

# Proton Transfers to DNA in Native Electrospray Ionization Mass Spectrometry: A QM/MM Study

*Mirko Paulikat<sup>a</sup>, Juan Aranda<sup>b</sup>, Emiliano Ippoliti<sup>a</sup>, Modesto Orozco<sup>b,c,\*</sup>, Paolo Carloni<sup>a,d,\*</sup>*

<sup>a</sup> Computational Biomedicine (IAS-5/INM-9), Forschungszentrum Jülich GmbH, Wilhelm-  
Johnen-Straße, 52428 Jülich, Germany

<sup>b</sup> Institute for Research in Biomedicine (IRB) Barcelona, The Barcelona Institute of Science and  
Technology, Baldíri Reixac 10, 08028 Barcelona, Spain

<sup>c</sup> Department of Biochemistry and Biomedicine, University of Barcelona, Avinguda Diagonal  
645, 08028, Spain

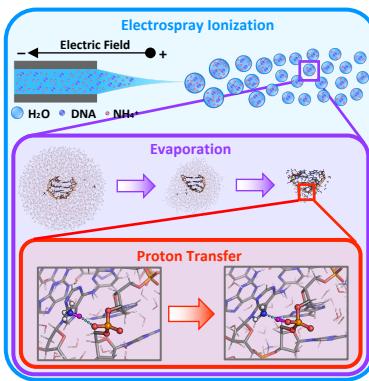
<sup>d</sup> Department of Physics, RWTH Aachen University, Otto-Blumenthal-Straße, 52062 Aachen,  
Germany

## Corresponding Authors

\* Email: [modesto.orozco@irbbarcelona.org](mailto:modesto.orozco@irbbarcelona.org), [p.carloni@fz-juelich.de](mailto:p.carloni@fz-juelich.de)

**ABSTRACT.** Native electrospray ionization - ion mobility mass spectrometry (N-ESI/IM-MS) is a powerful approach for low-resolution structural studies of DNAs in the free state and in complex with ligands. Solvent vaporization is coupled with proton transfers from ammonium ions to the DNA resulting in a reduction of the DNA charge. Here we provide insight on these processes by classical MD and QM/MM free energy calculations on the (GpCpGpApApGpC) heptamer, for which a wealth of experiments is available. Our multiscale simulations, consistent with experimental data, reveal a highly complex scenario: the proton either sits on one of the molecules or is fully delocalized on both, depending on the level of hydration of the analytes and on size of the droplets formed during the electrospray experiments. This work complements our previous study on the *intramolecular* proton transfer on the same heptamer occurring after the processes studied here, and, together, provide a first molecular view of proton transfer in N-ESI/IM-MS.

## TOC GRAPHICS



**KEYWORDS.** QM/MM calculations, mass spectrometry, molecular dynamics, DNA, ammonium ions

Native electrospray ionization - ion mobility mass spectrometry (N-ESI/IM-MS) is a very powerful tool for low-resolution structural studies of DNA oligomers and their complexes with ligands.<sup>1-4</sup> It requires much smaller quantities than high resolution techniques, such as NMR, X-ray crystallography or cryo-electron microscopy.<sup>5,6</sup> The biomolecular analytes are sprayed from solution in quasi-physiological conditions (aqueous ammonium solutions at neutral pH, without organic cosolvents)<sup>7,8</sup> into the gas phase through a capillary applied to an electric field.<sup>9,10</sup> Water droplets are formed and eventually evaporate up to a complete loss of solvation.<sup>11</sup> Originally, ammonia is positively and phosphates are negatively charged, but as solvent evaporates the ammonium ions transfer protons to the DNA phosphates, consistently with the gas phase basicities of NH<sub>3</sub> (819 kJ·mol<sup>-1</sup>)<sup>12</sup> and dimethyl phosphate (DMP; 1,360 kJ·mol<sup>-1</sup>).<sup>13</sup> This leads to a decrease in absolute value of the analyte's charge (along with a loss of NH<sub>3</sub>), in turn associated with a high signal-to-noise ratio.<sup>14</sup>

Molecular simulation (MD) studies have provided important insight on DNA oligomers in the gas phase.<sup>15</sup> They suggested that duplex,<sup>16</sup> triplex,<sup>17</sup> G-quadruplex<sup>18</sup> oligomers maintain several structural determinants on passing from the solution to the gas phase. They have also described the impact of the vaporization process in the structural ensemble.<sup>19</sup> Furthermore, when combined with quantum calculations, MD simulations show that once protons are transferred to the DNA, they can jump from different basic sites in the oligos, leading to some structural changes and to a complex fuzzy charge density in the analyte.<sup>20</sup> The remaining question is then how the intermolecular proton transfer events, from ammonium to DNA, occur during vaporization and what is their structural impact.

Here we perform molecular simulations on d(GpCpGpApApGpC) heptamer and ammonium ions in water and gas phase to investigate ammonium-DNA proton transfers. First, we simulate

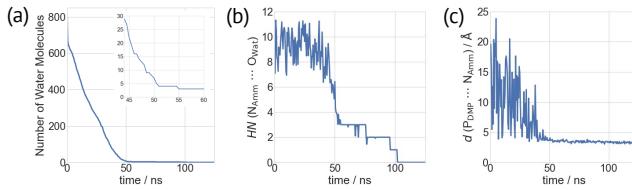
the evaporation process under native ESI-MS conditions by batches of classical MD simulations of the heptamer immersed in a water droplet containing ammonium ions. We remove dissociating water molecules progressively. Second, we study the proton transfer phenomena between the ammonium ion and the heptamer -represented either as DMP (**I**, Fig. 2) or in its integrity (**II**, Fig. 4)- in the partially dehydrated droplets. To this aim, we use quantum mechanics/molecular mechanics (QM/MM) MD and umbrella-sampling-based free energy simulations, exploiting QM/MM interfaces recently developed with the support of the EU projects BioExcel and BioExcel-2 ([www.bioexcel.eu](http://www.bioexcel.eu)).

