Molecular Mechanisms of Phosphoester Bond Formation in Water Using Tight-Binding Ab Initio Molecular Dynamics

Zakarya Benayad, Matthias Bova Saint-André, Guillaume Stirnemann

Molecular mechanisms of phosphoester bond formation in water using tight-binding ab-initio molecular dynamics

Zakarya Benayad, Matthias Saint-André, and Guillaume Stirnemann*

CNRS Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, PSL University, Université de Paris, 13 rue Pierre et Marie Curie, 75005, Paris, France

E-mail: stirnemann@ibpc.fr

Abstract

Phosphate groups are ubiquitous in biomolecules, and are usually incorporated through phosphoester bonds between alcohol groups and orthophosphate. The formation of this bond is exceptionally difficult, with associated barriers of 30–45 kcal/mol in the absence of catalysts. In abiotic conditions, polymerizing nucleic acids without enzymes remains very challenging and it is still a partly-unsolved problem that severely question the RNA-World hypothesis for the origins of life. Offering a solution to this problem would involve a detailed knowledge of the reaction energetics and mechanisms. Yet, these remain not fully understood at a molecular level, especially because of the very slow reaction rates that represent a significant challenge for the experiments. The number of involved reaction coordinates and the possible role of the solvent in assisting the reaction are challenging for computational studies. Here, we use extensive ab-initio molecular dynamics simulations using semi-empirical tight-binding methods and enhanced sampling to address these issues. We first show that the choice of the tight-binding method is greatly limited by the instability of the water liquid phase for most DFTB generations and parameter sets that are widely available. We then focus on a model reaction involving methanol and orthophosphate, for which the two protonation states (mono- and di-anionic) that are dominant around neutral pH are considered. We compare different proton coordinates that enable (or not) the participation of solvent water molecules. Our simulations suggest that in all cases, a dissociative associative mechanism, with an intermediate metaphosphate, is favored. The main difference between the two phosphate species is that reaction with the monoanion is assisted by the substrate while that with the dianion involves solvent water molecules. Our results are in agreement with early experimental measurements but the reaction barriers are underestimated in our framework. We believe that our approach provides an interesting perspective on how to sample the reaction phase space efficiently, but it calls for future studies using more accurate descriptions of chemical reactivity.

Introduction

Phosphate groups are ubiquitous in living systems. They are part of the backbone of nucleic acids, they are present in phospholipids that are the key constituants of cellular membranes, they are often involved in the regulation of protein function and activity by post-translational phosphorylation, and as polyphosphates, they allow to store and to redistribute free-energy in many biological processes. Their incorporation into the organic material is usually ob-

tained by forming a bond between an oxydized form of phosporus and an alcohol group, which is the phosphoester bond. From a kinetic point of view, these bonds are exceptionally stable with lifetimes potentially exceeding millions of years.[1] This allows to store free-energy in the form of phosphate anhydrides such as ATP, which are kinetically stable molecules, and thus to channel free-energy in the cell without destruction of the free-energy carrier. From a thermodynamic point of view, the hydrolysis of phosphoester bonds is usually exergonic — in particular, the hydrolysis of P–O–P bonds found in the ATP to ADP reaction releases vast amounts of free-energy. But since the reaction is kinetically not favored, the temporal regulation of the cell's functions, which critically depend on the formation and breaking of P-O-C and P-O-P bonds, requires very efficient catalysts. In fact, the corresponding enzymes (phosphatases, phosphodiesterases, kinases, etc.) have some of the largest catalyzing power known to date.[1,2]

A consequence of these very high barriers is that polymerizing nucleic acids without enzymes remains very challenging and still a partly-unsolved problem.[3,4] This observation is a major obstacle to the RNA World hypothesis,[5,6] which has gained support over the years, and which states that modern life would have evolved from RNA-based enzymatic systems preceding the modern protein world. Two essential steps involve the formation of a phosphoester bond in these systems: first, nucleoside phosphorylation to form the nucleotide after reaction between one alcohol group and a phosphate, and subsequently, the ligation between two nucleotides. These reactions are so slow and not thermodynamically favorable that efficient synthetic routes today employed to obtain short oligonucleotides involve the phosphoramidite (amides of phosphite) route[7] and a complex cascade of reactions that are not very plausible in the context of life's origins. Several strategies have been employed, sometimes synergistically, to accelerate phosphoester bond formation.[3,4] Chemical activation of the phosphate with good leaving groups, such as cyanogens or carbodiimides, that could have been present in abiotic conditions, is usually required[4,5,8] but is not sufficient.[8] Thermodynamic conditions and/or solvent effects have been found to improve the yield;[9–11] it is also acknowledged that the presence of metal ions, such as Mg^{2+} or Zn^{2+}, plays a crucial role.[3,5,11] The immobilization of activated monomers, e.g. with mineral surfaces[12] or by immobilized nucleic acid strand templates,[11,13] was found to favor the formation of RNA oligonucteotides.

The rationalization of this vast body of experimental results and the search for efficient yet plausible conditions is not easy, in particular because the energetics and the mechanism of phosphoester bond formation are not fully understood at a molecular level. Previous computational investigations on alkylphosphate monoesters,[14–22] which have exclusively focused on the reverse reaction (i.e., its hydrolysis), have led to often contradictory results, and a variety of mechanisms with their associated nomenclature — stepwise, concerted, associative, dissociative, solvent-assisted or not, etc. — have been evidenced depending on the level of quantum description, or on the inclusion of explicit water molecules. Experimental measurements, which were originally interpreted as an indication of a dissociative mechanism with a metaphosphate-like intermediate, were later questioned and could well indicate the presence of a loose pentacoordinated transition-state.[2,23–26]

The study of this reaction is challenging, because it involves at least two proton transfers (deprotonation of the nucleophile and protonation of the leaving group) in addition to the exchange of oxygen groups at the phosphorus centers. Previous computational studies have mostly focused on the determination of *energy* landscapes, often relaxing the proton coordinates for fixed phosphorus–oxygen distances; these calculations were performed either in the gas phase, or in the presence of an implicit solvent model. In such a scenario, a mechanism assisted by the water solvent, which can act as a proton donor/acceptor, is not allowed; other studies have evidenced that such a mechanistic pathway may be favorable, and that even if water molecules were not explicitly involved

in the mechanism, the inclusion of explicit water molecules had a significant impact on the energy barrier.[15,21] Finally, including entropic effects (and therefore obtaining a molecular picture in terms of a free-energy) can only be done a posteriori based on the energy calculations, solvent entropic effects being at best taken into account into the implicit solvation model.

These previous studies have brought decisive and unprecedented insights about the plausible mechanisms and the energetics of the reaction ; however, accurate quantum descriptions significantly limit the possibility of thermal sampling and to account for solvent effects, which does not allow for a direct estimation of the free-energy landscape of the reaction that is more relevant than the energy landscape at finite, non-zero temperatures. As compared to static quantum calculations in a solvent continuum, ab-initio Molecular Dynamics (aiMD) simulations account for solvent entropy and reorganization, and they can provide a full molecular picture of the reactions pathways. Given the complexity of the system, and of the possible reaction coordinates, propagating aiMD trajectories would prove itself very challenging and computationally intensive, even if performed at a low DFT level. Ideally, a mixed quantum/classical approach, where the reactive molecules would be treated at a quantum level, and the solvent classically, would allow to access longer timescales (and thus better sampling), but would not permit explicit proton transfers with the solvent.

Here, we adopt a different perspective where we favor the sampling over the accuracy of the quantum description. To that end, we choose to use extensive aiMD simulations using semi-empirical tight-binding methods, which a priori offer a good compromise between chemical accuracy and computational efficiency. However, the description of phosphorus based species is notoriously difficult,[19] and reactivity in the water liquid phase is not routinely studied. The current contribution addresses this challenge and poses the question whether this approach can provide valuable information about phosphoester bond formation (or hydrolysis). It is organized as follows. As we shall see in details, the choice of a TB method is greatly limited by the instability of the water liquid phase for most of its generations and parameter sets that are widely available, including a recent reparametrization. This effect had been occasionally seen before but it is not, to the best of our knowledge, widely reported.

We then focus on the phosphoester bond formation for a model reaction involving orthophosphate and methanol and compare different proton coordinates allowing the participation of solvent water molecules or not. Two protonation states for phosphate (mono- and dianionic) are considered, as these are the most dominant species around neutral pH. Overall, aiMD simulations performed at a SCC-DFTB level (with two different sets of parameters) and with a combination of metadynamics and umbrella-sampling propagated for several nanoseconds in each case, tend to show that a dissociative mechanism, with an intermediate metaphosphate, is favored. While this is in agreement with early experimental measurements, the reaction barriers are underestimated in our framework. Such a reaction pathway has already been evidenced for the dianion using DFT energy maps,[15,21] but the metaphosphate intermediate appears much more stabilized in the current approach. As a conclusion, we believe that our approach provides an interesting perspective on how to sample the reaction phase space efficiently, but our results call for future studies where higher-accuracy corrections could be performed on the reaction paths in order to compensate for possible deficiencies of the employed semi-empirical approaches.

Theoretical considerations and simulation strategy

The SCC-DFTB method and the employed atomic parameters

The central idea of the density functional tight-binding (DFTB) method is to write the Kohn-Sham density function theory (DFT) total energy as a sum of two-body interactions between localized electrons. In its self consistent-

charge (SCC) extension, atomic charges are determined in a self-consistent manner. Pairwise interactions are computed in advance and tabulated as a function of interatomic distances. Such a strategy considerably reduces the cost of the energy calculation at each simulation timestep, which is typically 10^2–10^3 faster as compared to a low-level (Generalized Gradient Approximation) DFT approach. We now detail the variants of SCC-DFTB that we have considered in this work. The reader is referred to recent reviews of this approach for more details,[27,28] which are only very briefly mentioned here.

