JCTC Journal of Chemical Theory and Computation

LETTER

pubs.acs.org/JCTC

Methyl Phosphate Dianion Hydrolysis in Solution Characterized by Path Collective Variables Coupled with DFT-Based Enhanced Sampling Simulations

Davide Branduardi,[†] Marco De Vivo,[†] Nadia Rega,[‡] Vincenzo Barone,[§] and Andrea Cavalli*[,†,∥]

[†]Department of Drug Discovery and Development, Italian Institute of Technology, Via Morego 30, I-16163 Genoa, Italy

[‡]Dipartimento di Chimica, Università "Federico II", Complesso Universitario di Monte S.Angelo, via Cintia, I-80126 Napoli, Italy

[§]Scuola Normale Superiore, piazza dei Cavalieri 7, I-56126 Pisa, Italy; INFN, Sezione di Pisa, Italy

[∥]Department of Pharmaceutical Sciences, University of Bologna, Via Belmeloro 6, I-40126 Bologna, Italy

Supporting Information

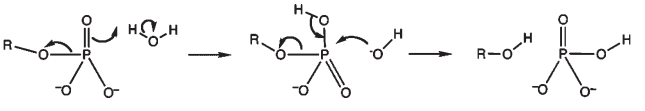
ABSTRACT: Herein, we propose a conceptually innovative approach to investigating reaction mechanisms. This study demonstrates the importance of considering explicitly the effects of large amplitude motions, aside from the intrinsic reaction coordinate, when tuning the free energy landscape of reaction pathways. We couple the path collective variables method with DFT-based enhanced sampling simulations to characterize the associative mechanism of the hydrolysis of the methyl phosphate dianion in solution. Importantly, energetics and mechanistic differences are observed when passing from the potential to the free energy surface.

Traditionally, chemical reaction mechanisms are unraveled in terms of stationary points of potential energy surfaces (PESs) and their harmonic environment.[1] It is then straightforward to recover the free energy differences using standard statistical thermodynamics equations in the rigid rotor/harmonic oscillator approximation. Such calculations provide an order-parameter-free description and allow researchers to study concerted mechanisms, providing reliable reaction rates. However, even simple chemical reactions can experience anharmonic effects that influence the range of applicability of such approaches. While small anharmonic effects experienced by semirigid systems can be effectively considered using perturbative approaches,[2] proper consideration of large-amplitude motions for large systems is much more involved. Conversely, molecular dynamics (MD)-based free energy calculations can deal with arbitrarily large amplitude motions. However, such calculations are usually carried out using simple geometrical descriptors. Subsequently, complex free energy landscapes must be split into separated steps, preventing accurate investigation of multievent mechanisms. Here, we apply an MD-based approach that can identify the free energy landscape using a single collective descriptor. This approach continuously captures the complexity of concerted multievent reaction mechanisms. We use this novel approach to identify the free energy landscape of the associative mechanism of a phosphoryl transfer reaction, using methyl phosphate dianion hydrolysis as a prototypical case study. Our approach reveals energetics and mechanistic features of this important reaction with unprecedented detail.

Phosphoryl transfers are ubiquitous in biology since they represent the crucial chemical process of many metabolic pathways. The hydrolysis and subsequent transfer of phosphates is efficiently catalyzed by a huge number of enzymes involved in energy production, replication of genetic material, biosynthesis, and protein control mechanisms.[3,4] One possible phosphoryl transfer mechanism displays an associative transition state (TS). Here, the nucleophilic attack precedes the departure of the leaving group, which leads to a phosphorane-like TS geometry. This mechanism can be concerted (A_ND_N, see Scheme 1) when stable intermediates are not identified along the reaction pathway. In the past few decades, there has been significant computational and experimental effort to elucidate the different mechanisms of phosphoryl transfer in vacuo, in solution, and in enzymes.[5−11] However, the energetics and mechanistic details are still under active investigation.[5−11] As a paradigmatic case study, here we focus on the A_ND_N mechanism of the methyl phosphate dianion hydrolysis, shedding further light on the mechanistic details related to the reaction's free energy profile.