**MD simulations of complex I.** Here, the heptamer is represented as dimethyl phosphate (DMP). First, we run 10ns-long force field-based simulations at 300 K and 1 bar of ammonium and DMP (as ionic moieties) in water. Both are fully solvated, as shown by the plots of radial distribution functions (rdf, Fig. S1). Integration of the latter shows that the average number of waters surrounding the ammonium ion (hydration number, *HN*) is 8, and the DMP solvation structure shows 4–5 water molecules in the first shell for each anionic oxygen atom. The total hydration number of DMP, estimated from the oxygen-phosphate rdf, counts 10 water molecules (Fig. S1). The distance between ammonium nitrogen and DMP phosphorus atom (*d*) varies widely from 4 (contact ion pairs) to 30 Å (size of the simulation box). The ion pairs are of very transient nature, lasting typically less than 1 ns (Fig. S2).

Next, we let three droplets containing ~800 water molecules (extracted from selected MD snapshots) evaporate by MD simulations at 300 K. Although most waters dissociate quickly from the droplet (Fig. 1a, Fig. S3), the ions remain fully solvated until ~45 ns (8 coordinating water molecules). Further evaporation leads to only 3 coordinating water molecules (Fig. 1b). The last water molecule dissociates after 102 ns at an increased temperature of 450 K. The distance *d*

decreases as the droplet shrinks; transient ion pairs are formed several times during the dynamics.

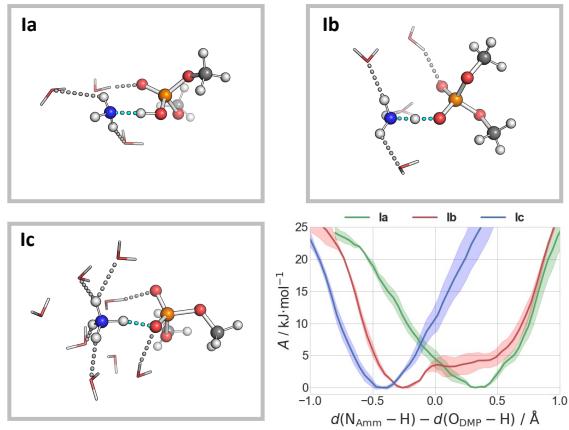
After  $\sim 45$  ns, the two ions form direct H-bonds (Fig. 1c).



**Figure 1:** Gas phase MD simulations of complex **I**: different quantities plotted as a function of simulated time: (a) Number of water molecules present in the droplet. Inset: Enlarged view from 45 ns to 60 ns, when permanent ion pairs are formed. (b) Hydration number ( $HN$ ) of the ammonium ion. (c) Distance  $d$  between the DMP-phosphorus atom ( $P_{DMP}$ ) and the  $\text{NH}_4^+$ -nitrogen atom ( $N_{Amm}$ ).

**QM/MM free energy calculations of complex I.** After 2.4 ps-long QM/MM MD re-equilibration of the system (see SI, Fig. S4), we carry out US free energy calculations at 300 K with different water content, from 3 (**Ia**) to 4 (**Ib**) and 8 (**Ic**) (Fig. 2). The collective variable (CV) used here is the difference between the N–H breaking bond distance and the O–H forming bond distance (Chart S1): negative and positive values of the CV are associated with ionic and neutral states, respectively. The free energy profile is well converged, as shown by its time evolution (Fig. S5).

In **Ia**, the proton is localized on DMP with a poor hydration of the ammonia moiety ( $HN$  of  $\sim 2$ –3, see Fig. 2). By slightly increasing the hydration ( $HN \sim 3.5$ ) the proton is shared by the two moieties (**Ib**). Further increasing the hydration ( $HN \sim 5$  in **Ic**) leads to the stabilization of the ammonium ion (Fig. 2). Thus, the protonation state is very sensitive to the local solvation environment: only if the ammonium is largely solvated does the cation keep the proton.



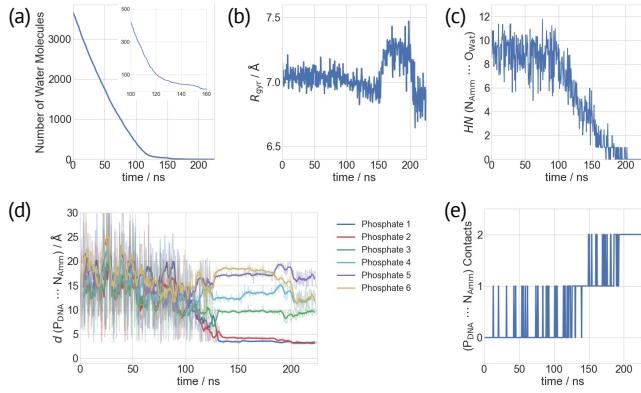
**Figure 2:** Snapshots of the three systems **Ia**, **Ib**, and **Ic**, indicating the preferred protonation state of the ammonium and DMP moieties. H-bonds are shown as dashed lines. Lower right: Free energies as a function of the difference of the breaking/forming  $\text{N}_{\text{Amm}}-\text{H}$  and  $\text{H}-\text{O}_{\text{DMP}}$  bond distances.

**MD simulations of complex II.** A 500ns-long classical MD simulation, in the same conditions as above, is carried out on d(GpCpGpApApGpC) and six ammonium ions in aqueous solution. The overall charge of the system is zero.

The results are similar to previous studies on the biomolecule alone in water (see SI).<sup>21,22</sup> In particular, the oligonucleotide retains its compact hairpin structure, which consists of a short B-DNA segment and a sharp turn within the d(G<sub>3</sub>A<sub>4</sub>A<sub>5</sub>)-triloop with one non-canonical d(G·A) base pair.<sup>21,22</sup> The radius of gyration ( $R_{\text{gyr}}$ ) is  $7.05 \pm 0.06$  Å (Fig. S6–7). Three and two H-bonds are observed regularly for the two canonical base pairs (G<sub>1</sub>–C<sub>7</sub> and C<sub>2</sub>–G<sub>6</sub>) and the non-canonical base pair G<sub>3</sub>–A<sub>5</sub>, respectively.