The Taylor expansion of the Kohn-Sham energy is written as follows at the third order:

$$E_{total} = \hat{H}_{core} + E_{rep} + E^{2^{nd}} + E^{3^{rd}} \quad (1)$$

$\hat{H}_{core}$ and E_{rep} are respectively the core and repulsive energies, written as a sum of tabulated pairwise interactions. All SCC-DFTB variants contain at least the second-order in the Taylor expansion of the Kohn-Sham energy:

$$E^{2^{nd}} = \frac{1}{2}\sum_{\alpha,\beta} \Delta q_\alpha \gamma_{\alpha\beta}(r_{\alpha\beta}) \Delta q_\beta \quad (2)$$

where the sum runs over the atoms α and β. The charge fluctuations $\Delta q = q - q^0$ are determined self-consistently. In the original DFTB2 formulation,[29,30] $\gamma_{\alpha\beta}$ is written as

$$\gamma_{\alpha\beta}(r_{\alpha\beta}) = \frac{1}{r_{\alpha\beta}} - S_{\alpha\beta}(r_{\alpha\beta}) \quad (3)$$

where $r_{\alpha\beta}$ is the interatomic distance between α and β, ans $S_{\alpha\beta}$ an exponentially decaying function of $r_{\alpha\beta}$ to account for short-term corrections. In the following, "DFTB2" will refer to this reference SCC approach.

Subsequent work have proposed an improved description of hydrogen-bonding interactions[31] by modifying the expression of $\gamma_{\alpha\beta}(r_{\alpha\beta})$ for interactions involving hydrogen atoms (α or β = H):

$$\gamma_{\alpha\beta}(r_{\alpha\beta}) = \frac{1}{r_{\alpha\beta}} - f_{\alpha\beta}(r_{\alpha\beta}) S_{\alpha\beta}(r_{\alpha\beta}) \quad (4)$$

with

$$f_{\alpha\beta}(r_{\alpha\beta}) = \exp\Big[-\Big(\frac{U_\alpha + U_\beta}{2}\Big)^\zeta r_{\alpha\beta}^2 \Big] \quad (5)$$

where U_α and U_β are the atomic Hubbard parameters, and ζ is determined by fitting the results of higher-level calculations. In the following, we will use the notation "-γ" to refer to calculations using this correction.

Further improvement of the SCC-DFTB description was then proposed by including third-order terms.[31,32] The current implementation of DFTB in cp2k only allows for on-site third-order corrections that are written as follows:

$$E^{3^{rd}}_{on\ site} = \frac{1}{6}\sum_\alpha \frac{\partial \gamma_{\alpha\alpha}}{\partial q_\alpha}\Big|_{q^0_\alpha} \Delta q_\alpha^3 \quad (6)$$

which will be referred to as "DFTB3-diag" in the context of this work.

Finally, we also tested the use if dispersion-corrections for a better description of weak intermolecular interactions. Originally developed in the framework of DFT calculations, these corrections have been shown to improve the results of DFTB calculations as well.[33] As we used the state-of-the-art, third generation of such corrections,[34] we will be using the notation "-D3" to refer to simulations employing these corrections.

All the variants mentioned above require a set of parameters that are determined based on higher-level quantum calculations and by numerical fitting of these data. In most of this work, we used the state-of-the-art SCC-DFTB mio1-1 parameter set,[29] which includes specific reparametrization for phosphorus.[19,32] As shown by Cui and workers,[19] interactions involving phosphorus pose considerable challenges. For a selection of investigated reactions, the molecular geometries were shown to be in good agreement with higher level quantum descriptions, but with sometimes large energy deviations. The employed parameter set (mio1-1) readily include some updated parameters for phosphorus interactions in order to address these issues. Subsequent reparametrizations of another parameter set (3ob) showed further improvements,[35] but is only compatible

with the DFTB3 framework, which as discussed in the main text, does not allow for a reliable description of dense aqueous environments.

We also used another, specific reparametrization for interactions involving oxygen and hydrogen,[36] as discussed in the results section. Following previous work, the DFTB2-γ used a γ value of 4.50 taken from refs.[37,38] and DFTB3-γ a value of 4.53 as in ref.[32] .

Simulation details

All simulations presented here employed cp2k (version 7.1), a versatile, well-parallelized simulation code which allows for aiMD simulations at various levels of quantum descriptions.[39] It was also used for initial equilibration of the systems at the classical level. All simulations were performed using cubic boxes and periodic boundary conditions. We used a simulation timestep of 0.5 fs. Long-range electrostatics was treated using smooth particle mesh-Ewald (PME), with 25 grid points in each direction, a 5-th order beta-Euler spline, and a α parameter equal to 0.7. While α values on the order of $\approx$ 0.3 Å^{-1} are generally recommended for short-range cutoffs on the order of 10 Å, higher values are employed for TB methods in cp2k. We systematically increased the α value until convergence of the energy was observed, and found that a value of 0.7 Å^{-1} yielded converged results that did not evolve anymore when by further increase.

The self-consistent calculations employed the cp2k default values, except for the maximum number of cycles that was fixed to 20, and the tolerance threshold for a relative change in energy fixed to 10^{-7}.

All enhanced-sampling strategies (well-tempered metadynamics or umbrella sampling) as well as specific constraints were applied in the cp2k simulation using the Plumed library (version 2.3).[40] Simulations were performed in the canonical ensemble at a target temperature of 300 K.

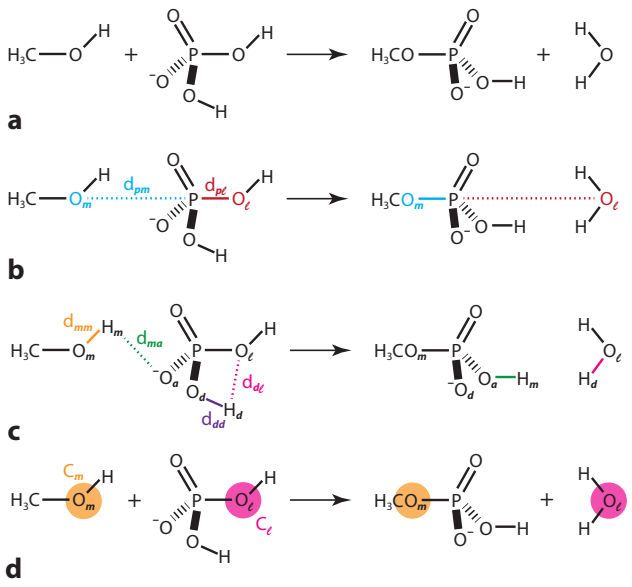

Figure 1: Reaction between methanol and $H_2PO_4^-$. (a) Scheme of the reaction ; (b) Definition of the distances used to define the distance-based oxygen CV 1 from $d(\mathrm{P}-\mathrm{O_l})$ (d_{pm}, blue) and $d(\mathrm{P}-\mathrm{O_m})$ (d_{pl}, red); (c) Definition of the distances used to define the distance-based hydrogen CV 2 from $d(\mathrm{O_m}-\mathrm{H_m})$ (d_{mm}, orange), $d(\mathrm{O_a}-\mathrm{H_m})$ (d_{ma}, green), $d(\mathrm{O_d}-\mathrm{H_d})$ (d_{dd}, purple), and $d(\mathrm{O_l}-\mathrm{H_d})$ (d_{dl}, pink); (d) Definition of the atoms involved in the hydrogen coordination-based CV 3 from $C(\mathrm{O_m})$ (C_m, orange) and $C(\mathrm{O_l})$ (C_l, pink).

Simulation workflow

Three different systems were investigated here: pure water, {methanol + $H_2PO_4^-$ + water}, and {methanol + HPO_4^{2-} + water}. In each case, 128 water molecules (and 1 molecule of each reactant) were randomly inserted in a 15.7 Å-side cubic box using Packmol.[41] A first classical simulation was performed using the SPC/E water forcefield,[42] and CgenFF forcefields for the reactants.[43] After minimization, the trajectory was propagated in the isothermal-isobaric ensemble for 2 ns. The first 200 ps were discarded and the remaining 1.8 ns was used to determine the average box volume. A subsequent 6-ns long trajectory was then propagated in the canonical ensemble at this volume. The last frame of this set of classical simulations was used as a starting configuration for the aiMD trajectory.

When investigating the different SCC-DFTB generations and corrections in the pure water box, a first aiMD equilibration was performed in the canonical ensemble for 10 ps and then propagated for 90 ps.

For the three reactive systems, a 50-ps equilibration phase was followed by a 2D well-tempered metadynamics (wt-metaD[44,45]) simulation for 1 ns, along 2 collective variables (CVs) that were defined using Plumed. In this work, 3 types of CVs were used. We first define them based on the {methanol + $H_2PO_4^-$ + water} system (Figure 1a). The first CV is a distance-based coordinate that is the difference of distances between the phosphorus atom and the nucleophile/leaving oxygens:

$$CV1 : \delta_{\mathrm{O}} = d(\mathrm{P} - \mathrm{O_l}) - d(\mathrm{P} - \mathrm{O_m}) \qquad (7)$$

where $\mathrm{O_l}$ is the (assigned) phosphate leaving oxygen and $\mathrm{O_m}$ the methanol oxygen (Figure 1b). The second CV is a distance-based coordinate that accounts for substrate-assisted proton transfer:

$$CV2 : \delta_{\mathrm{Hd}} = \delta\mathrm{H}_A + \delta\mathrm{H}_B \qquad (8)$$

with

$$\delta\mathrm{H}_A = d(\mathrm{O_m} - \mathrm{H_m}) - d(\mathrm{O_a} - \mathrm{H_m}) \qquad (9)$$

where $\mathrm{H_m}$ the methanol labile hydrogen and $\mathrm{O_{acc}}$ the phosphate proton acceptor oxygen (Figure 1c). It is assigned at the beginning of the simulation as being one of the non-protonated phosphate oxygens. We also introduce

$$\delta\mathrm{H}_B = d(\mathrm{O_d} - \mathrm{H_d}) - d(\mathrm{O_l} - \mathrm{H_d}) \qquad (10)$$

where $\mathrm{O_d}$ the phosphate proton donor oxygen and $\mathrm{H_d}$ its proton (Figure 1c). This group is assigned at the beginning of the simulation as being the non-leaving phosphate OH. Finally, a third CV is a coordination number-based coordinate that accounts for proton transfer (with any donor and acceptor):

$$CV3 : \delta_{\mathrm{Hc}} = C(\mathrm{O_l}) - C(\mathrm{O_m}) \qquad (11)$$

where C(...) is the coordination number of the oxygens in hydrogens (Figure 1d) using Plumed default switching function with a 1.3 Å cutoff value.

