Nucleophilic substitution at di- and triphosphates: leaving group ability of phosphate versus diphosphate

Bas van Beek †, Marc A van Bochove †, Trevor A Hamlin †,‡ and F Matthias Bickelhaupt †,‡,★

1 Department of Theoretical Chemistry and Amsterdam Center for Multiscale Modeling (ACMM), Vrije Universiteit Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands
2 Amsterdam University of Applied Sciences, Faculty of Technology, Forensic Science Programme, Leeuwenbug, Weesperzijde 190, 1097 DZ Amsterdam, The Netherlands
3 Institute for Molecules and Materials (IMM), Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands
4 Author to whom any correspondence should be addressed. E-mail: t.a.hamlin@vu.nl and f.m.bickelhaupt@vu.nl

Keywords: DNA replication, di- and triphosphate hydrolysis, nucleophilic substitution, magnesium, activation strain model

Abstract
Adenosine triphosphate (ATP) is the universal energy carrier in biochemical processes. Herein, we aim for a better understanding of the origin of the high-energy content of the triphosphate moiety involved, the influence of various physicochemical factors thereon, and implication for the actual S<u><sub>N</sub></u><sup>2@@P</sup>-induced hydrolysis, which drives uphill biochemical processes, such as, DNA replication. To this end, we have investigated the S<u><sub>N</sub></u><sup>2@@P</sup>-induced hydrolysis of triphosphate (PPP) versus that of diphosphate (PP) using density functional theory (DFT) at COSMO(H<sub>2</sub>O)-ZORA-OLYP/TZ2P. We find that S<u><sub>N</sub></u><sup>2@@P</sup>-induced hydrolysis of PPP is favored over that of PP, both kinetically and thermodynamically. The energetic advantage of PPP over PP is slightly diminished by the coordination of Mg<sup>2+</sup> counterions. Our activation strain and energy decomposition analyses reveal that the activation barrier for PPP hydrolysis is lower compared to that for PP due to a weaker P<sup>α</sup>–O leaving group bond.

Introduction
The bimolecular nucleophilic substitution reaction (S<u><sub>N</sub></u><sup>2</sup>) at phosphorus is a ubiquitous reaction in biological systems [3, 4]. Its mechanism and the nature of the stationary points are strongly affected not only by the nucleophile and leaving group, but also by solvation and the substituents at the electrophilic phosphorus atom [5–9]. Hydrolysis of phosphate derivatives is an essential reaction central to the existence of life on Earth [10–19]. Phosphate hydrolysis constitutes one of the elementary reactions in numerous enzyme-catalyzed processes, including the phosphorylation of sugars [20–24], carboxylation of pyruvate [25], and replication of DNA [26–28]. For example, DNA replication involves a key S<u><sub>N</sub></u><sup>2@@P</sup> reaction that forms a new phosphodiester bond between the hydroxyl group of the primer strand and phosphate on the incoming nucleotide (scheme 1) [29]. Formation of the new phosphodiester bond simultaneously displaces a pyrophosphate leaving group. The hydrolysis of phosphates, both under enzymatic and non-enzymatic conditions, has been subject of many studies and reviews over the past decades by Westheimer, Herschlag and Warshel [30–32], and very recently by Kamerlin [33].

Suggested mechanisms for DNA replication [34–46], usually based on crystal structures, generally include a deprotonated phosphate chain with an initially protonated nucleophile (sugar moiety on the primer), which is deprotonated prior to the S<u><sub>N</sub></u><sup>2</sup> reaction (proton hopping). The S<u><sub>N</sub></u><sup>2</sup> reaction itself occurs via a pentacoordinate phosphorus species, likely a transition state (TS), with two magnesium counterions located in the vicinity of the reactive center, one interacting with the incoming phosphate chain (nucleotide-binding) and one bridging the nucleophile and phosphate-chain (catalytic; helps positioning the nucleophilic oxygen for in-line attack). We recently explored the behavior of a triphosphates in aqueous S<u><sub>N</sub></u><sup>2@@P</sup> reactions involving methoxyl-based...
nucleophiles, where both the number and position of Mg$^{2+}$ counterions and the degree of protonation were systematically varied [8]. The reactions were found to proceed via pentacoordinate transition states that generally are more stable if the nucleophile (ROH) is deprotonated (RO$^-$) [5], in line with the suggested proton hopping to the enzyme surroundings of the primer 3′ hydroxyl proton in the DNA polymerase active site [34–46]. Deprotonation of, and thus buildup of negative charge on, the triphosphate substrate resulted in an increase in barrier height, whereas inclusion of Mg$^{2+}$ counterions (counteraction of negative charge) on the deprotonated triphosphate stabilizes the transition state, resulting in a lower $S_N2@P$ barrier.

Despite the importance of DNA replication, the exact details of the mechanism are still not fully understood. Inspired in part by Westheimer’s work on why nature chose phosphates [30], we now address why nucleotide triphosphates, rather than diphosphates, are the energy carriers for the ubiquitous DNA replication. We accomplish this by extending upon our work on solvated triphosphates in $S_N2@P$ [8] with equivalent reactions involving diphosphates. Our analysis addresses the leaving group ability of phosphate and diphosphate, assesses the kinetics and thermodynamics of $S_N2@PP$ and $S_N2@PPP$, and investigates the effect protonation and Mg$^{2+}$ counterions on these model hydrolysis reactions.