The two ionic moieties are well separated: Only 1.4% of these are transient ion pairs (Fig. S6), with an average hydration number equal to 8 for both species (Fig. S1).

Three water droplets with three ammonium ions (Fig. S8, total charge of  $-3e$ , the main charge state observed in ESI-MS experiments),<sup>20</sup> extracted by the MD simulations, underwent 225 ns of classical MD simulations. As all simulations showed very similar features, only one is discussed here (details for the others in the SI, Fig. S9–S13). During the first 150 ns (at 300 K), several water molecules strip out (Fig. 3a), but the oligomer maintains its intramolecular H-bonds as in water solution (Fig. S11–S13). The ions are fully solvated within the first 100 ns, but in the next 50 ns, the ammonium  $HN$  decreases from 8 to 2. Simultaneously, the oligomer becomes more compact ( $R_{\text{gyr}} \sim 6.9 \text{ \AA}$  compared to  $\sim 7.05 \text{ \AA}$  in solution). The two ions get closer as the water droplets shrink, as observed for **I** (Fig. 3d). The ammonium ion forms two types of direct H-bonds with the phosphate moieties, associated with different hydration numbers (Fig. 3e): with only one phosphate group (interaction mode (**P**), Fig. 4a-b and Fig. S14a) or with two phosphate groups from adjacent nucleotides (interaction mode (**P2**), Fig. 4c-d and Fig. S14b). After  $\sim 150$  ns, about 30 water molecules are present. By increasing the temperature up to 450 K, the last water molecules evaporate. The oligomer becomes even more compact ( $R_{\text{gyr}} \sim 6.8 \text{ \AA}$ ). The canonical and non-canonical base pairs may dissociate forming other hydrogen bonds (see Fig. S11–S13). Such compact conformation is similar to that of the same system without the ammonium group.<sup>20</sup>



**Figure 3:** Gas phase MD simulations of complex **II** – different quantities plotted as a function of simulated time: (a) Number of water molecules. Inset: Enlarged view from 100 ns to 160 ns. (b) Radius of gyration ( $R_{\text{gyr}}$ ) of the DNA oligo. (c)  $HN$  values of  $\text{NH}_4^+$ -nitrogen atoms ( $N_{\text{Amm}}$ ) surrounded by water-oxygen atoms ( $O_{\text{Wat}}$ ). (d) Distances of the DNA-phosphorus atoms ( $P_{\text{DNA}}$ ) and one  $N_{\text{Amm}}$ -atom. Data at the end of each 500ps-long batch of MD simulations are shown as transparent lines, while their moving averages (window size of 10) are represented as bold lines. (e) Number of contacts of one  $\text{NH}_4^+$  ion with the DNA phosphate groups.

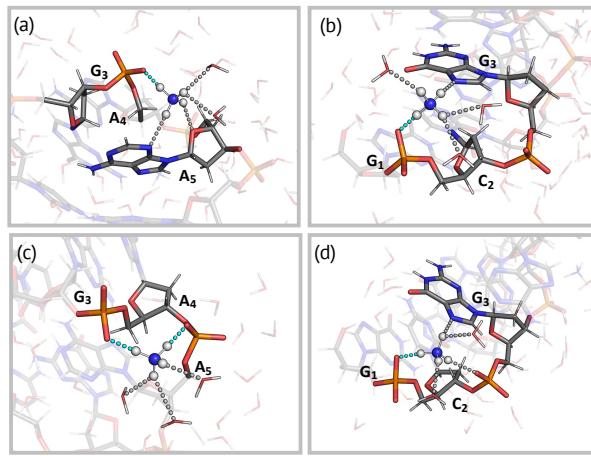
**QM/MM MD free energy calculations of complex **II**.** We consider here 10 systems, differing in the hydration level (from  $\sim 80$  to  $\sim 30$  water molecules, as obtained in our MD simulations, see Tab. 1)<sup>1</sup> and on whether they interact with one phosphate (interaction mode (**P**)) or with two (interaction mode (**P2**)). We calculate the free energy associated with proton transfer in each mode.

<sup>1</sup> As expected, the larger the number of water molecules, the higher the hydration number values for the ammonium ion and the DNA tend to be.

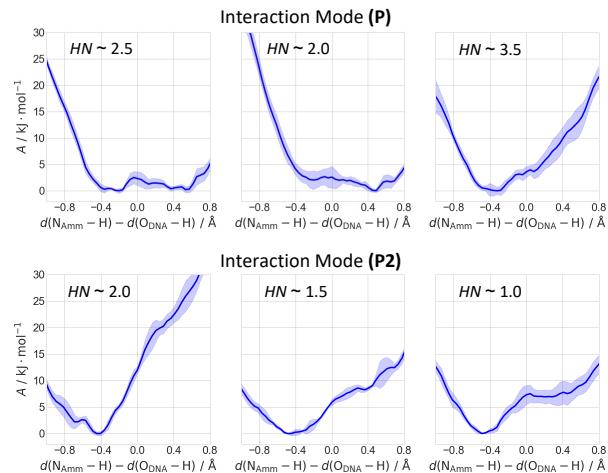
**Table 1:** Complex **II** featuring interaction modes (**P**) or (**P2**) at different levels of hydration, undergoing QM/MM simulations.  $N_{\text{Wat}}$  is the total number of water molecules. The  $HN$  values are calculated from the last 1.25 ps of the equilibration phase. The interaction modes are depicted in Figure 4.