Wt-metaD trajectories were propagated along CV1/CV2 (substrate-assisted proton transfer) or CV1/CV3 (substrate- or solvent-assisted proton transfer), using a Gaussian height of 5 kcal/mol every 50 timesteps, of width σ = 0.07 Å along CV1/CV2 and σ = 0.07 for CV3, and, a (well-tempered) bias factor of 80 and a temperature of 300 K. For HPO_4^{2-}, we only performed wt-metaD simulations along CV1/CV3, which do not require any modification in their definition. When specifically investigating the differences between the dissociative or associative-dissociative mechanisms, we imposed the order in which the two proton transfers occurred by adding constraints on the sign of $\delta\mathrm{H}_B - \delta\mathrm{H}_A$ (see Supplementary information). Additional constraints (mostly upper limits for a number of other collective variables) were used to ensure that the energy injected into the system would not lead to spurious reactions. For each system, a detailed list can be found in the Supplementary Information (SI).

Data of the wt-metaD trajectory was then used to determine a first guess of the reaction path along the two CVs using the string method[46] using 30 to 40 nodes (see below). We then randomly selected, along the wt-medaD trajectory, molecular configurations close each of the string nodes, that were subsequently used as starting points of umbrella sampling[47] (US) trajectories. Each of these trajectories was equilibrated for 10 ps (data was discarded from subsequent analyses) and then propagated for 10 ps that were used for the final PMF reconstruction. A first set of US windows was generated using weak force constants of 10 kcal/mol/Å^2 for the harmonic potentials centered on the node positions. Non-sampled areas along the two CVs were filled by creating additional windows and higher force constants (up to 5000 kcal/mol/Å^2). We carefully assessed the overlap between adjacent windows in terms of CV1/2/3, as well as along each of their components, to check, for example, that adjacent windows would not suddenly differ in

one of these components. Details about all the corresponding windows are given in the Supplementary Information for each system and each combination of CVs.

Data analysis

The molecular dynamics trajectories were visualized using VMD[48] (version 1.9.4), which was also employed to compute the radial distribution functions in pure water. The two-dimensional potentials of mean force (PMFs) were estimated from the history of Gaussian insertion in wt-metaD simulations by using Plumed. From that point, we used an in-house implementation of the string-method[46] in Python, which aims to find the minimum free energy path (MFEP) between two basins of a free energy surface. It consists in letting evolve an initial string of nodes to the most probable transition path, by displacing each node along the free energy gradient, and enforcing constraints and reparametrization in order to prevent entanglement of the string. The two-dimensional potentials of mean force extracted from the US windows were determined using 2D-WHAM.[49] Structural data was computed from wt-metaD or US trajectories using in-house analyses codes in Python, using in part the MDAnalysis tools.[50,51] For representation purposes and the discussion of the reaction energetics and mechanisms along one dimension reporting on the progress of the reaction, we defined the following generalized CV for our studies of the substrate-assisted mechanisms:

$$\Delta_{\mathrm{O.H}} = \frac{1}{2}\big(\frac{1}{2}\delta_{\mathrm{O}} + \frac{1}{4}\delta_{\mathrm{Hd}}\big) \qquad (12)$$

whereas we used the following definition when using the coordination-based CV3 for proton transfer:

$$\Delta_{\mathrm{O.H}} = \frac{1}{2}\big(\frac{1}{2}\delta_{\mathrm{O}} + \frac{1\text{Å}}{2}(\delta_{\mathrm{Hc}} - 1)\big) \qquad (13)$$

The coefficients were chosen such that the oxygen- and proton-transfer coordinate contribute to the global CV with approximately equal weights, with $\Delta \approx -1$ Å and $\Delta \approx +1$ Å in the reactant and product states, respectively.

Results and discussion

Ability to reproduce a stable water bulk

We first consider the possibility to simulate a stable water bulk phase using DFTB. As detailed in the Methods section, the DFTB simulation framework has seen many improvements along the years, either by including higher-order terms in the energy function,[31,32] corrections for a better description of hydrogen-bonds,[31] and improved parameter sets. One can also include improved descriptions of the dispersion interactions,[34] as commonly applied for aiMD DFT simulations. Our initial goal was to base our study on a state-of-the-art simulation framework, using the latest developments and parameters sets, especially because phosphorus chemistry is notoriously challenging in DFTB and major improvements have been made.[19,32] A first limitation pertains to the employed aiMD code (cp2k), which was at the time where the simulations presented here much faster than DFTB+,[52] and which also easily allows to later switch to aiMD simulations at at the DFT level. The nanosecond-timescale required by our enhanced sampling simulations, as detailed later, required an efficient and well-parallelized simulation code. The only setback is that cp2k currently does not offer the implementation of the off-diagonal DFTB3, as well as that of the latest generations of the xTB approach[53] (which we have also tested). As we incidentally realized, a second major issue is that using the most recent strategy that is currently available in cp2k resulted in very unstable simulation boxes. Large voids would appear after a few tens of picoseconds, and then remain present throughout the simulation (Figure 2). We then tested various combinations where we varied the DFTB generation, the inclusion of HB corrections, and that of dispersion interactions; in all cases but one, we observe such bubbles.

As such an observation is very surprising, a possibility could have been that this was due to an issue in the DFTB implementation in cp2k. But this is unlikely to be the case. First, we compared single-point calculations on a water

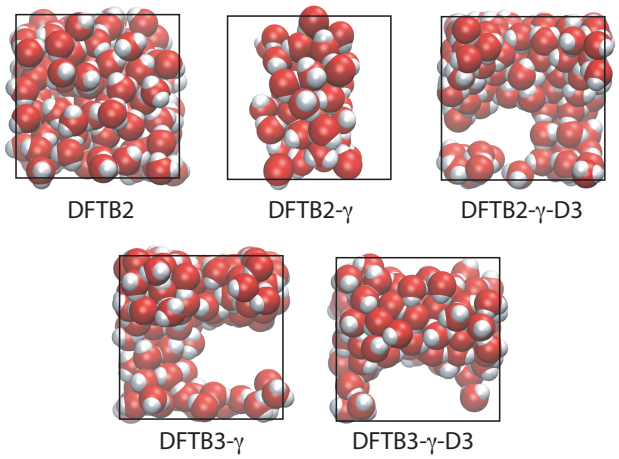

Figure 2: Water void formation in DFTB simulations. Snapshots of the pure water system obtained after 100 ps of simulations using various DFTB versions (see main text for the employed parameters). The black square indicates the periodic box. Only the simple SCC-DFTB (DFTB2) without hydrogen damping nor dispersion corrections seems to lead to a stable liquid water phase.

dimer using DFTB+ and cp2k, and found no difference in the calculated energies. Second, a similar problem has been reported by Voth and co-workers,[38] who used DFTB+. Water voids were seen in all systems, but a visual example ass only given for a specific DFTB3 scheme.[38] A major difference with our work is that we did not observe voids when employing the most simple strategy, namely SCC-DFTB with no corrections for HB or dispersion. As noted in Voth's work,[38] these voids are most likely a consequence of the overcoordination of DFTB water (Figure 3). In that respect, SCC-DFTB is less overcoordinated than the other versions of DFTB, with a height of the first peak close to that of the experimental radial distribution function (Figure 3), but with clear and important deficiencies in describing the position of the first peak, the first minimum and then the position of the second hydration shell (which are not really improved in further versions of DFTB). Tests performed in the NPT ensemble resulted in a worrying increase in density, reaching sometimes values close to 2 g.cm^{-3}. As an intermediate conclusion, we thus note that if one sticks to the the official DFTB/cp2k/parameter set release, only the SCC-DFTB with the canonical mio1-1 parameter sets is able to give a relatively stable water phase — albeit with a structure not well described beyond the first solvation shell.

A reparametrization of the parameter sets for SCC-DFTB have been carried out in order to better reproduce the repulsive part of the SCC-DFTB potential.[36] These parameters were used in particular for the liquid water phase, and allowed for a better description of the water structure.[36] However, as this reparametrization was made such that it reproduces DFT water structure better, it presents the same flaws as DFT water, i.e., the peak intensities are too strong, although they are located at the correct distances (Figure 3). A simple classical forcefield such as SPC/E largely outperforms the DFTB results in that respect (Figure 3). However, the corresponding water phase seems stable, with no voids observed in the simulation box. The Voth group also tested another parameter set (3OB) with DFTB3-diag; while the water structure was slightly improved as compared to DFTB3-diag with mio1-1, voids well still observed as a result of overcoordination.[38] We also tested the xTB approach, whose first generation GFN1-xTB[53] is implemented in cp2k; it also resulted in similar voids.

In the following, we will thus employ the SCC-DFTB setup together with the mio1-1 parameter set, which allows for stable simulations of a condensed phase of water and can therefore allow the study of reactivity in aqueous solutions. None of the other generations of DFTB tested here and before would be adequate for such a task, because the formation of bubbles could locally result in the formation of interfaces that would strongly affect the reaction. We will compare our results with that obtained with the reparametrization of the O-O, O-H and H-H interactions[36] mixed with the regular parameters of mio1-1 for the other pair interactions. However, such a hybrid combination should be taken with care and we will only use it to discuss whether it improves the results of the regular SCC-DFTB with mio1-1 as compared to the experimental and other computational studies.

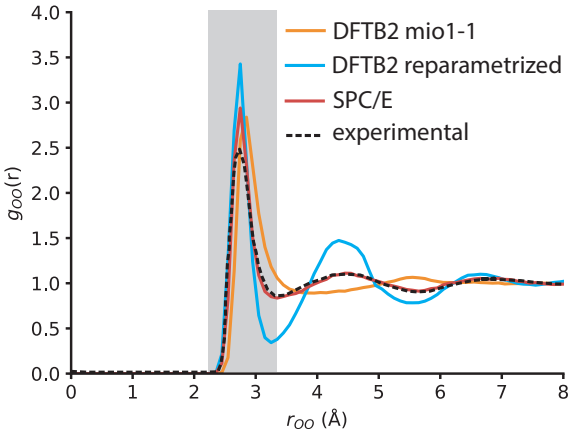

Figure 3: Oxygen-oxygen radial distribution functions in bulk water. Comparison between experimental data (dashed black line), classical SPC/E (red), DFTB2 with the mio1-1 parameter set (orange), and with the reparametrized set of O and H parameters (blue). The grey zone corresponds to the location of the first hydration shell; the mio1-1 parameter set clearly leads to a significant overcoordination in that region.

