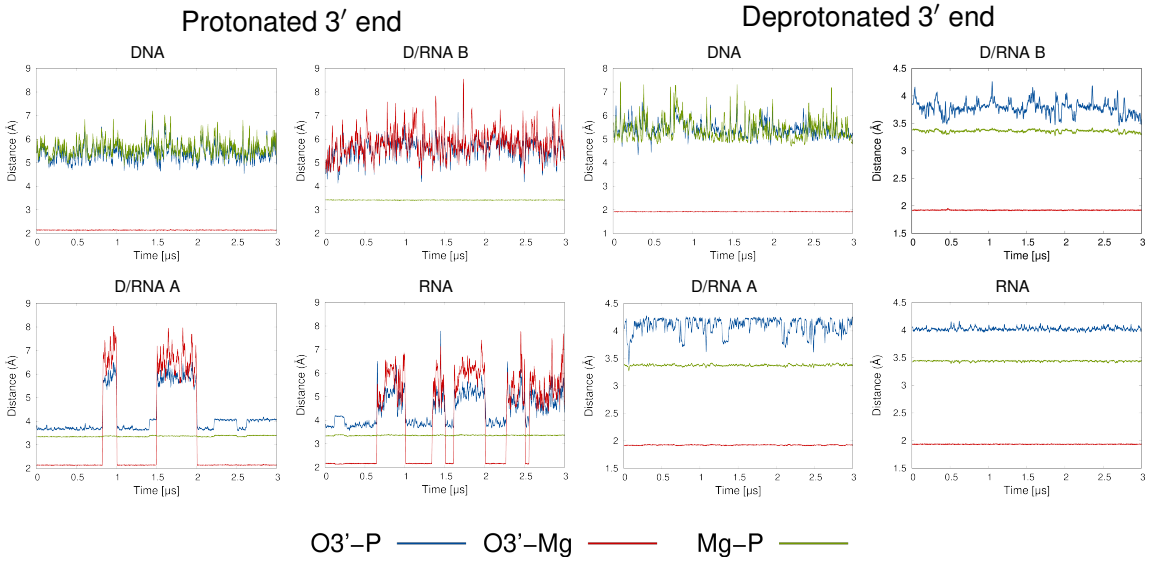

Fig. S 16: Plots representing changes of the distances between O3′–P, O3′–Mg^{2+} and Mg^{2+}–P atoms over simulation time.

6 Examination of the helix type - deprotonated system

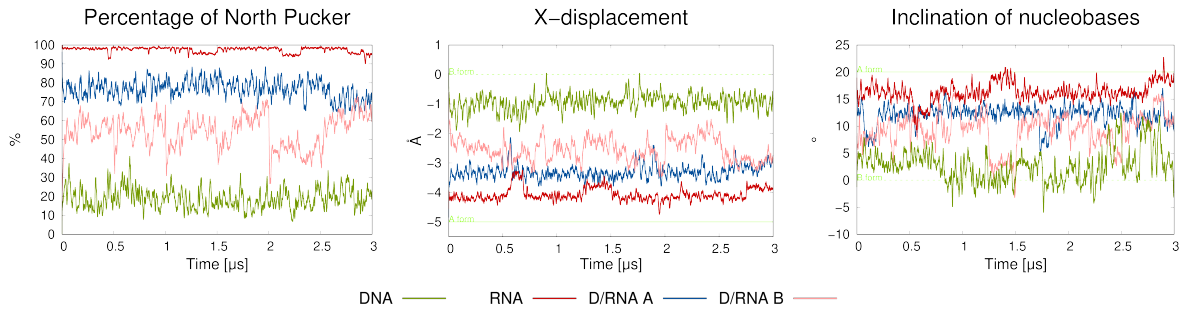

Fig. S 17: Changes of inclination, x-displacement, and percentage of north puckering throughout the simulation for deprotonated systems.

7 Correlation of sugar puckering and O3′–P distance - RNA

The north puckering is defined by angles less than 90° or greater than 270°, while the south puckering corresponds to angles between 90° and 270°.