|            | $N_{\text{Wat}}$ | $HN$ ( $\mathbf{N}_{\text{Amm}} \cdots \mathbf{O}_{\text{Wat}}$ ) | $HN$ ( $\mathbf{P}_{\text{DNA}} \cdots \mathbf{O}_{\text{Wat}}$ ) | Interaction Mode | Interacting Nucleotides                              |
|------------|------------------|---|---|------------------|--|
| <b>IIa</b> | 82               | 2.5   | 2.0   | <b>P</b>         | $\text{G}_3\text{-p-}\text{A}_4$                     |
| <b>IIb</b> | 55               | 2.5   | 1.5   | <b>P</b>         | $\text{G}_3\text{-p-}\text{A}_4$                     |
| <b>IIc</b> | 50               | 2.0   | 2.0   | <b>P</b>         | $\text{G}_3\text{-p-}\text{A}_4$                     |
| <b>IId</b> | 79               | 3.5   | 4.0   | <b>P</b>         | $\text{G}_1\text{-p-}\text{C}_2$                     |
| <b>IIe</b> | 64               | 4.0   | 0.5   | <b>P</b>         | $\text{G}_1\text{-p-}\text{C}_2$                     |
| <b>IIf</b> | 67               | 2.0   | 2.0/2.0   | <b>P2</b>        | $\text{G}_3\text{-p-}\text{A}_4\text{-p-}\text{A}_5$ |
| <b>IIg</b> | 44               | 1.5   | 2.0/1.5   | <b>P2</b>        | $\text{G}_3\text{-p-}\text{A}_4\text{-p-}\text{A}_5$ |
| <b>IIh</b> | 32               | 1.0   | 1.0/2.5   | <b>P2</b>        | $\text{G}_3\text{-p-}\text{A}_4\text{-p-}\text{A}_5$ |
| <b>IIi</b> | 49               | 1.5   | 2.0/5.0   | <b>P2</b>        | $\text{G}_1\text{-p-}\text{C}_2\text{-p-}\text{G}_3$ |
| <b>IIj</b> | 32               | 1.5   | 0.5/2.5   | <b>P2</b>        | $\text{G}_1\text{-p-}\text{C}_2\text{-p-}\text{G}_3$ |

In the first case (Fig. 4a), for  $HN \sim 2.5$ , the proton is delocalized between the oligomer and the ammonium (**IIa-b**, Fig. 5 and S16). For  $HN \sim 2$ , the proton is still delocalized between the two moieties, but it mostly sits onto the oligomer (**IIc**, Fig. 5). Thus, decreasing solvation leads the proton towards the oligonucleotide, likely because of the lack of stabilization of the charged ammonium group from the surrounding water molecules.



**Figure 4.** Snapshots of proton transfer configurations showing the interactions of the ammonium ion with (a) the G<sub>3</sub>-p-A<sub>4</sub> (**IIa-c**), (b) the C<sub>2</sub>-p-G<sub>3</sub> (**IId-e**), (c) the G<sub>3</sub>-p-A<sub>4</sub>-p-A<sub>5</sub> (**IIf-h**), and (d) the G<sub>1</sub>-p-C<sub>2</sub>-p-G<sub>3</sub> (**IIIi-j**) moieties. H-bond interactions are shown as dashed lines. The H-bonds involved in the proton transfer are colored in cyan.



**Figure 5:** Free energy as a function of the difference of the bond breaking/forming N<sub>Amm</sub>-H and H-O<sub>DNA</sub> bond distances. Top: Representative profiles of interaction mode (**P**) at different HNs, showing systems **IIa** (left), **IIc** (middle) and **IId** (right). Bottom: Representative profiles of interaction mode (**P2**) at different HNs, showing the protonation of the G<sub>3</sub>-p-A<sub>4</sub> moiety in systems **IIIf** (left), **IIIf** (middle) and **IIIf** (right). The complete set of free energy profiles are shown in the SI (Fig. S16).

In the second case (**IIId-e** in Fig. 4b), the ammonium cation is markedly more hydrated than in the first ( $HN \sim 3.5\text{--}4.0$ , see Tab. 1). As a result, the proton is now *only* localized on the ammonium group (highly stabilized by its solvation shell) and the oligomer is not protonated (Fig. 5).<sup>2</sup>

In the interaction mode **P2**, the ammonium group is less hydrated than in **P**. Either it interacts with the  $G_3\text{-p-}A_4\text{-p-}A_5$  moiety (**IIIf-h**, Fig. 4c) or with  $G_1\text{-p-}C_2\text{-p-}G_3$  (**IIIi-j**, Fig. 4d).<sup>3</sup> In the first case, it can transfer the proton to either  $G_3\text{-p-}A_4$  (phosphate “ $p_{3,4}$ ”) or  $A_4\text{-p-}A_5$  (“ $p_{4,5}$ ”). At  $HN \sim 2$  (the largest value here), the proton is localized on the ammonium group. This contrasts with the interaction mode **P**, where the proton is delocalized. Thus, the interaction of the ammonium group with a second phosphate stabilizes the ionic state. Lowering  $HN$  slightly increases the probability of  $p_{3,4}$  to be protonated (Fig. 5 and Tab. 1;  $HN \sim 1.5$  for **IIg** and  $\sim 1$  for **IIIh**), while the protonation on  $p_{4,5}$  remains unlikely (see SI for further details). Thus, the decrease of solvent interactions with the complex only slightly destabilizes the ionic state. In the second case, the proton is localized only at the ammonium group even with  $HN \sim 1.5$ . The ionic state is, as described above, stabilized by water and the presence of an additional phosphate.

In conclusions, our simulations suggest that low hydration leads to full protonation of the oligo in interaction mode **P** and it causes an increase of probability to be protonated in **P2**.

In N-ESI-MS experiments of DNA, the evaporation of the droplets leads eventually to protonated oligomers and ammonia.<sup>3,10,14</sup> The molecular determinants of the process are not known. Here, we address this issue by multiscale simulations on the d(GpCpGpApApGpC)

---

<sup>2</sup> This discussion assumes that the influence of the different nucleobases on the free energy profile is negligible with respect to the influence of the  $HN$ .