Phosphate monoanion reactivity with substrate-assisted proton transfer

We begin our investigation by considering mechanisms that do not involve proton transfers with the solvent. In that case, we proceed as follows: we specifically define as a leaving group one -OH group (out of 2) of the diprotonnated phosphate (Figure 1b); the second -OH group on the phosphate can transfer its hydrogen to the leaving oxygen; and a third phosphate oxygen can accept the proton of the methanol nucleophile (Figure 1c). In a framework that is as simple as possible, this leaves us with 6 reaction coordinates: the distance between the leaving oxygen and phosphorus, $d(\mathrm{P}-\mathrm{O_l})$; the distance between the methanol oxygen and phosphorus, $d(\mathrm{P}-\mathrm{O_m})$; the distance between the methanol proton and oxygen, $d(\mathrm{O_m}-\mathrm{H_m})$; the distance between the methanol proton and the phosphate oxygen acceptor (which is pre-defined), $d(\mathrm{O_a}-\mathrm{H_m})$; the distance between the second phosphate proton and its original oxygen, $d(\mathrm{O_d}-\mathrm{H_d})$; and the distance between the leaving oxygen and the second phosphate proton, $d(\mathrm{O_l}-\mathrm{H_d})$. Note that this imposes the phosphate oxygens that would accept or donate their hydrogens; the fact that the phosphate has another oxygen that could do so would only bring a very minor contribution to the actual barrier for the proton transfer, on the order of $k_BT\ln(2)$, which is negligible as compared to the involved barriers, as we shall see below. In this approach, we also eliminate the pathway along which the $\mathrm{H_m}$ would be the only traded hydrogen. At best, this also brings a very minor contribution to the free-energy barrier, that could be much higher anyway because it imposes a specific order for proton transfer as well as some reorientation of the OH group.[21]

An additional reduction of dimensionality is required as the free-energy landscape exploration cannot be effectively done in 6 dimensions. We thus further group the two O−P distances into one coordinate (their difference δ_{O}, Equation 7) that can discriminate between the reactant, the transition state(s) and/or intermediate(s), and the product states. Similarly, we define a proton transfer coordinate δ_{Hd} (Equation 8), bearing in mind that a detailed analysis of the involved O−H and O−P distances would then be required to characterize the molecular geometries along the reaction path(s). For example, $\delta_{\mathrm{O}} = 0$ Å can well be degenerate and correspond to several distinct states (such as, a metaphosphate with large $d(\mathrm{P}-\mathrm{O_l})$ and $d(\mathrm{P}-\mathrm{O_m})$ distances, or, a pentacoordinated intermediate with shorter phosphorus–oxygen bonds).

We propagated a first metaD trajectory for 1 ns by sampling along these two coordinates. A detailed investigation of the time-course of $d(\mathrm{P}-\mathrm{O_m})$, $d(\mathrm{P}-\mathrm{O_l})$ and of the protonation states, reveals that the reaction indeed occurred several times on this time scale, with several transitions that were observed between the reactant and the product. However, two types of intermediate species are observed and can be clearly characterized by e.g. the distributions of $d(\mathrm{P}-\mathrm{O_m})$ and $d(\mathrm{P}-\mathrm{O_l})$ close to the reaction midpoint ($\delta_O \approx 0$ Å). First, a metaphosphate-

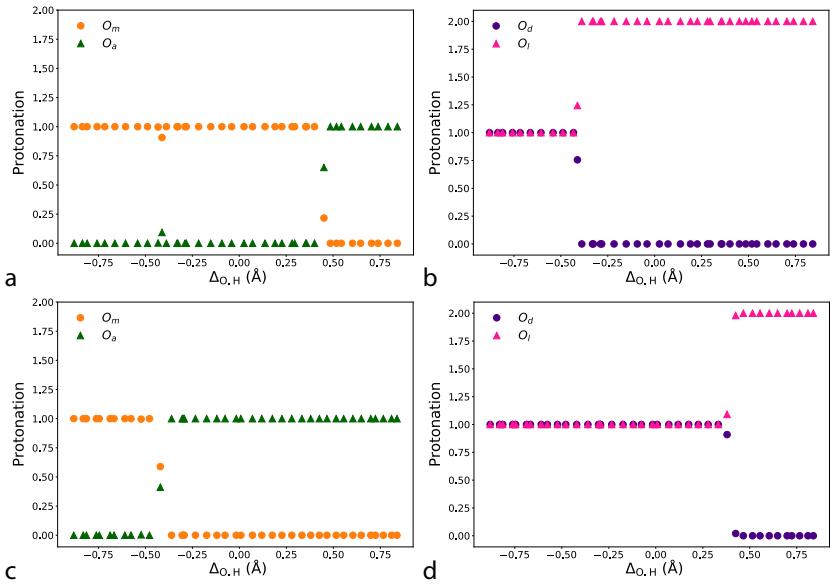

Figure 4: Protonation state and proton transfers among the reactive {methanol + $H_2PO_4^-$} system with substrate-assisted transfer. Each point corresponds to the average number of protons for a small region in the CV1/CV2 space centered on the position of each string node, determined from the US 2D-PMF (a) and (b) $D_N + A_N$ mechanism, where the transfer between the leaving group and phosphate (b) occurs before that between the phosphate and the nucleophile (a). (c) and (d) A_ND_N mechanism, where the transfer between the phosphate and the nucleophile (c) occurs before that between the leaving group and phosphate (d).

like species, with loose $d(P-O_m)$ and $d(P-O_l)$ distances, where the proton transfer between the leaving group and the phosphate has already taken place, but the methanol oxygen is still protonated; and a pentacoordinated species, with short $d(P-O_m)$ and $d(P-O_l)$, with a deprotonnated methanol but where the leaving group is still a hydroxyl group. For the standard mio1-1 parameter sets, the metaphosphate intermediate is largely present (97% of molecular geometries correspond to the two distances being larger than 2.0 Å). When using the reparametrized oxygen/hydrogen interactions, this fraction falls to 66%. These observations point toward the co-existence of two distinct mechanisms, that differ from the geometries of their intermediate species and by the order of the proton transfers: a associative-dissociative mechanism (A_ND_N) where methanol is deprotonated first and the leaving hydroxyl group in a second step; a dissociative mechanism (S_N1, or $D_N + A_N$), with a metaphosphate intermediate and where the methanol oxygen is deprotonated last. Note that we never observed concerted deprotonation and protonation of the nucleophile and leaving groups, respectively.

These two mechanisms have been observed and discussed before, based on the experimental measurements[2] and computational approaches;[14–21] a consensus picture that seems to emerge is that they are very close in energy.[21] However, an exploration of the free-energy landscape along two coordinates for which different types of mechanisms are degenerate, does not allow for a direct determination of the free-energy barriers corresponding to each of this reaction pathways. We therefore repeated our wt-metaD exploration by imposing the order in which the proton transfers occurred, which would then determine the reaction pathway, as now discussed. The order in which the proton transfers occurred is controlled by adding an additional constraint on the O−H distances (see Methods). In each case, a long wt-metaD simulation was performed along the same two coordinates as before. From this simulation, a first minimum free-energy path (MFEP) was determined by

optimizing the position of a discretized version of the path (a string[46]) on the dimensional landscape representing the system 2D PMF along δ_O and δ_{Hd}. Molecular geometries that correspond to each point of the string were then identified and served as starting points of subsequent umbrella sampling windows, following a protocol described in the Methods section. A converged 2D PMF was then reconstructed, and a reaction path was optimized again on this surface. Finally, the molecular configurations corresponding to each point of the string were identified, which provides the molecular details of the two reaction pathways, as now described.

Before we examine the (US) free-energy landscapes, we first discuss the characteristics of each reaction mechanism. For convenience, we represent most of the data in the following as a function of a reduced reaction coordinate that is a weighted linear combination of the protons and oxygen coordinates, $\Delta_{O.H}$ (Equation 12). We first note that proton transfers occur as expected and they are never concerted: for the D_N + A_N mechanism, the transfer of the methanol proton occurs after that of the phosphate (Figure 4a and b), and the reverse is observed for the second A_ND_N simulation (Figure 4c and d). The evolution of the O−P distances (Figure 5) reveals that in the second scenario, protonation of the leaving group occurs first (Figure 4b) and it is concerted with a sudden increase in $d(P - O_l)$, leading to a metaphosphate intermediate with O-P distances that is larger than 3 Å (Figure 5a). The two O-P distances reach a symmetric situation while they have very large values (2.5 Å). Eventually, $d(P - O_m)$ suddenly drops while the methanol proton is transferred to the phosphorus (Figure 4a). When the metaphosphate intermediate is present, we observe large fluctuations of the distances around their maximum probability values. The second scenario indeed corresponds to an associative-dissociative mechanism (Figure 5b): after deprotonation of the methanol when $d(P - O_m)$ is about 2.9 Å (Figure 4c), the attacking oxygen keeps approaching the phosphorus and eventually replaces the leaving hydroxyle in a very graduate manner (Figure 5a). For intermediate values of the reaction coordinates, the presence of a tight pentacoordinate phosphate (with both O−P distances below 1.8 Å) is observed. Protonation of the leaving group eventually occurs when the O-P distance is about 2.8 Å (Figure 4d).

The 2D PMFs reconstructed by the US simulations are shown in Figure 6. They reveal that the two concerted proton/oxygen transfers occurring during the dissociative mechanism are associated with noticeable barriers of about 15 and 13 kcal/mol respectively (Figure 6a). Here, we approximate the free-energy barriers with the barriers seen in the 2D potentials of mean force, although these are in principle slightly different. However, the absolute values as well as the relative differences between the different mechanisms are large compared to the small errors made by such approximations, that are expected to be on the order of 1 kcal/mol at most.[54]

The metaphosphate species appears to be a stable reaction intermediate within the DFTB framework. The associative mechanism is significantly higher in free-energy (Figure 6b). The pentacoordinated species corresponds to a flat region of the free-energy surface, but does not appear to be a very stable intermediate and it would rather correspond to a first transition state species. The proton transfer to the leaving group corresponds to a second barrier of 13 kcal/mol, with a second transition state corresponding to the second proton transfer. These relative differences between the two mechanisms probably explain why the dominant mechanism in the metadynamics exploration was the dissociative mechanism.