### Computational details

#### Density functional and basis set

All calculations are based on density functional theory (DFT) [47–53] and have been carried out using the Amsterdam Density Functional program (ADF 2017.111) [54–56]. Geometries, single points and vibrational analyses were computed with a generalized gradient approximation (GGA) of DFT. All computations use the OLYP functional, which consists of the optimized exchange (OPTX) functional proposed by Handy and coworkers [57, 58], and the Lee–Yang–Parr (LYP) correlation functional [59]. Previously we have shown that OLYP reproduces $S_N2$ barriers from highly correlated ab initio benchmarks within a few kcal mol$^{-1}$ [60, 61]. Stationary points were confirmed as either minima (no imaginary frequencies) or transition states (one imaginary frequency) through a vibrational analysis. The transition states were connected to their associated minima by following the intrinsic reaction coordinate (IRC) [62]. The molecular orbitals (MOs) were expanded in a large TZ2P-type basis set consisting of uncontracted Slater-type orbitals (STOs). The TZ2P basis set is of triple-$\zeta$ quality and has been augmented with two polarization functions: 2$p$ and 3$d$ on hydrogen; 3$d$ and 4$f$ on carbon, oxygen, phosphorus and magnesium [63]. The core shells of carbon, oxygen, phosphorus and magnesium (1$s$) were treated with the frozen-core approximation. Scalar relativistic effects were taken into account by means of the zero order regular approximation (ZORA) [64–67]. The accuracy parameter of both the Becke grid integration and ZLMfit density fitting of the IRCs were set to GOOD, whereas geometry optimizations and frequencies analyses were calculated at VERYGOOD [68]. Spurious imaginary frequencies were set to 20 cm$^{-1}$, amounting to a correction of $-2.0$ and 0.6 kcal mol$^{-1}$ to $-T*S_v$ and $U_v$ (298.15 K, 1 atm), respectively. Structures were illustrated using CYLview [69].

#### Solvation model

Aqueous solvation was taken into account in all calculations using the COSMO implicit solvation model [70–72]. Solvent effects were explicitly used in the solving of the SCF equations and during the optimization of
the geometry and the vibrational analysis. The solvent radius \( R_s \) for water was taken from experimental data for the macroscopic density \( \rho \) and molecular mass \( M_m \) with the formula \( R_s^2 = 2.6752 \cdot M_m/\rho \), leading to a \( R_s \) value of 1.9 Å for water; a value of 78.4 was used for the dielectric constant of water. Atomic radii values were taken from the MM3 van der Waals radii [73], which are available for almost the whole periodic system, and scaled by 0.8333 (the MM3 radii are 20% larger than the normal van der Waals radii due to the specific form for the van der Waals energy within the MM3 force field). The surface charges at the Delley surface [74] were corrected for outlying charges. A finer grid was employed for the construction of the surface to minimize numerical noise, increasing the maximum number of points per atom for hydrogen from 194 to 302. This setup provides a ‘non-empirical’ approach to including solvent effects with a dielectric continuum, and works well for solvation processes [75]. Trends in triphosphate hydrolysis reactions are unchanged when either modeled in implicit solvation or with the inclusion of discrete water molecules to the coordination sphere of the Mg\(^{2+}\) counterions [9].

**Solution-phase activation strain analyses**

The activation strain model (ASM) [76–81], also known as the distortion/interaction model [82, 83], was used to obtain quantitative insight into the activation barriers of the studied S_{2}2 reactions in aqueous solution. This analysis involves decomposing the solution-phase potential energy surface \( \Delta E_{\text{solution}}(\zeta) \) along the reaction coordinate \( \zeta \) into the strain \( \Delta E_{\text{solution-strain}}(\zeta) \), associated with deforming the reactants from their equilibrium geometry, and interaction \( \Delta E_{\text{solution-int}}(\zeta) \) terms, accounting for interaction between the deformed reactants (equation (1)).

\[
\Delta E_{\text{solution}}(\zeta) = \Delta E_{\text{solution-strain}}(\zeta) + \Delta E_{\text{solution-int}}(\zeta).
\]  

(1)

The destabilizing \( \Delta E_{\text{solution-strain}}(\zeta) \) is determined by the rigidity of the reactants in solution and by the extent to which they must distort in order to achieve their geometry at a given point along the reaction coordinate. The \( \Delta E_{\text{solution-int}}(\zeta) \) is related to the electronic structure of the reactants in solution and how they are mutually oriented over the course of the reaction. These strain and interaction energy terms were calculated in the potential of the (implicit) solvent.

The solution-phase potential energy surface, \( \Delta E_{\text{solution}}(\zeta) \), was further decomposed into the \( \Delta E_{\text{solution}}(\zeta) \), which accounts for the interaction between solute and solvent, and the \( \Delta E_{\text{solvent}}(\zeta) \), which is the reaction system in vacuum with the solution-phase geometry (equation (2)) [9].

\[
\Delta E_{\text{solution}}(\zeta) = \Delta E_{\text{solvent}}(\zeta) + \Delta E_{\text{solute}}(\zeta).
\]  

(2)

The solute term, \( \Delta E_{\text{solvent}}(\zeta) \), is subsequently decomposed into a solvent-free strain and interaction energy (equation (3)), which we refer to as solute strain \( \Delta E_{\text{solvent-strain}}(\zeta) \) and solute interaction \( \Delta E_{\text{solvent-int}}(\zeta) \), respectively, to distinguish between the two solution-phase activation strain schemes of equations (1) and (2).

\[
\Delta E_{\text{solution}}(\zeta) = \Delta E_{\text{solvent}}(\zeta) + \Delta E_{\text{solvent-strain}}(\zeta) + \Delta E_{\text{solvent-int}}(\zeta).
\]  