<sup>3</sup> **IIIi-j** are formed during the QM/MM-MD equilibration phase (see SI).

heptanucleotide, for which experimental N-ESI-MS data are available, as well as the model DMP, in the presence of ammonium counterions. The simulations of the model systems (Fig. 2) clearly indicate the impact on hydration on the proton transfer free energy profile: the proton is transferred to DMP if the ammonium is not fully hydrated. MD simulations on the actual heptanucleotide show that the ammonium forms two different interactions with the biomolecule, involving either one (**P** in Fig. 4) or two phosphates (**P2**). Both underwent QM/MM US-based free energy calculations. For (**P**), where only one phosphate interacts with the ammonium, we find that as soon the ammonium hydration is sufficiently low the proton is transferred to reduce the total charge state of the oligonucleotide. For (**P2**), where two phosphates interact with the ammonium ion, the oligomer is still charged for poorly hydrated ammonium ions, however the probability of being protonated does increase with a decrease of the hydration of the cation. This suggests that the DNA oligomer is eventually protonated when fully dehydrated, consistent with experimental evidence.<sup>14,20</sup>

## COMPUTATIONAL METHODS

**Force fields.** The parmBSC1<sup>23</sup> and TIP3P<sup>24</sup> force fields were used for the biomolecules (**I** and **II**) and water molecules, respectively. Atomic partial charges for DMP and ammonium ions were derived to be compatible with the employed AMBER force field by a RESP fit at the HF/6-31G\* level of theory.<sup>25,26</sup> Bonded and van-der-Waals parameters for the ammonium ions were taken from the AMBER parm99 library.<sup>27</sup> The same force field was used for the MM part in the QM/MM calculations.

**MD in aqueous solution.** Model **I** consists of DMP and one ammonium ion in a (43.5 × 41.9 × 43.2) Å<sup>3</sup> water box, containing 5,646 atoms, while **II** consists of the heptamer and six ammonium

ions in a  $(66.7 \times 75.9 \times 79.1)$  Å<sup>3</sup> water box, containing 39,714 atoms. The initial structures were taken from the model with PDB code 1PQT<sup>21</sup> and solvated with the GROMACS solvate module.<sup>28</sup>

The electrostatic interactions were calculated with the smooth particle mesh Ewald algorithm using a real space cutoff of 10 Å.<sup>29,30</sup> The same cutoff was employed for the van-der-Waals interactions. The center of mass motion was removed every 100 steps. Bond distances involving covalently bound hydrogen atoms were constrained with the LINCS algorithm.<sup>31</sup> Periodic boundary conditions and a time step of 1 fs was used throughout. The systems first underwent 5,000 steps of steepest descent minimization. Then, they were heated up to 300 K with 5ns-long simulations at constant volume using the velocity rescaling thermostat algorithm with a coupling constant  $\tau$  of 0.1 ps.<sup>32</sup> Next, the systems' density was equilibrated for 5 ns by performing NPT simulations employing the same thermostat algorithm as in the previous step and the Berendsen barostat algorithm (coupling constant of 1 ps) to achieve a pressure of 1 bar.<sup>33</sup> Finally, production simulations were carried out within the NPT ensemble at 300 K and 1 bar. For this purpose, the Nosé-Hoover thermostat<sup>34,35</sup> and Parrinello-Rahman barostat<sup>36</sup> were employed with coupling constants of 0.5 ps and 1.0 ps, respectively. We collected 10 and 500 ns for the **(I)** DMP and **(II)** DNA systems in the water solution, respectively. The following quantities were determined as a function of time: root-mean-square deviation, radius of gyration, and hydrogen bond interactions of the DNA oligo, the distance between NH<sub>4</sub><sup>+</sup> ions and phosphate groups for the ion pair characterization and radial distribution functions for the NH<sub>4</sub><sup>+</sup> ions and phosphate groups with respect to water oxygen atoms (see SI for details).

**MD and QM/MM in the gas phase.** The initial structures were taken from the last snapshots of the MD simulations. For **I** (**II**), we selected three droplets of approximately 800 (3,680) water molecules within a radius of 15 Å (30 Å) around the center of mass of the biomolecules. They also

turned out to contain 1 and 3 ammonium ions, and to bear a charge of 0 and  $-3$ , respectively for **I** and **II**. Electrostatic interactions were treated by direct Coulomb summation, whereby no cutoff is applied. The same holds for the Lennard-Jones interactions. Both the center of mass translational and rotational velocity was removed every 100 steps. The target temperature was controlled by a velocity rescaling algorithm ( $\tau = 0.1$  ps).<sup>32</sup> The systems underwent 50 and 150 ns of NVT simulations at 300 K, following Ref. (19,37), re-assigning the velocities at the beginning of each batch according to the Maxwell-Boltzmann distribution and removing the water molecules located at 60 Å or more from the center of mass of the biomolecules at the end of each simulation. The last sticky water molecules were then evaporated by three consecutive 25ns-long simulations at 350, 400 and 450 K. The same quantities were monitored as above, along with the hydration number (*HN*) of the ammonium and of the phosphate moieties (see SI for details).

The initial QM/MM configurations were taken from the gas phase MD simulations, in which ion pairs between  $\text{NH}_4^+$  and DNA were observed (see SI for the definition of ion pairs). The systems were inserted in a large box of  $(200 \times 200 \times 200)$  Å<sup>3</sup>. In **I**, the  $\text{NH}_4^+$  and DMP ions were treated at the QM level, while the water molecules were described at the MM level. In **II**, the ammonium ion and the DNA backbone atoms, which were involved in the studied proton transfer, were enclosed in the QM region (28–41 atoms, Fig. S15). The covalent C4'–C5' and glycosidic bonds were cut and the dangling bonds saturated with hydrogen link atoms. The QM problem was solved within the density functional theory (see below). Born-Oppenheimer MD and umbrella sampling (US)<sup>38</sup> was carried out using a time step of 0.5 fs. We used the WHAM analysis to calculate the free energy from the US calculations.<sup>39</sup>

For the simulations of **I**, a planewave basis set up to a cutoff of 100 Ry was used,<sup>40</sup> with Troullier-Martins pseudopotentials describing the valence shell-core electron interactions.<sup>41</sup> The PBE

exchange-correlation functional was employed.<sup>42</sup> Periodic images were decoupled from the unit cell with the Martyna-Tuckerman solver.<sup>43</sup> The system was heated up with a rate of  $0.12 \text{ K}\cdot\text{fs}^{-1}$  to 300 K in 2.42 ps using the Berendsen thermostat (coupling strength of 5,000 a.u.)<sup>33</sup> and then equilibrated for other 2.42 ps using the Nosé-Hoover thermostat (coupling frequency of  $3,500 \text{ cm}^{-1}$ ).<sup>34,35</sup> The collective variable (CV) for the US calculations was chosen as the difference distance of the breaking/forming  $\text{N}_{\text{Amm}}-\text{H}$  and  $\text{H}-\text{O}_{\text{DMP}}$  bonds (Chart S1). Initial configurations were generated by performing a 2.66 ps-long simulation, scanning the CV with a moving restraint. Then, 11 equidistant windows of 14.5 ps in the interval  $[-1.0; 1.0] \text{ \AA}$  were simulated. A harmonic restraint with a force constant  $k = 40 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$  was applied. The first 1.1 ps are discarded from the analysis.