In our approach so far, proton transfers are taken into account via differences of distances, which requires to know the proton donors and acceptors. They are therefore limited to intra-reactant proton transfers, and the involvement of solvent water molecules cannot be taken into account this way. We now switch to a sampling framework allowing for proton transfers with the solvent.

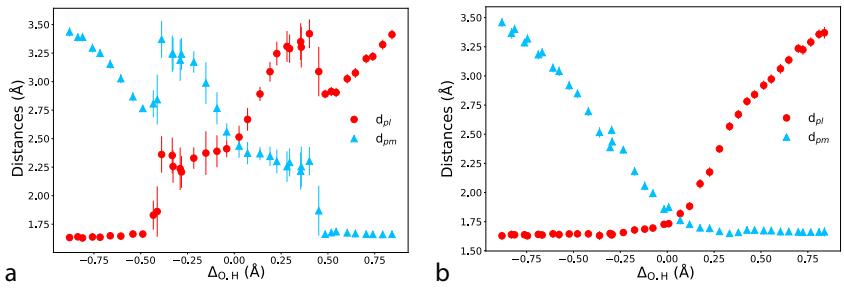

Figure 5: Distances between the attacking/leaving oxygens and phosphorus ($d(\mathrm{P}-\mathrm{O_m})$ (d_{pm}, blue) and $d(\mathrm{P}-\mathrm{O_l})$ (d_{pl}, red)) in the reactive {methanol + $H_2PO_4^-$} system with substrate-assisted transfer. Each point corresponds to the maximum in the distance distribution for a small region in the CV1/CV2 space centered on the position of each string node, determined from the US surface. The bars indicate the standard deviation around this maximum. (a) $\mathrm{D_N}$ + $\mathrm{A_N}$ mechanism and (b) $\mathrm{A_N D_N}$ mechanism.

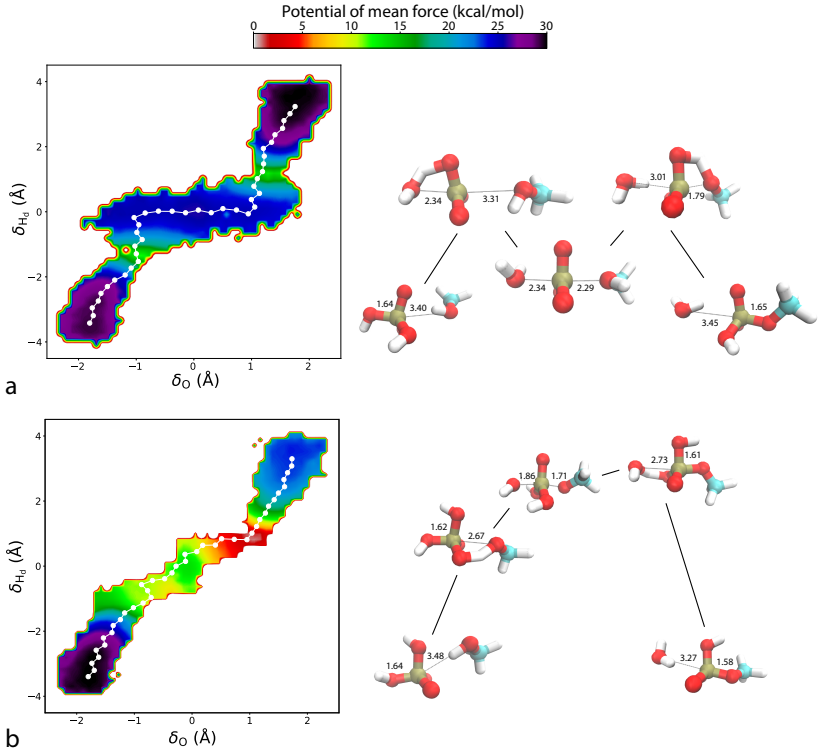

Figure 6: 2D potentials of mean force (PMF) along the two CV1 (δ_{O}) and CV2 (δ_{Hd}) for the reactive {methanol + $H_2PO_4^-$} system with substrate-assisted transfer. (a) $\mathrm{D_N}$ + $\mathrm{A_N}$ and (b) $\mathrm{A_N D_N}$ mechanisms. Note that the relative diffrerences between the reactant and product states is slightly overestimated in the $\mathrm{A_N D_N}$ simulation, and probably arises from errors accumulated along the PMF reconstruction. The points on the 2D landscape correspond to the MFEP string (see Methods). On the right hand side, we indicate for each scenario, a schematic reaction path with typical structures along the path corresponding to the proton and oxygen transfers, with the numbers indicating the O–P distances in these structures.

Phosphate monoanion reactivity with possibility of proton transfer with the solvent

Distance-based coordinates are not adequate in that case, as the acceptor of the methanol proton and the donor for the leaving group proton can now not only be a phosphate oxygen, but also the oxygen of any water molecule in

the system. Therefore, we turn to an alternate collective variable based on the proton coordination number of the involved oxygens (Figure 1d). In the reactant state, both oxygens are protonated; in the product, the leaving group is a water molecule and the nucleophile oxygen has no more hydrogen. It therefore seems natural to consider the difference in proton coordination numbers between these two groups, which would replace the distance-based proton coordinate in the two-dimensional sampling.

We then repeated our simulations following the same workflow as before: a first, long exploration using 2d-metadynamics; a check that the reaction was indeed sampled successfully multiple times; the identification of a putative reaction pathways; and specific exploration of the free-energy landscape around the initial pathway using 2D-US. Our initial exploration revealed that the reaction occurred exclusively through a dissociative mechanism (Figure 7a-b), and we therefore did not treat separately a dissociative and associative-dissociative mechanism. The 2D PMFs then obtained with US strongly ressembles the one obtained previously with distance-based proton coordinates (Figure 7d): the presence of a reaction intermediate and two barriers. Although reduced by a few kcal/mol, these barriers are consistent with those obtained with substrate-assisted proton transfer. The evolution of the O−P distances as a function of a reduced reaction coordinate (Figure 7a) reveals that these two barriers are associated with the departure of the leaving group, and then the binding of the nucleophile. As shown in Figure 7, these oxygen transfers are concerted with first, the protonation of the leaving group, and second, the deprotonation of methanol (Figure 7b).

The question now is whether solvent molecules are involved in these transfers. Our simulations unambiguously show that this is not the case; the total number of protons on the {methanol+phosphate} system is always equal to 3 along the reaction path (Figure 7c). In the US windows, no hydroxyde or hydronium ions are observed. These results suggest that within the SCC-DFTB framework, solvent molecules do not actively participate to the reaction. We can finally compare the energetics of the reaction to that obtained with the distance-based coordinate for proton transfer. While the observed pathway is similar in both cases, the barrier heights are smaller when using a coordination-number based coordinates (Figure 7d), and are reduced by 30-50% as compared to the results of Figure 6a. Despite these variations, the observation of a metaphosphate intermediate and the fact that the reaction mechanisms are identical with both set of coordinates suggests a certain robustness of our results with respect to the description of the proton transfer.

Phosphate dianion reactivity

Since the pKa of orthophosphate is equal to 7.2, at neutral pHs the phosphate monoanion coexists with its dianionic form, which only carries one proton, HPO_4^{2-}. Most of the previous computational studies have in fact focused on the hydrolysis of this form. Experimentally, the hydrolysis of HPO_4^{2-} is much slower as compared to that of $H_2PO_4^-$, mainly because of electrostatic repulsion between the nucleophile and the phosphate group. We would thus expect the involved free-energy barrier(s) to be larger in that case. We applied a similar strategy to the reaction between methanol and HPO_4^{2-}, using the δ_C and δ_O coordinates in order to allow the transfer of protons to the solvent water molecules (Figure 8).

The obtained PMFs evidence the presence of a reaction intermediate similar to that obtained with the monoanion; and two barriers of respective heights of 5 and 6 kcal/mol in the condensation direction, and 10 and 2 kcal/mol for the hydrolysis (Figure 8d). A major difference with the reaction intermediate of $H_2PO_4^-$ is that the corresponding free-energy local minimum is much more shallow. A detailed analysis of the involved O-P distances (Fig. 8a) and of the protonation states of the reactants (Figure 8b) and of the solvent water molecules neighbors (Figure 8c) reveals that this intermediate corresponds to a protonated leaving OH group, a deprotonated metaphosphate, a deprotonated water molecule, and methanol. However, as op-

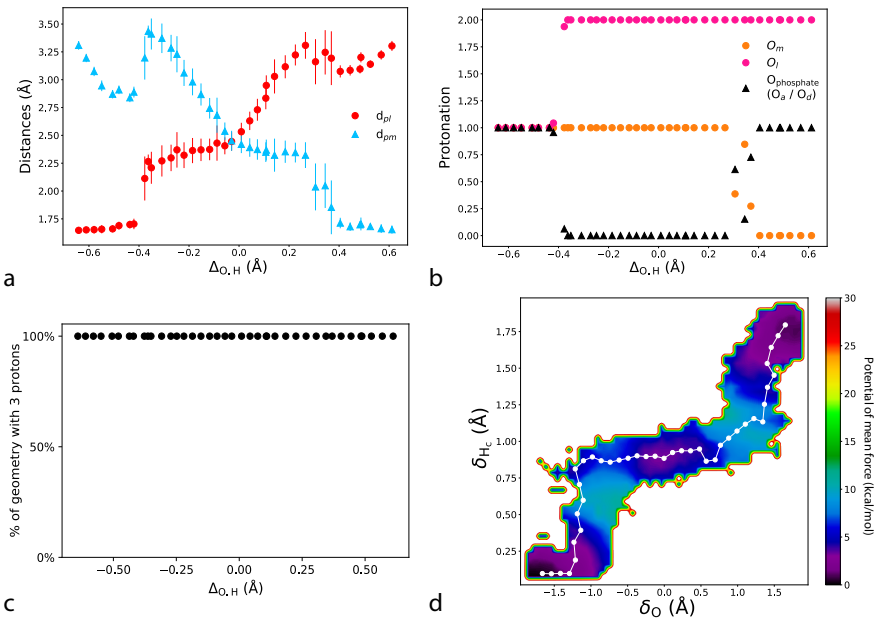

Figure 7: Reactive {methanol + $H_2PO_4^-$} system with possibility of solvent-assisted transfer. No order is imposed on the proton transfer; in that case, the reaction proceeds exclusively via a D_N + A_N mechanism. (a) Distances between the attacking/leaving oxygens and phosphorus ($d(P - O_m)$ (d_{pm}, blue) and $d(P - O_l)$ (d_{pl}, red)). Each point corresponds to the average distance for a small region in the CV1/CV2 space centered on the position of each string node, determined from the US 2D PMFs. The bars indicate the standard deviation around this average. (b) Protonation state for the oxygens involved in the transfer between methanol and phosphate, and the leaving group and phosphate. The symmetry between the respective protonation states suggest that the solvent molecules do not assist the transfers, as confirmed by (c) the total number of protons in the reactive system, excluding water molecules. (d) 2D PMFs along the two CV1 (δ_O) and CV3 (δ_{Hc}), and MFEP string. As no solvent molecule is involved in the proton transfers, the reaction proceeds via the same intermediate species as in Fig. 6.