(3)

**Energy decomposition analysis (EDA)**

In the conceptual framework provided by the Kohn–Sham molecular orbital (KS-MO) model, the solute interaction energy \( \Delta E_{\text{solvent-int}}(\zeta) \) can be further decomposed into three physically meaningful terms (equation (4)) [84–88].

\[
\Delta E_{\text{solvent-int}}(\zeta) = \Delta V_{\text{elstat}}(\zeta) + \Delta E_{\text{Pauli}}(\zeta) + \Delta E_{\text{P}}(\zeta).
\]  

(4)

The \( \Delta V_{\text{elstat}}(\zeta) \) term corresponds to the classical electrostatic interaction between unperturbed charge distributions \( \rho_A(r) + \rho_B(r) \) of the deformed fragments A and B and is usually attractive. The Pauli repulsion \( \Delta E_{\text{Pauli}}(\zeta) \) involves destabilizing interactions between occupied orbitals and is responsible for steric repulsion. The orbital interaction \( \Delta E_{\text{P}}(\zeta) \) accounts for charge transfer (interaction between occupied orbitals on one fragment with unoccupied orbitals of the other fragment, including HOMO–LUMO interactions) and polarization (empty–occupied orbital mixing on one fragment due to the presence of another fragment).

In the activation strain and energy decomposition plots, the IRC is projected onto the stretching of the P–O\(^{n}\) bond. This critical reaction coordinate \( \zeta \) undergoes a well-defined change over the course of the reaction from the reactant complex to the transition state and product complex [89]. The activation strain analysis was performed with the aid of the PyFrag program along the intrinsic reaction coordinate [90].

**Thermochemistry**

Enthalpies at 298.15 K and 1 atm (\( \Delta H^\circ \)) were calculated from the electronic bond energies and vibrational frequencies by using the thermochemistry relation for an ideal gas (equation (5)) [91, 92].

\[
\Delta H^\circ = \Delta E_{\text{trans},298} + \Delta E_{\text{rot},298} + \Delta E_{\text{vib},0} + \Delta (\Delta E_{\text{vib},0})_{298} + \Delta (p V).
\]  

(5)
Here, $\Delta E_{\text{trans,298}}$, $\Delta E_{\text{rot,298}}$, and $\Delta E_{\text{ vib,0}}$ are the differences between the fully protonated substrate (1dR and 1tR) and the fragments that result from the heterolytic cleavage of the P–O bond (i.e. the P$^+$ cation and the anionic leaving group) in translational, rotational, and zero-point vibrational energy, respectively. $\Delta (\Delta E_{\text{ vib,0}})_{298}$ is the change in the vibrational energy difference as one goes from 0 to 298.15 K. The vibrational energy corrections are based on our frequency calculations. The molar work term $\Delta(pV)$ is $\Delta G_T$; $\Delta n = +1$ for one substrate dissociating into two fragments. Thermal corrections for the electronic energy are neglected. The change in the Gibbs free energy ($\Delta G$) was calculated at 298.15 K at 1 atm ($\Delta G^\circ$) (equation (6)).

$$\Delta G^\circ = \Delta H^\circ - T \Delta S^\circ.$$  

(6)

**Results and discussion**

The $S_n2@P$ reactions and substrates investigated in the present study are summarized in table 1 and figure 2. The substrate in reactions 1d and 1t are the parent, fully protonated PP and PPP molecules. In reactions 2d and 2t, all the hydrogens on the substrate are removed, resulting in anionic substrates with the following notation, [PP–3H]$^-$ and [PPP–4H]$^{3-}$, respectively. To counteract the negative charge on the fully deprotonated phosphates, a Mg$^{2+}$ counterion is added to diphosphate in reaction 3d [PP–3H]$^-$ (Mg$^{2+}$)$_n$, a single Mg$^{2+}$ is added to triphosphate in reaction 3t [PPP–4H]$^{3-}$ (Mg$^{2+}$)$_2$, and finally two Mg$^{2+}$ are added to triphosphate in reaction 4t [PPP–4H]$^{3-}$ (Mg$^{2+}$)$_2$.

Unlike previously suggested mechanisms [34–46], our simple model system, based on a relatively small methoxide nucleophile, highlights a preferential interaction of the Mg$^{2+}$ ion(s) with the phosphate chain to produce tridentate phosphate–(Mg$^{2+}$)$_n$ complexes [8]. In the absence of the surrounding enzyme active site, the mutual stabilization of both the phosphate negative charge and magnesium positive charge is preferred, contrasting to the catalytic nucleophile–bridging function adopted by one of the Mg$^{2+}$ counterions in e.g. DNA replication [35]. Nevertheless, the placement of Mg$^{2+}$ counterions mimics the positions in the reported crystal structure of DNA polymerase [46].

We show how deprotonated and therefore negatively charged substrates give rise to higher central reaction barriers under attack by a deprotonated organic nucleophile, methoxide, and how Mg$^{2+}$ counterions can effectively reduce these barriers again. Whereas in the gas phase, the pentacoordinate intermediates can potentially rise to highly stable product complexes.

The $S_n2@P$ reactions and substrates investigated in the present study are summarized in table 1 and figure 2. The substrate in reactions 1d and 1t are the parent, fully protonated PP and PPP molecules. In reactions 2d and 2t, all the hydrogens on the substrate are removed, resulting in anionic substrates with the following notation, [PP–3H]$^-$ and [PPP–4H]$^{3-}$, respectively. To counteract the negative charge on the fully deprotonated phosphates, a Mg$^{2+}$ counterion is added to diphosphate in reaction 3d [PP–3H]$^-$ (Mg$^{2+}$)$_n$, a single Mg$^{2+}$ is added to triphosphate in reaction 3t [PPP–4H]$^{3-}$ (Mg$^{2+}$)$_2$, and finally two Mg$^{2+}$ are added to triphosphate in reaction 4t [PPP–4H]$^{3-}$ (Mg$^{2+}$)$_2$.