For the simulations of **II**, the quantum problem was solved using the mixed Gaussian-planewave DFT approach.<sup>44,45</sup> We employed the PBE-D3(BJ)<sup>42,46,47</sup> functional with the DZVP-MOLOPT basis set<sup>48</sup> and GTH pseudopotentials.<sup>49</sup> Four grids were used for the planewave expansion. A density cutoff of 500 Ry was used for the finest grid, and a relative cutoff was set to 80 Ry to specify the coarser grids.<sup>45</sup> The QM region was electrostatically coupled to the MM potential within the Gaussian expansion of the electrostatic potential approach.<sup>50,51</sup> After a short minimization, we heated the system to 300 K within 2.5 ps using the velocity rescale algorithm ( $\tau = 0.1 \text{ ps}$ ).<sup>32</sup> Next, we run an equilibration QM/MM MD at 300 K for 2.5 ps. We finally employed umbrella sampling to predict the free energy for the proton transfer using the same CV as for **I**. 10 equidistant windows of 30.0 ps in the interval  $[-1.0; 0.8] \text{ \AA}$  were used ( $k = 50 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$ ). The first 5 ps were discarded from the analysis.

The classical MD simulations and the MM simulations in the QM/MM calculations were performed using the GROMACS program package.<sup>28</sup> The QM calculations were performed either by CP2K<sup>45</sup> (for **II**, using the API QM/MM interface of GROMACS) or by CPMD<sup>52</sup> (for **I**, using the MiMiC QM/MM interface).<sup>53-55</sup> The PLUMED plugin was used to introduce the biases in the umbrella sampling QM/MM simulations.<sup>56,57</sup>

## ASSOCIATED CONTENT

### Supporting Information

Details on the calculated properties; MD analysis of the solution structure of model **II**; additional analysis of the gas phase MD simulations of both models; details on QM/MM-MD re-equilibration; time evolution of all simulated free energy profiles (PDF).

## AUTHOR INFORMATION

### Corresponding Authors

Paolo Carloni – *Computational Biomedicine (IAS-5/INM-9), Forschungszentrum Jülich GmbH, , 52428 Jülich, Germany; Department of Physics, RWTH Aachen University, Otto-Blumenthal-Straße, 52062 Aachen, Germany; Email: [p.carloni@fz-juelich.de](mailto:p.carloni@fz-juelich.de)*

Modesto Orozco – *Institute for Research in Biomedicine (IRB) Barcelona, The Barcelona Institute of Science and Technology, 08028 Barcelona, Spain; Department of Biochemistry and Biomedicine, University of Barcelona, 08028, Spain; Email: [modesto.orozco@irbbarcelona.org](mailto:modesto.orozco@irbbarcelona.org)*

### Notes

The authors declare no competing financial interests.

## ACKNOWLEDGMENT

This work was supported by the BioExcel CoE ([www.bioexcel.eu](http://www.bioexcel.eu)), a project funded by the European Union Contract H2020-INFRAEDI-02-2018-823830.

## REFERENCES

- (1) Baker, E. S.; Dupuis, N. F.; Bowers, M. T. DNA Hairpin, Pseudoknot, and Cruciform Stability in a Solvent-Free Environment. *J Phys Chem B* **2009**, *113* (6), 1722–1727. <https://doi.org/10.1021/jp807529m>.
- (2) Clemmer, D. E.; Jarrold, M. F. Ion Mobility Measurements and Their Applications to Clusters and Biomolecules. *Journal of Mass Spectrometry* **1997**, *32* (6), 577–592. [https://doi.org/10.1002/\(SICI\)1096-9888\(199706\)32:6<577::AID-JMS530>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1096-9888(199706)32:6<577::AID-JMS530>3.0.CO;2-4).
- (3) Largy, E.; König, A.; Ghosh, A.; Ghosh, D.; Benabou, S.; Rosu, F.; Gabelica, V. Mass Spectrometry of Nucleic Acid Noncovalent Complexes. *Chem Rev* **2022**, *122* (8), 7720–7839. <https://doi.org/10.1021/acs.chemrev.1c00386>.
- (4) van Dyck, J. F.; Burns, J. R.; le Huray, K. I. P.; Konijnenberg, A.; Howorka, S.; Sobott, F. Sizing up DNA Nanostructure Assembly with Native Mass Spectrometry and Ion Mobility. *Nat Commun* **2022**, *13* (1), 3610. <https://doi.org/10.1038/s41467-022-31029-5>.
- (5) Barth, M.; Schmidt, C. Native Mass Spectrometry—A Valuable Tool in Structural Biology. *Journal of Mass Spectrometry* **2020**, *55* (10). <https://doi.org/10.1002/jms.4578>.
- (6) Chance, M. R.; Farquhar, E. R.; Yang, S.; Lodowski, D. T.; Kiselar, J. Protein Footprinting: Auxiliary Engine to Power the Structural Biology Revolution. *J Mol Biol* **2020**, *432* (9), 2973–2984. <https://doi.org/10.1016/j.jmb.2020.02.011>.
- (7) Leney, A. C.; Heck, A. J. R. Native Mass Spectrometry: What Is in the Name? *J Am Soc Mass Spectrom* **2017**, *28* (1), 5–13. <https://doi.org/10.1007/s13361-016-1545-3>.
- (8) Tamara, S.; den Boer, M. A.; Heck, A. J. R. High-Resolution Native Mass Spectrometry. *Chem Rev* **2022**, *122* (8), 7269–7326. <https://doi.org/10.1021/acs.chemrev.1c00212>.
- (9) Gaskell, S. J. Electrospray: Principles and Practice. *Journal of Mass Spectrometry* **1997**, *32* (7), 677–688. [https://doi.org/10.1002/\(SICI\)1096-9888\(199707\)32:7<677::AID-JMS536>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1096-9888(199707)32:7<677::AID-JMS536>3.0.CO;2-G).
- (10) Grégoire, G. Transferring Nucleic Acids to the Gas Phase. In *Nucleic Acids in the Gas Phase*; Gabelica, V., Ed.; Springer Berlin, Heidelberg, 2014; pp 21–54. [https://doi.org/10.1007/978-3-642-54842-0\\_2](https://doi.org/10.1007/978-3-642-54842-0_2).
- (11) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal Chem* **2013**, *85* (1), 2–9. <https://doi.org/10.1021/ac302789c>.