posed to the phosphate monoanion, an asymmetry exists between the molecular events corresponding to these two barriers. During the reaction, the protonation of the leaving -OH group is assisted by a solvent water molecule and is almost barrierless once the distance between the departing oxygen and phosphorus starts to increase. It leads to the departure of the H_2O leaving group and then to a shallow metaphosphate intermediate. The second barrier corresponds to the simultaneous proton transfer between methanol and a hydroxyde anion, and the formation of the bond between the methanol oxygen and phosphorus.

A major difference with the reaction with the phosphate monoanion is thus that the formation of the phosphoester bond with the phosphate dianion is assisted by the water solvent. The phosphate oxygens others than the leaving group are never protonated during the reaction, and a water molecule is assisting the two proton transfers. Note that the water molecule involved in the first transfer is not necessarily the same as that involved in the second transfer, as clearly observed in the metadynamics trajectory.

We finally explored the same reaction but imposing a substrate-assisted proton transfer. Because only one proton is traded in that case (the phosphate non-bridging oxygens being deprotonated), the reaction has to proceed through a pathway that involves, first, deprotonation of methanol and protonation of a phosphate non-bridging oxygen, and second, protonation of the leaving oxygen atom. Such a scenario imposes a A_ND_N reaction mechanism, as we observed in

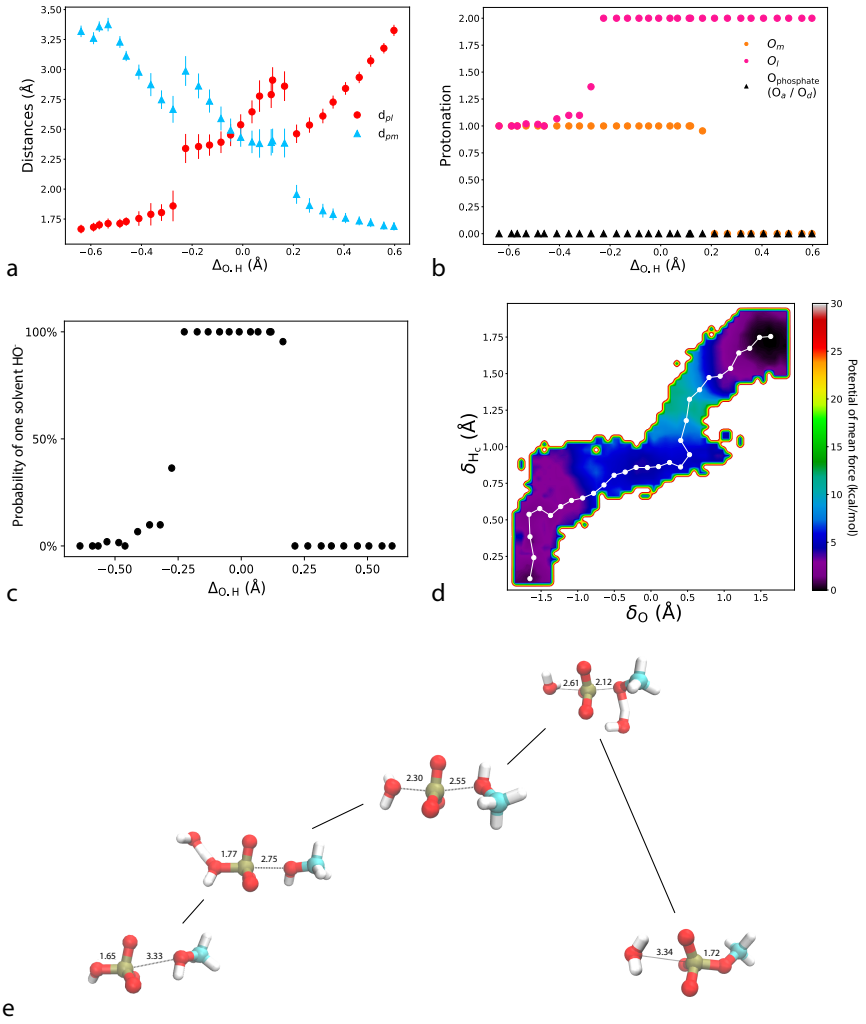

Figure 8: Reactive {methanol + HPO_4^{2-}} system with possibility of solvent-assisted transfer. No order is imposed on the proton transfer; in that case, the reaction proceeds exclusively via a D_N + A_N mechanism. (a) Distances between the attacking/leaving oxygens and phosphorus ($d(P - O_m)$ (d_{pm}, blue) and $d(P - O_l)$ (d_{pl}, red)). Each point corresponds to the average distance for a small region in the CV1/CV2 space centered on the position of each string node, determined from the US 2D PMF. The bars indicate the standard deviation around this average. (b) Protonation state for the oxygens involved in the transfer between methanol and phosphate, and the leaving group and phosphate. The absence of protonation of the phosphate non-bridging oxygens suggests that the mechanism is assisted by the water solvent, as evidenced by (c) the presence of one solvent hydroxide ion in the medium. (d) 2D PMFs along the two CV1 (δ_O) and CV3 (δ_{Hc}), and MFEP string. (e) Schematic reaction path with typical structures along the path corresponding to the proton and oxygen transfers, with the numbers indicating the O–P distances in these structures. The solvent water molecule involved in the proton transfers with the leaving and attacking group is shown for the two structures in which these transfers occur, and is not shown otherwise.

our metadynamics simulations, with an associated (single) barrier of ≈ 30 kcal/mol.

Comparison with experimental results and previous theoretical work

We now compare the results of our simulation with existing experimental and computational

data. Experimentally, phosphoester bond formation between phosphate and methanol are slightly thermodynamically unfavored and thus endergonic (+0.7 kcal/mol for HPO_4^{2-} and +2.0 kcal/mol for $H_2PO_4^-$).[14,55] The rate constant of the reverse hydrolysis reaction is very slow:[1] 2.0 10^{-20} $M^{-1}.s^{-1}$ for HPO_4^{2-} and 2.4 10^{-10} $M^{-1}.s^{-1}$ for $H_2PO_4^-$, with associated free-energy barriers that were estimated to 44.3 and 30.6 kcal.mol^{-1}. There has been an intense debate in the literature about the nature of the reaction mechanism.[2] Following early investigations, many experimental results seemingly supported the existence of a stepwise, dissociative mechanism with a metaphosphate intermediate.[23–26] Such a mechanism could explain the enhanced reactivity of the phosphate monoanion as compared to the dianion,[56] the results obtained regarding the small activation entropy, as well as the kinetic isotope effect (KIE) and linear free-energy relationship (LFER) results showing that the bond between the phosphorus and the leaving group was considerably extended during the reaction.[23–26] But as clearly reviewed by Herschlag and coworkers,[2] these results do not necessarily demonstrate the existence of a dissociative mechanism, but instead, suggest that the reaction proceeds through a loose transition state, with elongated P−O bonds. In fact, stable metaphosphate intermediates have not been evidenced in the experimental studies.[57,58] However, experiments mostly rely on indirect trends extracted from thermodynamic or kinetic studies, as they can hardly characterize the molecular geometries of reaction intermediates and/or transition states, which are by essence either unstable or highly metastable. Computational studies therefore appear as an ideal complement in order to rationalize and to provide a molecular picture of experimental results.

These theoretical work have led to contradictory and diverse conclusions regarding the reaction mechanism, depending on the level of theory used, on the computational methodology, and on the inclusion or not of explicit solvent molecules.[14–21] There has thus been reports of associative[14,20,21] and dissociative[14–16,18] reaction pathways, with either stepwise[17] or concerted[14,20,21] mechanisms, assisted by the water solvent or not. A recent extensive study compared some of these possibilities for the hydrolysis of the methanol phosphoester dianion using single-point calculations at a high level of DFT description, including explicit water molecules in an implicit water solvent.[21] Two pathways were found that were close in free-energy: a concerted, solvent-assisted mechanism with a very loose transition state (corresponding to large P–O distances) ; and a step-wise associative-dissociative mechanism with substrate-assisted proton transfer, exhibiting much tighter transition states. A dissociative pathway, with a metastable metaphosphate intermediate, was also found to be favored in the aqueous phase by a number of other, independent studies on the methanol phosphoester mono- and dianions that used different levels of DFT theory, explicit water molecules and an implicit PCM solvent.[15,16]

This associative-dissociative pathway in the dianion was also studied using metadynamics in implicit solvent[20] (thus forbidding any possibility of transfer with the solvent), which demonstrated an associative-dissociative mechanism but in much more concerted way, without any metastable intermediate, and with a phosphorane-like transition state that is to contrast with the intermediates found in other studies.[15] While different approaches may thus agree on the overall geometries encountered along the reaction pathway, they need not lead to clear conclusions regarding the presence or the absence of even a shallow well, which would be the signature of a metastable intermediate. This illustrates the possible differences between single-point energy energy calculations by adiabatic mapping at 0 K, with chosen reaction coordinates that usually rely on the P–O distances and that usually neglect an explicit description of the barriers associated with the proton transfers (that could well be a slow coordinate), ad-hoc inclusion of entropic effects, and direct explorations of free-energy reaction pathways using enhanced sampling techniques. Using a similar level of DTFB description, Cui and coworkers reported results[19] regarding the first step of the dissociation of the monoanion of

the methanol monoester that illustrate the major differences between adiabatic mapping calculations and potential of mean force through enhanced sampling simulations, as well as the influence of explicit (but non reactive) classical water molecules.