Unlike previously suggested mechanisms [34–46], our simple model system, based on a relatively small methoxide nucleophile, highlights a preferential interaction of the Mg$^{2+}$ ion(s) with the phosphate chain to produce tridentate phosphate–(Mg$^{2+}$)$_n$ complexes [8]. In the absence of the surrounding enzyme active site, the mutual stabilization of both the phosphate negative charge and magnesium positive charge is preferred, contrasting to the catalytic nucleophile–bridging function adopted by one of the Mg$^{2+}$ counterions in e.g. DNA replication [35]. Nevertheless, the placement of Mg$^{2+}$ counterions mimics the positions in the reported crystal structure of DNA polymerase [46].

We show how deprotonated and therefore negatively charged substrates give rise to higher central reaction barriers under attack by a deprotonated organic nucleophile, methoxide, and how Mg$^{2+}$ counterions can effectively reduce these barriers again. Whereas in the gas phase, the pentacoordinate intermediates can potentially be stable minima, solvation changes the nature of the central species to a labile transition state for all reactions. For the systems most resembling biological systems [PPP–4H]$^{3-}$ (Mg$^{2+}$)$_2$, a deprotonated triphosphate chain with two Mg$^{2+}$ counter ions proceeds with the lowest computed energy barrier. Relative energies with respect to the reactants for all reactions, both di- and triphosphates in solution, are summarized in table 1. Trends in relative electronic energy and relative Gibbs free energy are fully consistent and trends are unchanged. Schematic structures with $O^{\text{nucl}}$–P$^+$ and P$^+$–O$^-$ distances for the reactions are given in figure 1. Cartesian coordinates and additional details are provided in the supporting information.

As previously shown, stable transition complexes (TC) in the gas phase can turn into labile transition states (TS) in solution, the transition from TC to TS depending on the solvent polarity and spatial extent of the substrate LUMO [9]. All reactions investigated in the current study proceeds via a labile central TS in the aqueous phase. Only reactions involving the fully protonated species (1d and 1t) proceed from a reactant complex (RC), to the transition state and then to a product complex (PC). While still lacking an RC, Mg$^{2+}$ (3d, 3t and 4t) gives rise to highly stable product complexes.

Nucleophilic attack by CH$_3$O$^-$ at the neutral PP (1d) and PPP (1t) substrates must compete with proton transfer from the $\alpha$-proton to the nucleophile, rendering these reactions unviable from a biochemical standpoint. The barrierless and highly exothermic proton transfer (−44.8 and −46.6 kcal mol$^{-1}$ for 1d and 1t, respectively) converts the reactive methoxide nucleophile into methanol, forming a stable hydrogen bonded RC with the deprotonated substrate. No RC is observed for reactions lacking a protonated substrate (i.e. 2d, 2t, 3d, 3t, and 4t).

The pentavalent transition states (TS) are greatly influenced by the presence of Mg$^{2+}$ counter ions, systematically lowering barriers for $S_n2@PP$ from 34.9 [2t – (Mg$^{2+}$)$_0$], to 23.3 [3t – (Mg$^{2+}$)$_1$], to 9.9 [4t – (Mg$^{2+}$)$_2$] kcal mol$^{-1}$. Earlier TSs on the reaction coordinate (longer $O^{\text{nucl}}$–P$^+$ and shorter P$^+$–O$^-$ distances) are observed for reactions with lower activation barriers (figure 1). The catalytic effect of Mg$^{2+}$ was previously traced back to (i) an enhanced interaction between nucleophile and triphosphate with more stabilizing electrostatic and orbital interactions and (ii) weakening of the leaving-group bond leading to a reduced activation strain [8]. Mg$^{2+}$ has the same catalytic effect on PP as it does on PPP, decreasing the $S_n2@PP$ barriers from 42.1 [2d – (Mg$^{2+}$)$_0$] to 27.4 [3d – (Mg$^{2+}$)$_1$] kcal mol$^{-1}$. All transition states, products and reactant/product complexes associated with PP are destabilized with respect to its PPP counterpart, an effect which is most pronounced in the two (Mg$^{2+}$)$_0$ systems (2d and 2t), barriers and reaction energies respectively being 8.2 and 10.0 kcal mol$^{-1}$ lower for PPP.

Weakly bound product complexes (PC) are observed for the neutral PP (1d) and PPP (1t) systems, stabilized by hydrogen bonding between dimethylphosphate and the protonated mono-/diphosphate leaving group.
and PP. Coordination of Mg2+ phase PES (EΔ3(2d, 2t, 3d, 3t, and 4t) and S1(1d and 1t) (stacks.iop.org/EST/1/024001/mmedia)) decompose the solution-strain) and subsequent interaction between the deformed reactants (ΔE_solution-int) in (implicit) solvation. The activation strain analysis is performed along a reaction coordinate defined by the geometry (ΔE_solution-strain) and subsequent interaction between the deformed reactants (ΔE_solution-int) in (implicit) solvation. The activation strain analysis is performed along a reaction coordinate defined by the