(12) Hunter, E. P. L.; Lias, S. G. Evaluated Gas Phase Basicities and Proton Affinities of Molecules: An Update. *J Phys Chem Ref Data* **1998**, *27* (3), 413–656. <https://doi.org/10.1063/1.556018>.

(13) Lum, R. C.; Grabowski, J. J. Trimethyl Phosphate: The Intrinsic Reactivity of Carbon versus Phosphorus Sites with Anionic Nucleophiles. *J Am Chem Soc* **1992**, *114* (22), 8619–8627. <https://doi.org/10.1021/ja00048a041>.

(14) Stults, J. T.; Marsters, J. C.; Carr, S. A. Improved Electrospray Ionization of Synthetic Oligodeoxynucleotides. *Rapid Communications in Mass Spectrometry* **1991**, *5* (8), 359–363. <https://doi.org/10.1002/rcm.1290050806>.

(15) Arcella, A.; Portella, G.; Orozco, M. Structure of Nucleic Acids in the Gas Phase. In *Nucleic Acids in the Gas Phase*; Gabelica, V., Ed.; Springer Berlin, Heidelberg, 2014; pp 55–75. [https://doi.org/10.1007/978-3-642-54842-0\\_3](https://doi.org/10.1007/978-3-642-54842-0_3).

(16) Rueda, M.; Kalko, S. G.; Luque, F. J.; Orozco, M. The Structure and Dynamics of DNA in the Gas Phase. *J Am Chem Soc* **2003**, *125* (26), 8007–8014. <https://doi.org/10.1021/ja0300564>.

(17) Arcella, A.; Portella, G.; Ruiz, M. L.; Eritja, R.; Vilaseca, M.; Gabelica, V.; Orozco, M. Structure of Triplex DNA in the Gas Phase. *J Am Chem Soc* **2012**, *134* (15), 6596–6606. <https://doi.org/10.1021/ja209786t>.

(18) Rueda, M.; Luque, F. J.; Orozco, M. G-Quadruplexes Can Maintain Their Structure in the Gas Phase. *J Am Chem Soc* **2006**, *128* (11), 3608–3619. <https://doi.org/10.1021/ja055936s>.

(19) Porrini, M.; Rosu, F.; Rabin, C.; Darré, L.; Gómez, H.; Orozco, M.; Gabelica, V. Compaction of Duplex Nucleic Acids upon Native Electrospray Mass Spectrometry. *ACS Cent Sci* **2017**, *3* (5), 454–461. <https://doi.org/10.1021/acscentsci.7b00084>.

(20) Arcella, A.; Dreyer, J.; Ippoliti, E.; Ivani, I.; Portella, G.; Gabelica, V.; Carloni, P.; Orozco, M. Structure and Dynamics of Oligonucleotides in the Gas Phase. *Angewandte Chemie International Edition* **2015**, *54* (2), 467–471. <https://doi.org/10.1002/anie.201406910>.

(21) Padrtá, P.; Štefl, R.; Králík, L.; Žídek, L.; Sklenář, V. Refinement of d(GCGAAGC) Hairpin Structure Using One- and Two-Bond Residual Dipolar Couplings. *J Biomol NMR* **2002**, *24* (1), 1–14. <https://doi.org/10.1023/A:1020632900961>.

(22) Portella, G.; Orozco, M. Multiple Routes to Characterize the Folding of a Small DNA Hairpin. *Angewandte Chemie International Edition* **2010**, *49* (42), 7673–7676. <https://doi.org/10.1002/anie.201003816>.

(23) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J. L.; González, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M. Parmbsc1: A Refined Force Field for DNA Simulations. *Nat Methods* **2016**, *13* (1), 55–58. <https://doi.org/10.1038/nmeth.3658>.

(24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **1983**, *79* (2), 926–935. <https://doi.org/10.1063/1.445869>.

(25) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J Phys Chem* **1993**, *97* (40), 10269–10280. <https://doi.org/10.1021/j100142a004>.

(26) Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F.-Y. R.E.D. Server: A Web Service for Deriving RESP and ESP Charges and Building Force Field Libraries for New Molecules and Molecular Fragments. *Nucleic Acids Res* **2011**, *39* (suppl\_2), W511–W517. <https://doi.org/10.1093/nar/gkr288>.

(27) Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J Comput Chem* **2000**, *21* (12), 1049–1074. [https://doi.org/10.1002/1096-987X\(200009\)21:12<1049::AID-JCC3>3.0.CO;2-F](https://doi.org/10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F).

(28) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.

(29) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An  $N \cdot \log(N)$  Method for Ewald Sums in Large Systems. *J Chem Phys* **1993**, *98* (12), 10089–10092. <https://doi.org/10.1063/1.464397>.

(30) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J Chem Phys* **1995**, *103* (19), 8577–8593. <https://doi.org/10.1063/1.470117>.

(31) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J Comput Chem* **1997**, *18* (12), 1463–1472. [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H).

(32) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J Chem Phys* **2007**, *126* (1), 014101. <https://doi.org/10.1063/1.2408420>.

(33) 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* **1984**, *81* (8), 3684–3690. <https://doi.org/10.1063/1.448118>.