7.1 Protonated system

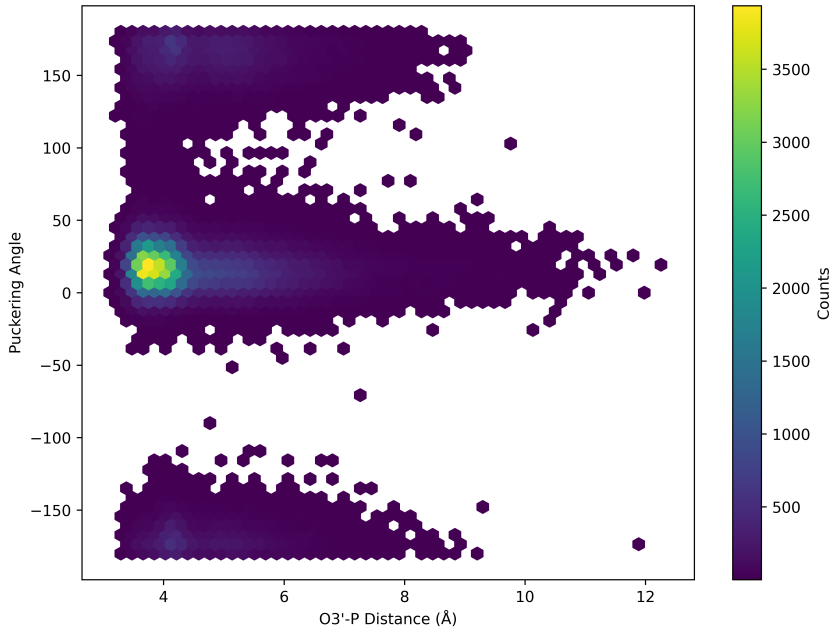

Fig. S 18: Density plots demonstrating the dependence of the puckering angle of 3′ terminal nucleotide and the O3′–P distance generated for the MD simulation of the protonated system.

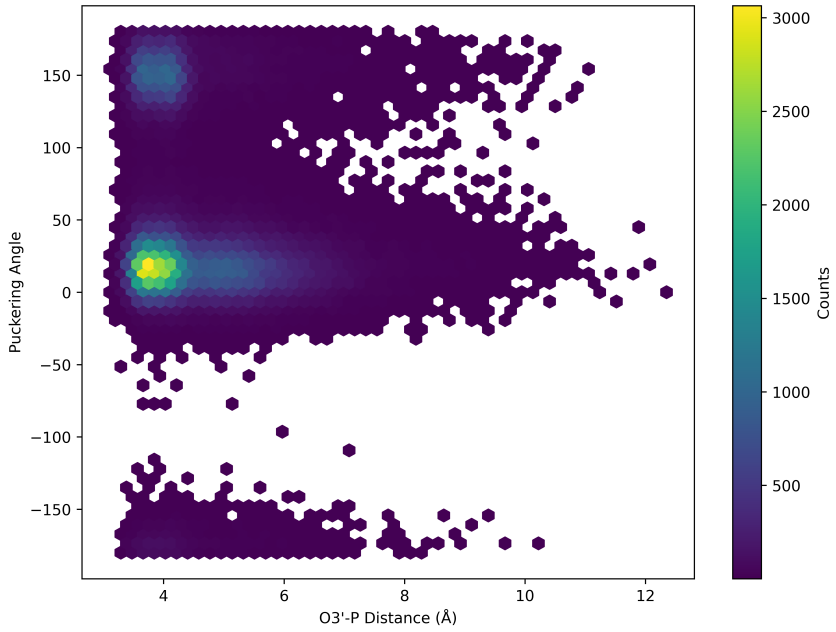

Fig. S 19: Density plots demonstrating the dependence of the puckering angle of adjacent nucleotide and the O3′–P distance generated for the MD simulation of the protonated system.

7.2 Deprotonated system

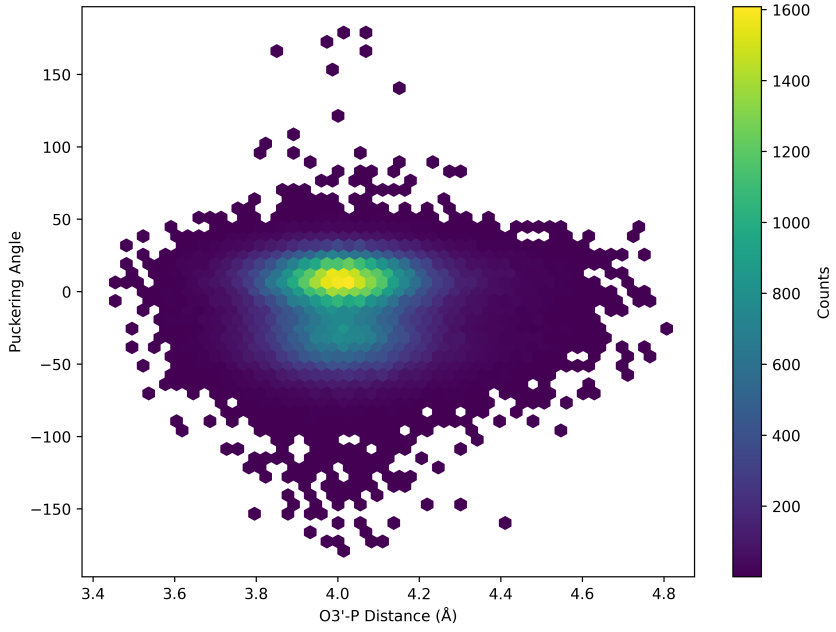

Fig. S 20: Density plots demonstrating the dependence of the puckering angle of 3′ terminal nucleotide and the O3′–P distance generated for the MD simulation of the deprotonated system.