- **Table 1.** Model S₈₂@P reactions in aqueous solution with relative energies of stationary points (in kcal mol⁻¹) and overall charge Q (in a.u.). Gibb's free energies (298.15 K, 1 atm) are bolded.²

| Rxn | Designator | Q | RC | TS | PC | Pb |
|-----|-------------|---|----|----|----|----|
| 1d  | PP          | −1| 0.0 (−44.8) | 0.0 (−32.9) | 29.9 | 31.2 | 21.1 | 24.6 | 21.0 | 10.6 |
| 1t  | PPP         | −1| 0.0 (−46.6) | 0.0 (−35.0) | 23.7 | 24.2 | 14.0 | 17.5 | 14.8 | 3.7  |
| 2d  | [PP-3H]+    | −4| —  | 42.1 | 56.2 | —  | −12.9 | −10.1 |
| 2t  | [PPP-4H]+   | −5| —  | 34.9 | 49.6 | —  | −22.9 | −22.0 |
| 3d  | [PP-3H]+(Mg²⁺) | −2| —  | 27.4 | 37.8 | −34.9 | −25.3 | −2.1  | −2.9 |
| 3t  | [PPP-4H]+(Mg²⁺) | −3| —  | 23.3 | 34.6 | −35.0 | −22.9 | −6.2  | −7.5 |
| 4t  | [PPP-4H]+(Mg²⁺) | −1| —  | 9.9  | 23.9 | −63.5 | −49.7 | −12.5 | −14.1 |

² Computed at COSMO(H₂O)-ZORA-OLYP/TZ2P. If a bound reactant complex (RC) exists, its energy relative to separate reactants, i.e. E(RC)−E(R), is shown in parentheses.

³ The reaction energy, i.e. E(P)−E(RC), in the presence of the counterion is calculated based on chelation of Mg²⁺ to the leaving group.

as is observed in the RCs, phosphate deprotonation (2d and 2t) causes the disappearance of the PC by removing the phosphates capability as a hydrogen bond donor. In contrast, the introduction of Mg²⁺ (3d, 3t, and 4t) gives rise to highly stabilized PCs, chelation of the Mg²⁺ counterion by both dimethylphosphate and the mono-/diphosphate leaving group yielding PC energies as low as −63.5 kcal mol⁻¹ (4t). Note that this stable PC is likely bypassed completely in the enzyme active site (e.g. DNA polymerase), as the interaction of Mg²⁺ with key amino acid residues, an interaction unaccounted for in our simplified model system, facilitates the stabilization and retention of Mg²⁺.

**Activation strain and EDA**

In order to shed light on the factors leading to the different diphosphate and triphosphate S₈₂ reactivities, we next performed an activation strain analysis for all reactions. The resulting activation strain diagrams (figures 3(2d, 2t, 3d, 3t, and 4t) and S1(1d and 1t) (stacks.iop.org/EST/1/024001/mmedia)) decompose the solution-phase PES (ΔE_solution) into the energy associated with distorting the reactants from their equilibrium geometry (ΔE_solution-strain) and subsequent interaction between the deformed reactants (ΔE_solution-int) in (implicit) solvation. The activation strain analysis is performed along a reaction coordinate defined by the stretching of the P⁻-O²⁻ bond [89]. The effect of Mg²⁺ on the PES (figure 3(a)) is similar for both PPP and PP. Coordination of Mg²⁺ counterions to the substrate facilitates a more stabilizing interaction between nucleophile and the di-/triphasphate substrate by lowering the di-/triphasphate LUMO and depleting its negative charge (figure 3(a)) [8].

Turning our attention to the reactivity differences between PP and PPP (2d and 2t) without Mg²⁺ counterions, we find that 2d proceeds with a higher barrier than 2t, due to more destabilizing activation strain along the entire reaction coordinate (figure 3(b)). The interaction curves are similar and, thus, reactivity differences can be attributed to differences in strain curves. Introduction of a single Mg²⁺ leads to lower, but more similar barriers for reactions 3d and 3t. The kinetic preference for the triphosphate is thus diminished (figure 3(c)). The counterion stabilizes the negative charge of the leaving group for both cases. 3t goes with a slightly more stabilizing interaction curve compared to 3d due to the more stabilized LUMO of [PP-3H]+(Mg²⁺), leading to more favorable charge transfer, but the differences in strain curves are again responsible for the differences in barrier height.

Our activation strain analysis reveals that the primary factor leading to the enhanced reactivity of triphosphates compared to diphosphates, both with or without the Mg²⁺ counterion, is the less destabilizing strain along the entire reaction coordinate. The origin of the more destabilizing activation strain for S₈₂@PP was traced to
the stronger bond between the electrophilic P° and the monophosphate leaving group. The computed heterolytic bond dissociation energies (BDE), which are energetically favored over homolytic bond cleavage in solution [93], between P° and the mono-/diphosphate leaving group for systems 1dR and 1tR are shown in tables 2 and S1. The BDE for the leaving group bond of PP is, indeed, larger ($\Delta H_{\text{solution}} = 49.9 \text{ kcal mol}^{-1}$) than for PPP ($\Delta H_{\text{solution}} = 40.2 \text{ kcal mol}^{-1}$).
The EDA in the gas phase was used to understand and rationalize the differing bond strengths of 1d and 1t (table 3). The interaction between Pα and the leaving group is more stabilizing for PP compared to PPP in reactants 1d and 1t. Analysis of the interaction energy at every point along the reaction coordinate also reveals a more stabilizing interaction for PP (figure S1). Taken together, both the interaction and bond dissociation energy point to the same conclusion: PP has a stronger Pα−Oα bond than PPP. Additionally, the trends observed for ΔE_{solute-int} (table 3) also hold for the interaction energy in solution ΔE_{solution-int} (table S2).