Our present results on the phosphoester bond formation between phosphate and methanol can be summarized as follows. For the monoanion phosphate, two pathways were identified that differ in the order of the substrate-assisted proton transfers: one concerted, associative-dissociative mechanism with a phosphorane-like transition state; and a dissociative mechanism involving a metastable metaphosphate intermediate. The second pathway exhibits much lower free-energy barriers, on the order of 10-15 kcal/mol (depending on the chosen proton coordinate), as compared to 25 kcal/mol for the associative-dissociative one. For the dianion phosphate, proton transfers are assisted by solvent water molecules that first donate a proton to the leaving hydroxyle group and eventually deprotonate the attacking methanol oxygen. The resulting reaction pathway is once again dissociative, with a metaphosphate intermediate, although much more metastable as compared to the monoanion intermediate.

We first comment on the presence of a metaphosphate intermediate. As mentioned before, such an intermediate had been evidenced in many previous theoretical work,[15] and is compatible with existing experimental data.[23–26] However, it has never been isolated experimentally,[2,57,58] which is probably not surprising because it is nonetheless very high in energy. A major difference between our current results and those using different computational approaches and models lies in the relatively high stability of this intermediate within the SCC-DFTB framework. This observation is actually consistent with the results obtained by Cui and coworkers[19] , who used a QM (phosphate+methanol) / MM (solvent water) description and who determined potentials of mean-force for the first steps of the associative and dissociative reaction pathways, respectively. Notably, they found that in the former case, a free-energy basin corresponding to a very stable metaphosphate intermediate was seen to be even lower in energy than the reactant. They attributed this surprising stabilization to an overpolarization effect of the solute because of the classical treatment of the surrounding explicit water molecules, and not to a SCC-DFTB deficiency by itself. Our current results demonstrate that an explicit description of the surrounding water molecules at the same level of theory does not significantly change this picture where the metaphosphate appears to be overstabilized as compared to previous theoretical work using higher levels of quantum description.

A second issue regarding the SCC-DTB description is that the reaction free-energy barriers are underestimated and significantly differ from what would be expected from the very slow rates of reaction. For the monoanion, the calculated barriers are at least 15 kcal/mol smaller, and the difference is even more pronounced for the dianion. Because the reaction involves two proton transfers, a reliable description of O–H interactions is critical; yet, the water structure is only very poorly described in this framework, and not reliably reproduced at all for most of DFTB's generations. To check whether other, and probably better, parameters for all interactions involving O and H would provide a better report on the reaction barriers, we tested a recent reparametrization of the O–O, O–H, and H–H interactions that was seen to greatly enhance water description.[36] But this does not significantly change the barriers, and they even appear slightly smaller. We also tested whether including a systematic correction along the reactions paths in order to include the differences in energy between a chosen DFT description and SCC-DFTB (calculated in the gas phase) would impact the validity of our results. As discussed in the SI, although such a strategy is probably too simplistic, the key qualitative features of our results remain unchanged, and the dissociative-associative mechanism is favored in all cases. A more careful examination of the impact of the quantum description level would definitely deserve future attention.

One could then argue that the employed

methodology could lead to artifactual effects in the determination of the reaction free-energy surface. But several arguments reinforce the validity of our estimations, at least in the SCC-DFTB framework: first, we always find a good overlap between adjacent US windows, both in terms of the collective variables involved in the reaction coordinate(s) and each of their components. It is thus unlikely that we are missing some slow variables that would have totally different values in adjacent windows. Second, the PMFs, obtained by piece-wise reconstruction using the WHAM algorithm, suggest that the reactants and the products have almost the same free-energy, in all our systems, which is consistent with experimental data. If some artefacts would be introduced in the reconstruction, it is very likely that we would end up with totally shifted surfaces in favor of the products or of the reactants. Finally, the fact that our barriers and the presence of a stable metaphosphate intermediate are seen in different sets of simulations, involving different proton transfer descriptions and different protonation states for the phosphate, suggests that these conclusions seem pretty robust — at least, within the SCC-DFTB framework.

Conclusions

The formation of phosphoester bonds in abiotic conditions is a considerable challenge as it is associated with important free-energy barriers and thus very slow rates.[1] In this work, we make an attempt to shed light on its molecular mechanism of phosphoester bond formation between methanol and phosphate in bulk water using an approximate density functional theory framework and enhanced sampling techniques in ab-initio molecular dynamics simulations. From a computational perspective, difficulties arise from the number of collective variables that are required to take into account the changes occurring during the reaction (both in terms of protonation states and phosphorus-oxygen distances), and of the possible assistance of the proton transfers by the neighboring water solvent molecules. Most previous theoretical studies have employed higher level quantum descriptions,[14–18,20,21] often in implicit solvent and by taking into account entropic effects only a posteriori. By the determination of energy maps along the phosphorus-oxygen distances, the assumption is also often made that the proton coordinates are fast in comparison, which might not necessarily be the case. Here, we propose a computational strategy that allows to account for both oxygen and proton coordinates, in explicit and reactive water solvent, and that relies on extensive enhanced sampling the reaction pathways. Such an approach comes at the cost of the level of quantum description, because even low-level DFT would pose a significant computational challenge. Instead, we adopt the approximate DFT framework provided by DFTB, together with a combination of well-tempered metadynamics and umbrella sampling.

We first show that most generations of DFTB lead to an unstable liquid water phase, with the formation of water voids within a few picoseconds of simulations, in all cases but for the most simple SCC-DFTB (DFTB2) description. As compared to experimental data, water structure is still very approximately reproduced by the standard mio1-1 parameter sets, and could be improved by a recent reparametrization. While most of our work is based on the mio1-1 parameter set, we also propose some comparison with this alternative set of parameters.

We then successively focus on the two phosphate species that are predominant at neutral pH, the monoanion $H_2PO_4^-$ and the dianion HPO_4^{2-}. In the first case, we compare two types of proton coordinates: one based on oxygen-hydrogen distances, imposing a substrate-assisted mechanism, and another one based on coordination states, allowing also for solvent-assisted pathways. In all cases, we find that the reaction proceeds exclusively with proton transfers via the phosphate non-bridging oxygens. Two pathways were identified, one associative-dissociative A_ND_N concerted mechanism with a tight pentacoordinated transition state, and a stepwise dissociative-associative $D_N + A_N$ mechanism with a loose metaphosphate intermediate. We find the free-energy barriers for the later to be significantly lower

(10–15 kcal/mol) than that of the A_ND_N mechanism (20–25 kcal/mol). For the dianion HPO_4^{2-}, the most probable pathway is also a stepwise $D_N + A_N$, which now involves assistance by neighboring water molecule. Indeed, the leaving -OH group deprotonates a water molecule to give a hydroxide anion, that later deprotonates the nucleophile.

Our results are in qualitative agreement with previous experimental[23–26] and computational[15] results, even if the exact nature of the intermediate species along the reaction pathway and of their relative stability (transition states or intermediates) remain controversial. Our results suggest that there exists a metastable metaphosphate intermediate, which is compatible with the experimental results suggesting significant elongation of the P-O bond during the rate limiting step of the reaction[2,23–26] . However, as noted earlier,[19] this intermediate is likely to be overstabilized in the SCC-DFTB approach, and the barriers corresponding to its formation are probably underestimated accordingly (by 20-25 kcal/mol), which could explain the discrepancy between the barriers that we obtain and those estimated based on experimental results.

Our current effort thus faces some limitations probably inherent to the approximate quantum description, and the methodology presented therein could later be improved in several ways. First, assuming that the reaction pathway as well as the density of states in the free-energy landscape would little change if a higher level of theory would be used — which might be strong assumptions considering the description offered by the DFTB, energy corrections could be made on the free-energy surfaces. Second, the latest generation of extended tight-binding (xTB), which seems to offer a much higher accuracy as compared to other tight-binding approaches, could offer an interesting alternative. A third possibility that we are currently investigating is whether a reactive forcefield that would be deep-learned from high-level quantum calculations could lead to a reliable description of the reaction energies and forces while allowing for extensive sampling of the reaction pathways.

Supporting information

Simulation details, constraints used during the metadynamics and umbrella sampling runs, overlap along relevant collective variables among adjacent replicas, and comparison of the reaction energy paths with a higher level quantum method based on DFT.

Acknowledgement The research leading to these results has received funding from the European Research Council under the European Union's Eighth Framework Program (H2020/2014-2020)/ERC Grant Agreement No. 757111 (G.S.). This work was also supported by the "Initiative d'Excellence" program from the French State (Grant "DYNAMO", ANR-11-LABX-0011-01 to GS). ZB is funded by a PhD Fellowship from the CFM Research Foundation. The simulations presented here benefited from a local computing platform administered by G. Letessier, and was granted access to the HPC resources of TGCC under the allocation A0070811005 made by GENCI (Grand Equipement National de Calcul Intensif). We thank Élise Duboué-Dijon for useful discussions and her critical reading of the manuscript.

Data availability

The data that support the findings of this study are available from the corresponding author upon request.

Conflict of interests

The authors have no conflict of interests to declare.

References

(1) Schroeder, G. K.; Lad, C.; Wyman, P.; Williams, N. H.; Wolfenden, R. The time required for water attack at the phosphorus atom of simple phosphodiesters and of DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4052–4055.

(2) Lassila, J. K.; Zalatan, J. G.; Herschlag, D. Biological phosphoryl-Tranfer reactions: understanding mechanism and catalysis. *Annu. Rev. Biochem.* **2012**, *80*, 669–702.

(3) Szostak, J. W. The eightfold path to non-enzymatic RNA replication. *J. Sys. Chem.* **2012**, *3*, 2.

(4) Cafferty, B. J.; Hud, N. V. Abiotic synthesis of RNA in water: A common goal of prebiotic chemistry and bottom-up synthetic biology. *Curr. Opin. Chem. Biol.* **2014**, *22*, 146–157.

(5) Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99–123.