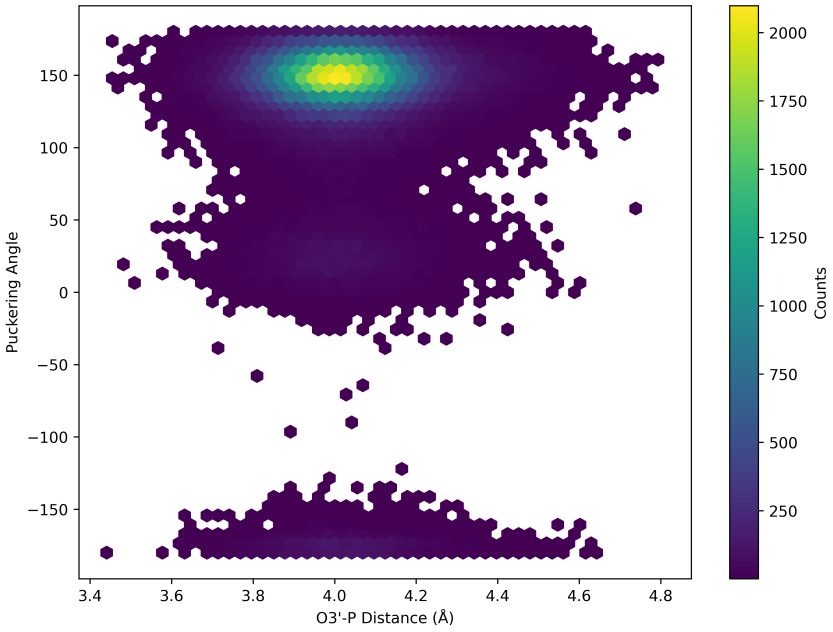

Fig. S 21: Density plots demonstrating the dependence of the puckering angle of adjacent nucleotide and the O3′–P distance generated for the MD simulation of the deprotonated system.

References

[1] David A. Case et al. "AmberTools". In: *J. Chem. Inf. Model.* 63.20 (Oct. 2023), pp. 6183–6191. ISSN: 1549-9596. DOI: 10.1021/acs.jcim.3c01153. (Visited on 12/15/2023).

[2] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. "The Missing Term in Effective Pair Potentials". In: *J. Phys. Chem.* 91.24 (Nov. 1987), pp. 6269–6271. ISSN: 0022-3654. DOI: 10.1021/j100308a038. (Visited on 12/15/2023).

[3] In Suk Joung and Thomas E. III Cheatham. "Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations". In: *J. Phys. Chem. B* 112.30 (July 2008), pp. 9020–9041. ISSN: 1520-6106. DOI: 10.1021/jp8001614. (Visited on 07/24/2024).

[4] G. Parisi. "Correlation Functions and Computer Simulations". In: *Nuclear Physics B* 180.3 (May 1981), pp. 378–384. ISSN: 0550-3213. DOI: 10.1016/0550-3213(81)90056-0. (Visited on 10/15/2024).

[5] Johan Åqvist et al. "Molecular Dynamics Simulations of Water and Biomolecules with a Monte Carlo Constant Pressure Algorithm". In: *Chemical Physics Letters* 384.4 (Jan. 2004), pp. 288–294. ISSN: 0009-2614. DOI: 10.1016/j.cplett.2003.12.039. (Visited on 10/15/2024).

[6] Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J. C Berendsen. "Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes". In: *Journal of Computational Physics* 23.3 (Mar. 1977), pp. 327–341. ISSN: 0021-9991. DOI: 10.1016/0021-9991(77)90098-5. (Visited on 12/15/2023).

[7] Shuichi Miyamoto and Peter A. Kollman. "Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models". In: *Journal of Computational Chemistry* 13.8 (1992), pp. 952–962. ISSN: 1096-987X. DOI: 10.1002/jcc.540130805. (Visited on 12/15/2023).

[8] Chad W. Hopkins et al. "Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning". In: *J. Chem. Theory Comput.* 11.4 (Apr. 2015), pp. 1864–1874. ISSN: 1549-9618. DOI: 10.1021/ct5010406. (Visited on 12/15/2023).

[9] Jun-hui Peng et al. "Clustering Algorithms to Analyze Molecular Dynamics Simulation Trajectories for Complex Chemical and Biological Systems†". In: *Chinese Journal of Chemical Physics* 31.4 (Aug. 2018), pp. 404–420. ISSN: 1674-0068. DOI: 10.1063/1674-0068/31/cjcp1806147. (Visited on 07/24/2024).

[10] John A Hartigan and Manchek A Wong. "A K-means Clustering Alrorithm". In: *Appl. Stat* 28 (1979), pp. 126–130.