By decomposing the ΔE_{solution-int}, the prominent role of electrostatics and, to a lesser extent, orbital interactions in contributing to the relative strength of the Pα−Oγ bonds becomes apparent. A relationship was found between the degree of charge localization on the leaving group and the magnitude of the electrostatic interaction (see scheme S1 for details). Negative charge localization on the leaving group results in an increasingly stabilizing electrostatic interaction with Pα and thus a stronger bond. The stabilizing electrostatic interaction associated with PP (1d), and by extension its stronger Pα−Oγ bond, are rationalized by its localized negative charge on the small monophosphate leaving group. Conversely, a more delocalized negative charge, as observed for the larger diphosphate leaving group of SN2@PPP (1t), leads to a less stabilizing electrostatic interaction and a weaker leaving group bond.

Moreover, we have also analyzed the bond strength of the Oγ−Pγ bond in PPP using the EDA method. The Oγ−Pγ bond is the key phosphodiester bond that is broken during SN2@Pγ hydrolysis reactions in ATPase.
B van Beek et al.

Comparable trends in $O^{\gamma}−O^{\gamma}$ and $P^{\alpha}−O^{\alpha}$ bond dissociation energies (table S3) and interaction energies (table S4) are found, highlighting the applicability of our findings for these other ubiquitous enzymatic reactions.

Conclusions

We have computationally investigated the potential energy surfaces of seven aqueous SN2@P reactions involving di- and triphosphates, substrates which play a key role in biological processes such as DNA replication. Our DFT study reveals that the hydrolysis SN2@PPP reactions proceed with lower activation free energies and are more exergonic than the corresponding SN2@PP reactions. These hydrolysis reactions all proceed via a pentavalent transition state, which is stabilized by the coordination of Mg$^{2+}$ counterions.

The enhanced reactivity of PPP compared to PP was traced back to the weaker $P^{\alpha}−O^{\alpha}$ bond of PPP. This reduced bond strength lowers the strain associated with breaking the $P^{\alpha}−O^{\alpha}$ bond over the course of the reaction and thus the overall reaction barrier. The weaker $P^{\alpha}−O^{\alpha}$ bond in PPP originates from a less stabilizing electrostatic interaction between the leaving group and electrophilic phosphorus center $P^{\alpha}$, compared to PP, caused by the more delocalized negative charge of the PPP leaving group. Mg$^{2+}$ counterions diminish the reactivity differences between PP and PPP by stabilizing the negative charge of the leaving group. Stabilization of the LUMO by Mg$^{2+}$, which is prominent in PPP, yields strain and interaction terms that still clearly favor PPP over PP in Mg$^{2+}$-catalyzed systems, although differences in reactivity slightly diminish when compared to the uncatalyzed SN2@P reactions.

In conclusion, we find that both SN2@PP and SN2@PPP are exothermic and exergonic. Importantly, however, our results underscore both the favorable kinetics and thermodynamics of the SN2@PPP reactions. Thus, PPP can fulfill its role as an energy-rich species that stores energy and fuels up-hill reactions in biological systems. This energetic preference is a factor that contributes to making PPP the universal energy storage molecule in biological processes, including DNA replication. Additionally, by also computing the bond dissociation energy along with the $P^{\gamma}−O^{\gamma}$ (for PPP) and $P^{\beta}−O^{\beta}$ (for PP) interaction energies, we find that our previous conclusions also hold, that is, PP is a better leaving group than P, due to a weaker leaving-group bond in these corresponding SN2@P$^3$ hydrolysis reaction central to ATPases and GTPases.

Acknowledgments

This work was supported by The Netherlands Organization for Scientific Research (NWO) through the Planetary and Exo-Planetary Science (PEPSci) program and the Dutch Astrochemistry Network (DAN). We thank SURFsara for use of the Lisa supercomputer.

| Table 2. Heterolytic $P^{\alpha}−O^{\alpha}$ bond dissociation energy (in kcal mol$^{-1}$) of 1dR and 1tR in aqueous solution$^a$. |
|---|---|---|
| System | $\Delta H_{\text{solution}}$ | $-T\Delta S_{\text{solution}}$ | $\Delta G_{\text{solution}}$ |
| 1dR | 49.9 | −14.0 | 35.8 |
| 1tR | 40.2 | −13.6 | 26.2 |

$^a$ Computed at COSMO(H$_2$O)-ZORA-OLYP/TZ2P at 298.15 K and 1 atm.

| Table 3. $P^{\alpha}−O^{\alpha}$ EDA (in kcal mol$^{-1}$) of 1dR and 1tR in the gas phase$^a$. |
|---|---|---|---|
| System | $\Delta E_{\text{solute-int}}$ | $\Delta V_{\text{elstat}}$ | $\Delta E_{\text{Pauli}}$ | $\Delta E_{\text{oi}}$ |
| 1dR | −232.7 | −309.1 | 324.5 | −248.1 |
| 1tR | −203.9 | −281.1 | 309.0 | −231.9 |

$^a$ Computed at ZORA-OLYP/TZ2P. See equation (4) for details on the energy decomposition terms.

and GTPase [13, 14, 94]. Comparable trends in $O^{\gamma}−P^{\gamma}$ and $P^{\alpha}−O^{\alpha}$ bond dissociation energies (table S3) and interaction energies (table S4) are found, highlighting the applicability of our findings for these other ubiquitous enzymatic reactions.
References