(6) Robertson, M. P.; Joyce, G. F. The origins of the RNA World. *Cold Spring Harbor Perspectives in Biology* **2012**, *4*, 1.

(7) Beaucage, S. L.; Caruthers, M. H. Deoxynucleoside phosphoramidites. A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **1981**, *22*, 1859–1862.

(8) Lohrmann, R.; Orgel, L. E. Prebiotic synthesis: phosphorylation in aqueous solution. *Science* **1968**, *161*, 64–66.

(9) Lohrmann, R.; Orgel, L. E. Urea-Inorganic Phosphate Mixtures as Prebiotic Phosphorylating Agents. *Science* **1971**, *171*, 490–494.

(10) Costanzo, G.; Saladino, R.; Crestini, C.; Ciciriello, F.; Di Mauro, E. Nucleoside phosphorylation by phosphate minerals. *J. Biol. Chem.* **2007**, *282*, 16729–16735.

(11) Monnard, P. A.; Szostak, J. W. Metal-ion catalyzed polymerization in the eutectic phase in water-ice: A possible approach to template-directed RNA polymerization. *J. Inorg. Chem.* **2008**, *102*, 1104–1111.

(12) Ferris, J. P. Mineral catalysis and prebiotic synthesis: montmorillonite-catalyzed formation of RNA. *Elements* **2005**, *1*, 145–150.

(13) Deck, C.; Jauker, M.; Richert, C. Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. *Nature Chem.* **2011**, *3*, 603–608.

(14) Florián, J.; Warshel, A. Phosphate ester hydrolysis in aqueous solution: Associative versus dissociative mechanisms. *J. Phys. Chem. B* **1998**, *102*, 719–734.

(15) Bianciotto, M.; Barthelat, J. C.; Vigroux, A. Reactivity of phosphate monoester monoanions in aqueous solution. 1. Quantum mechanical calculations support the existence of "anionic zwitterion" $MeO^{+}(H)PO_3^{2-}$ as a key intermediate in the dissociative hydrolysis of the methyl phosphate anion. *J. Am. Chem. Soc.* **2002**, *124*, 7573–7587.

(16) Wang, Y. N.; Topol, I. A.; Collins, J. R.; Burt, S. K. Theoretical studies on the hydrolysis of mono-phosphate and triphosphate in gas phase and aqueous solution. *J. Am. Chem. Soc.* **2003**, *125*, 13265–13273.

(17) Iché-Tarrat, N.; Ruiz-Lopez, M.; Barthelat, J. C.; Vigroux, A. Theoretical evaluation of the substrate-assisted catalysis mechanism for the hydrolysis of phosphate monoester dianions. *Chem. Eur. J.* **2007**, *13*, 3617–3629.

(18) Kamerlin, S. C. L.; Florián, J.; Warshel, A. Associative versus dissociative mechanisms of phosphate monoester hydrolysis: On the interpretation of activation entropies. *ChemPhysChem* **2008**, *9*, 1767–1773.

(19) Yang, Y.; Yu, H.; York, D.; Elstner, M.; Cui, Q. Description of phosphate hydrolysis reactions with the Self-Consistent-chage Density-Functional-Tight-Binding (SCC-DFTB) theory. 1. Parameterization. *J. Chem. Theo. Comput.* **2008**,

(20) Branduardi, D.; De Vivo, M.; Rega, N.; Barone, V.; Cavalli, A. Methyl phosphate dianion hydrolysis in solution characterized by path collective variables coupled with DFT-based enhanced sampling simulations. *J. Chem. Theo. Comput.* **2011**, *7*, 539–543.

(21) Duarte, F.; Åqvist, J.; Williams, N. H.; Kamerlin, S. C. L. Resolving apparent conflicts between theoretical and experimental models of phosphate monoester hydrolysis. *J. Am. Chem. Soc.* **2015**, *137*, 1081–1093.

(22) Martínez-Bachs, B.; Rimola, A. Prebiotic peptide bond formation through amino acid phosphorylation. Insights from quantum chemical simulations. *Life* **2019**, *9*, 75.

(23) Di Sabato, G.; Jencks, W. P. Mechanism and catalysis of reactions of acyl phosphates. II. Hydrolysis. *J. Am. Chem. Soc.* **1961**, *83*, 4400–4405.

(24) Di Sabato, G.; Jencks, W. P.; Whalley, E. The effect of pressure on the spontaneous hydrolysis of acetyl phosphate mono-anion and di-anion and of acetyl phenyl phosphate mono-anion. *Can. J. Chem.* **1962**, *40*, 1220.

(25) Kirby, A. J.; Varvoglis, A. G. The reactivity of phosphate esters. Monoester Hydrolysis. *J. Am. Chem. Soc.* **1967**, *89*, 415–423.

(26) Gorenstein, D. G.; Lee, Y.-g.; Kar, D. Kinetic isotope effects in the reactions of aryl-' 80-2,4-dinitrophenyl dibenzyl phosphate and aryl- 80-2,4-dinitrophenyl phosphate. Evidence for monomeric metaphosphate. *J. Am. Chem. Soc.* **1977**, *99*, 2264–2267.

(27) Seifert, G.; Joswig, J. O. Density-functional tight binding-an approximate density-functional theory method. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 456.

(28) Elstner, M.; Seifert, G. Density functional tight binding. *Phil. Trans. R. Soc. A* **2014**, *372*, 20120483.

(29) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**,

(30) Koskinen, P.; Mäkinen, V. Density-functional tight-binding for beginners. *Comput. Mat. Sci.* **2009**, *47*, 237.

(31) Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. Extension of the self-consistent-charge density-functional tight-binding method: Third-order expansion of the density functional theory total energy and introduction of a modified effective coulomb interaction. *J. Phys. Chem. A* **2007**,

(32) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J. Chem. Theo. Comput.* **2011**, *7*, 931–948.

(33) Brandenburg, J. G.; Grimme, S. Accurate modeling of organic molecular crystals by dispersion-corrected density functional tight binding (DFTB). *J. Phys. Chem. Lett.* **2014**,

(34) 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. *J. Chem. Phys.* **2010**, *132*, 154104.

(35) Gaus, M.; Lu, X.; Elstner, M.; Cui, Q. Parameterization of DFTB3/3OB for sulfur and phosphorus for chemical and biological applications. *Journal of Chemical Theory and Computation* **2014**, *10*, 1518–1537, PMID: 24803865.

(36) Doemer, M.; Liberatore, E.; Knaup, J. M.; Tavernelli, I.; Rothlisberger, U. In situ

parameterisation of SCC-DFTB repulsive potentials by iterative Boltzmann inversion. *Mol. Phys.* **2013**, *111*, 3595.

(37) Goyal, P.; Elstner, M.; Cui, Q. Application of the SCC-DFTB method to neutral and protonated water clusters and bulk water. *J. Phys. Chem. B* **2011**, *115*, 6790.

(38) Choi, T. H.; Liang, R.; Maupin, C. M.; Voth, G. A. Application of the SCC-DFTB method to hydroxide water clusters and aqueous hydroxide solutions. *J. Phys. Chem. B* **2013**, *117*, 5165–5179.

(39) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F. et al. CP2K: An electronic structure and molecular dynamics software package -Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020**, *152*, 194103.

(40) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604.

(41) Martinez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.

(42) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(43) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I. et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2009**, *31*, 671–690.

(44) Ensing, B.; De Vivo, M.; Liu, Z.; Moore, P.; Klein, M. L. Metadynamics as a tool for exploring free energy landscapes of chemical reactions. *Acc. Chem. Res.* **2006**, *39*, 73–81.

(45) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(46) E, W.; Ren, W.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B* **2002**, *66*, 052301.

(47) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.

(48) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33.

(49) Grossfeld, A. WHAM: the weighted histogram analysis method.

(50) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**,

(51) Gowers, R.; Linke, M.; Barnoud, J.; Reddy, T.; Melo, M.; Seyler, S.; Domański, J.; Dotson, D.; Buchoux, S.; Kenney, I. et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. Proceedings of the 15th Python in Science Conference. 2016.

(52) Hourahine, B.; Aradi, B.; Blum, V.; Bonafé, F.; Buccheri, A.; Camacho, C.; Cevallos, C.; Deshaye, M. Y.; Dumitric, T.; Dominguez, A. et al. DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *J. Chem. Phys.* **2020**, *152*, 124101.

(53) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.;

Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**, *11*, e1493.

(54) Dietschreit, J. C. B.; Diestler, D. J.; Hulm, A.; Ochsenfeld, C.; Gomez-Bombarelli, R. From Free-Energy Profiles to Activation Free Energies. *arXiv* **2022**, 10.48550/ARXIV.2206.02893.

(55) Guthrie, J. P. Hydration and dehydration of phosphoric acid derivatives: free energies of formation of the pentacoordinate Intermediates for phosphate ester hydrolysis and of monomeric metaphosphate. *J. Am. Chem. Soc.* **1977**, *99*, 3991–4001.

(56) Butcher, W. W.; Westheimer, F. H. The lanthanum hydroxide gel promoted hydrolysis of phosphate esters. *J. Am. Chem. Soc.* **1955**, *77*, 2420–2424.

(57) Buchwald, S. L.; Friedman, J. M.; Knowles, J. R. Phosphoric monoesters and a phosphoguanidine : The role of metaphosphate. *J. Am. Chem. Soc.* **1984**, *106*, 4911–4916.

(58) Herschlag, D.; Jencks, W. P. Evidence that metaphosphate nonoanion is not an intermediate in solvolysis reactions in aqueous solution. *J. Am. Chem. Soc.* **1989**, *111*, 7579–7586.

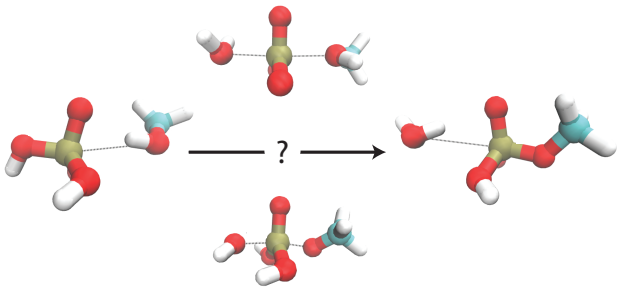

Figure 9: TOC Graphic