(34) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. *Mol Phys* **1984**, *52* (2), 255–268. <https://doi.org/10.1080/00268978400101201>.

(35) Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys Rev A (Coll Park)* **1985**, *31* (3), 1695–1697. <https://doi.org/10.1103/PhysRevA.31.1695>.

(36) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J Appl Phys* **1981**, *52* (12), 7182–7190. <https://doi.org/10.1063/1.328693>.

(37) Metwally, H.; McAllister, R. G.; Konermann, L. Exploring the Mechanism of Salt-Induced Signal Suppression in Protein Electrospray Mass Spectrometry Using Experiments and Molecular Dynamics Simulations. *Anal Chem* **2015**, *87* (4), 2434–2442. <https://doi.org/10.1021/ac5044016>.

(38) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J Comput Phys* **1977**, *23* (2), 187–199. [https://doi.org/10.1016/0021-9991\(77\)90121-8](https://doi.org/10.1016/0021-9991(77)90121-8).

(39) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J Comput Chem* **1992**, *13* (8), 1011–1021. <https://doi.org/10.1002/jcc.540130812>.

(40) Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics*; Cambridge University Press, 2009. <https://doi.org/10.1017/CBO9780511609633>.

(41) Troullier, N.; Martins, J. L. Efficient Pseudopotentials for Plane-Wave Calculations. *Phys Rev B* **1991**, *43* (3), 1993–2006. <https://doi.org/10.1103/PhysRevB.43.1993>.

(42) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys Rev Lett* **1996**, *77* (18), 3865–3868. <https://doi.org/10.1103/PhysRevLett.77.3865>.

(43) Martyna, G. J.; Tuckerman, M. E. A Reciprocal Space Based Method for Treating Long Range Interactions in *Ab Initio* and Force-Field-Based Calculations in Clusters. *J Chem Phys* **1999**, *110* (6), 2810–2821. <https://doi.org/10.1063/1.477923>.

(44) Lippert, G.; Parrinello, M.; Hutter, J. A Hybrid Gaussian and Plane Wave Density Functional Scheme. *Mol Phys* **1997**, *92* (3), 477–488. <https://doi.org/10.1080/002689797170220>.

(45) Kühne, T. D.; Iannuzzi, M.; del Ben, M.; Rybkin, V. v.; Seewald, P.; Stein, F.; Laino, T.; Khalilullin, R. Z.; Schütt, O.; Schiffmann, F.; Golze, D.; Wilhelm, J.; Chulkov, S.; Bani-Hashemian, M. H.; Weber, V.; Borštník, U.; Taillefumier, M.; Jakobovits, A. S.; Lazzaro, A.; Pabst, H.; Müller, T.; Schade, R.; Guidon, M.; Andermatt, S.; Holmberg, N.; Schenter, G. K.; Hehn, A.; Bussy, A.; Belleflamme, F.; Tabacchi, G.; Glöß, A.; Lass, M.; Bethune, I.; Mundy, C. J.; Plessl, C.; Watkins, M.; VandeVondele, J.; Krack, M.; Hutter, J. CP2K: An Electronic Structure and Molecular Dynamics Software Package - Quickstep: Efficient and Accurate Electronic Structure Calculations. *J Chem Phys* **2020**, *152* (19), 194103. <https://doi.org/10.1063/5.0007045>.

(46) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *Journal of Chemical Physics* **2010**, *132* (15). <https://doi.org/10.1063/1.3382344>.

(47) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J Comput Chem* **2011**, *32* (7), 1456–1465. <https://doi.org/10.1002/jcc.21759>.

(48) VandeVondele, J.; Hutter, J. Gaussian Basis Sets for Accurate Calculations on Molecular Systems in Gas and Condensed Phases. *J Chem Phys* **2007**, *127* (11), 114105. <https://doi.org/10.1063/1.2770708>.

(49) Goedecker, S.; Teter, M.; Hutter, J. Separable Dual-Space Gaussian Pseudopotentials. *Phys Rev B* **1996**, *54* (3), 1703–1710. <https://doi.org/10.1103/PhysRevB.54.1703>.

(50) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. An Efficient Real Space Multigrid QM/MM Electrostatic Coupling. *J Chem Theory Comput* **2005**, *1* (6), 1176–1184. <https://doi.org/10.1021/ct050123f>.

(51) Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. An Efficient Linear-Scaling Electrostatic Coupling for Treating Periodic Boundary Conditions in QM/MM Simulations. *J Chem Theory Comput* **2006**, *2* (5), 1370–1378. <https://doi.org/10.1021/ct6001169>.

(52) CPMD Version 4.3, <http://www.cpmd.org/>, Copyright IBM Corp 1990–2019, Copyright MPI Für Festkörperforschung Stuttgart 1997–2001.

(53) Olsen, J. M. H.; Bolnykh, V.; Meloni, S.; Ippoliti, E.; Bircher, M. P.; Carloni, P.; Rothlisberger, U. MiMiC: A Novel Framework for Multiscale Modeling in Computational Chemistry. *J Chem Theory Comput* **2019**, *15* (6), 3810–3823. <https://doi.org/10.1021/acs.jctc.9b00093>.

(54) Bolnykh, V.; Olsen, J. M. H.; Meloni, S.; Bircher, M. P.; Ippoliti, E.; Carloni, P.; Rothlisberger, U. Extreme Scalability of DFT-Based QM/MM MD Simulations Using MiMiC. *J Chem Theory Comput* **2019**, *15* (10), 5601–5613. <https://doi.org/10.1021/acs.jctc.9b00424>.

(55) Bolnykh, V.; Rossetti, G.; Rothlisberger, U.; Carloni, P. Expanding the Boundaries of Ligand–Target Modeling by Exascale Calculations. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. Blackwell Publishing Inc. July 1, 2021. <https://doi.org/10.1002/wcms.1535>.

(56) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput Phys Commun* **2014**, *185* (2), 604–613. <https://doi.org/10.1016/j.cpc.2013.09.018>.

(57) The PLUMED consortium. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nat Methods* **2019**, *16* (8), 670–673. <https://doi.org/10.1038/s41592-019-0506-8>.