[1] Smith M B 2013 March´s Advanced Organic Chemistry: Reactions, Mechanisms, and Structure 7th edn (New York: Wiley)
[2] Hamlin T A, Swart M and Bickelhaupt F M 2018 ChemPhysChem 19 1315–30
[3] Nakamura T, Zhao Y, Yamagata Y, Hua Y J and Yang W 2012 Nature 487 196–201
[4] Kottler J and Nair D T 2018 Nucleic Acids Res. 46 5875–85
[5] van Bochove M A, Swart M and Bickelhaupt F M 2006 J. Am. Chem. Soc. 128 10738–44
[6] van Bochove M A, Swart M and Bickelhaupt F M 2007 ChemPhysChem 8 2452–63
[7] van Bochove M A and Bickelhaupt F M 2008 Eur. J. Org. Chem. 2008 649–54
[8] van Bochove M A, Roos G, Fonseca Guerra C, Hamlin T A and Bickelhaupt F M 2018 Chem. Commun. 54 3448–51
[9] Hamlin T A, van Beek B, Wolters L P and Bickelhaupt F M 2018 Chem. Eur. J. 24 5927–38
[10] Barrozo A, Blaha-Nelson D, Williams N H and Kamerlin S C L 2017 Pure Appl. Chem. 89 715–27
[11] Kiani F A and Fischer S 2016 BMC Biochem. 17 12
[12] Perera L, Freudenthal B D, Beard W A, Shook D D, Pedersen I G and Wilson S H 2015 Proc. Natl Acad. Sci. USA 112 E5228–36
[13] Mironov V A, Khenrova M G, Lychko L A and Nemukhin A V 2015 Proteins 83 1046–53
[14] Prasad B R, Plotnikov N V, Lameira J and Warshel A 2013 Proc. Natl Acad. Sci. USA 110 20590–14
[15] Cleland W W and Hengge A C 2006 Chem. Rev. 106 3525–78
[16] Kumara Swamy K C and Satish Kumar N 2006 Acc. Chem. Res. 39 324–33
[17] Range K, McGrath M J, Lopez X and York D M 2004 J. Am. Chem. Soc. 126 1654–65
[18] Kamerlin S C L and Wilkie J 2011 Org. Biomol. Chem. 9 5394–406
[19] Duarte E, Barrozo A, Aqvist J, Williams N H and Kamerlin S C L 2016 J. Am. Chem. Soc. 138 10664–73
[20] Thoden J B, Timson D J, Reece R J and Holden H M 2005 J. Biol. Chem. 280 9662–70
[21] Aleshin A E, Zeng C, Bourenkov G P, Bartunik H D, Fromm H J and Honzatko R B 1998 Structure 6 39–50
[22] Cronin C N and Tipton K F 1987 Biochem. J. 247 41–6
[23] Jenkins B T and Hajra A K 1976 J. Neurochem. 26 377–85
[24] Flynn T G and McKay D 1972 Biochem. J. 128 135P
[25] Jitrapakdee S and Wallace J C 1999 Biochem. J. 340 1–16
[26] Berg J M, Tymoczko J L and Stryer L 2012 Biochemistry 7th edn (New York: W H Freeman)
[27] Alberts B 2003 Nature 421 431–5
[28] Kool E T 2001 Annu. Rev. Biophys. Biomol. Struct. 30 1–22
[29] Kuchta R D, Benkovic P and Benkovic S J 1988 Biochemistry 27 6716–25
[30] Westheimer F 1987 Science 235 1173–8
[31] Lassila J K, zalatan J G and Herschlag D 2011 Annu. Rev. Biochem. 80 669–702
[32] Kamerlin S C, Sharma P K, Prasad R B and Warshel A 2013 Q. Rev. Biophys. 46 1–132
[33] Petrovic D, Szelker K and Kamerlin S C L 2018 Chem. Commun. 54 3077–89
[34] Suzato C A et al 2008 Biochemistry 47 870–9
[35] Bojin M D and Schlöck T 2007 J. Phys. Chem. B 111 11244–52
[36] Radhakrishnan R and Schlöck T 2006 Biochem. Biophys. Res. Commun. 350 521–9
[37] Lin P, Pederson L C, Batra Y K, Beard W A, Wilson S H and Pedersen L G 2006 Proc. Natl Acad. Sci. USA 103 13294–9
[38] Beard W A and Wilson S H 2006 Chem. Rev. 106 861–82
[39] Beard W A and Wilson S H 2003 Structure 11 889–96
[40] Steitz T A 1999 J. Biol. Chem. 274 17395–8
[41] Steitz T A 1998 Nature 391 231–2
[42] Sawaya M R, Prasad R, Wilson S H, Krout J and Pelletier H 1997 Biochemistry 36 11205–15
[43] Steitz T A, Smardon S J, Jäger J and Joyce G M 1994 Science 266 2022–5
[44] Pelletier H, Sawaya M, Kumar A, Wilson S and Krout J 1994 Science 264 1891–903
[45] Besse L S and Steitz T A 1991 EMBO J. 10 29–35
[46] Batra V K, Beard W A, Shook D D, Kraln J M, Pedersen L C and Wilson S H 2006 Structure 14 757–66
[47] Koch W and Holthausen M C 2001 A Chemist’s Guide to Density Functional Theory (Weinheim: Wiley–VCH)
[48] Gross E K U and Dreizler R M 2013 Density Functional Theory (New York: Springer)
[49] Baerends E J and Grimseken O V 1997 J. Phys. Chem. A 101 5383–403
[50] Ziegler T 1995 Can. J. Chem. 73 743–61
[51] Ziegler T 1991 Chem. Rev. 91 651–67
[52] Parr R G and Yang W 1994 Density-Functional Theory of Atoms and Molecules 1st edn (New York: Oxford University Press)
[53] Bickelhaupt F M and Baerends E J 2000 Reviews in Computational Chemistry ed K B Lipkowitz and D B Boyd (Hoboken, NJ: Wiley)
[54] et al te Velde G, Bickelhaupt F M, Baerends E J, Fonseca Guerra C, van Gisbergen S J A, Snijders J G and Ziegler T 2001 J. Comput. Chem. 22 931–67
[55] Fonseca Guerra C, Snijders J G, et al te Velde G and Baerends E J 1998 Theor. Chem. Acc. 99 391–403
[56] ADF 2017.111 2010 SCM Theoretical Chemistry (Amsterdam: Vrije Universiteit) (http://scm.com)
[57] Lucks J B, Cohen A J and Handy N C 2002 Phys. Chem. Chem. Phys. 4 4612–8
[58] Handy N C and Cohen A J 2001 Mol. Phys. 99 403–12
[59] Lee C, Yang W and Parr R G 1988 Phys. Rev. B 37 785–93
[60] Bento A P, Solà M and Bickelhaupt F M 2008 J. Chem. Theory Comput. 4 929–40
[61] Bento A P, Solà M and Bickelhaupt F M 2005 J. Comput. Chem. 26 1497–504

ORCID iDs

Bas van Beek  https://orcid.org/0000-0003-2463-6559
Trevor A Hamlin  https://orcid.org/0000-0002-5128-1004
F Matthias Bickelhaupt  https://orcid.org/0000-0003-4655-7747
[62] Hare S R, Hudson B M and Tantillo D J 2018 Modeling organic reactions—general approaches, caveats and concerns Applied Theoretical Organic Chemistry ed D J Tantillo (Singapore: World Scientific)
[63] van Lenthe E and Baerends E J 2003 J. Comput. Chem. 24 1142–56
[64] van Lenthe E, Baerends E J and Snijders J G 1993 J. Chem. Phys. 99 4597–610
[65] van Lenthe E, Baerends E J and Snijders J G 1994 J. Chem. Phys. 101 9783–92
[66] van Lenthe E, van Leeuwen R, Baerends E J and Snijders J G 1996 Int. J. Quantum Chem. 57 281–93
[67] van Lenthe E, Ehlers A and Baerends E J 1999 J. Chem. Phys. 110 8943–53
[68] Franchini M, Filipson P H T and Visscher L 2013 J. Comput. Chem. 34 1819–27
[69] Legault C Y 2009 CYLview 1.0b (Sherbrooke, QC: Université de Sherbrooke) (http://cylview.org)
[70] Klamt A and Schüürmann G 1993 J. Chem. Soc. Perkin Trans. 2 799–805
[71] Klamt A 1995 J. Phys. Chem. 99 2224–35
[72] Pys C and Ziegler T 1999 Theor. Chem. Acc. 101 396–408
[73] Allinger N L, Zhou X and Bergsma J 1994 J. Mol. Struct.: Theochem. 312 69–83
[74] Delley B 2006 Mol. Sustain. 32 117–23
[75] Swart M, Rösel E and Bickelhaupt F M 2007 Eur. J. Inorg. Chem. 3646–54
[76] Bickelhaupt F M and Houk K N 2017 Angew. Chem., Int. Ed. 56 10070–86
[77] Bickelhaupt F M and Houk K N 2017 Angew. Chem. 129 10204
[78] Wolters L P and Bickelhaupt F M 2015 Wires Comput. Mol. Sci. 5 324–43
[79] Fernández I and Bickelhaupt F M 2014 Chem. Soc. Rev. 43 4953–67
[80] van Zeist W-J and Bickelhaupt F M 2010 Org. Biomol. Chem. 8 3118–27
[81] Bickelhaupt F M 1999 J. Comput. Chem. 20 114
[82] Ess D H and Houk K N 2007 J. Am. Chem. Soc. 129 10646–7
[83] Ess D H and Houk K N 2008 J. Am. Chem. Soc. 130 10187–98
[84] Bickelhaupt F M, Nibbering N M M, van Wezenbeek E M and Baerends E J 1992 J. Phys. Chem. 96 4864–73
[85] Bickelhaupt F M, Diefenbach A, de Visser S P, de Koning L J and Nibbering N M M 1998 J. Phys. Chem. A 102 9549–53
[86] Ziegler T and Rauk A 1979 Inorg. Chem. 18 1755–9
[87] Hamlin T A, Svatunek D, Yu S, Riddell L, Infante I, Visscher L and Bickelhaupt F M 2018 Eur. J. Org. Chem. 378–86
[88] Yu S, de Brujin H M, Svatunek D, Hamlin T A and Bickelhaupt F M 2018 ChemistryOpen 7 993–1004
[89] Galabov B, Koleva G, Schäfer H P III and Allen W D 2018 Chem. Eur. J. 24 11637–48
[90] van Zeist W-J, Fonseca Guerra C and Bickelhaupt F M 2008 J. Comput. Chem. 29 312–5
[91] Atkins P and de Paula J 2002 Atkins’ Physical Chemistry (Oxford: Oxford University Press)
[92] Jensen F 2007 Introduction to Computational Chemistry 2nd edn (West Sussex: Wiley)
[93] Osuna S, Swart M, Baerends E J, Bickelhaupt F M and Solà M 2009 ChemPhysChem 10 2955–65
[94] Liu Q, Nassar N and Wang J 2011 Chem. Phys. Lett. 516 233–8