We have performed a detailed description of the reaction's free energy landscape, coupling for the first time the path collective variables (PCVs) method[12] with ADMP (atom-centered density matrix propagation)[13] available in G09[14] for DFT-based MD simulations, which is similar to the Car−Parrinello scheme[15] in a localized basis set. The integration time step adopted was 0.2 fs. A

Scheme 1. Concerted Associative Mechanism (A_ND_N) for Phosphoryl Transfer



Received: September 23, 2010
Published: January 05, 2011

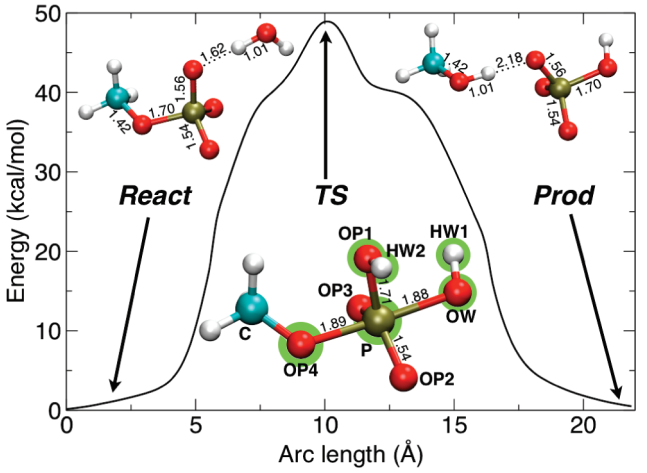


Figure 1. Reaction mechanism as extracted from TS optimization and IRC calculations. Bond distances are in Å. Green shadowed atoms are used for tracing the PCVs.

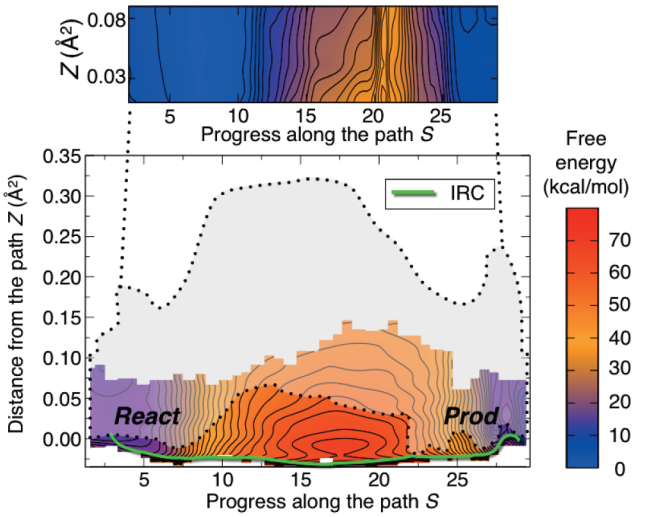


Figure 2. FES in S and Z for the first umbrella sampling run (bottom panel). The shaded gray region represents the sampled region with the second US, whose FES in the new S and Z is magnified in the upper panel. The projection of the IRC pathway on S and Z is displayed in light green. Negative values of distance Z are due to its definition,[12] as a result of a negative logarithm whose argument can be larger than one.

canonical ensemble was enforced by a stochastic rescaling thermostat[16] with a period of 50 fs. Solvation effects were accounted for using a C-PCM model,[17] one of the implicit solvation approaches which have been reported to reproduce experimental results well for these kinds of reactions.[6,10,18] The PLUMED package[19] was implemented in a special stand-alone version to add enhanced sampling (ES) capabilities to the G09 program. In this way, the ADMP module could perform the umbrella sampling (US) and steered-MD needed to depict the free energy landscape of the reaction in question. Importantly, the PCVs approach simultaneously considers proton transfer (PT) events, dipole moment reorientation, and conversion of the scissile phosphate along the reaction with a single collective variable.

The model system was formed by the methyl phosphate and the nucleophilic water molecule. The PCVs method aims to find low energy pathways that connect reactants and products. It requires an initial guess path in terms of Cartesian coordinates. Thus, we first identified a reactive path for the A_ND_N TS that was obtained by B3LYP calculations in continuum solvent, followed by intrinsic reaction coordinate (IRC) calculations initiated from the TS imaginary frequency.[20] This provided the reactants-to-products pathway on the PES (Figure 1). The reaction mechanism was characterized by a very compact TS, in which the lengths of the forming (OW−P) and breaking (OP4−P) bonds were 1.88 Å and 1.89 Å, respectively. The water molecule was already fully deprotonated in the TS, while the HW2 proton was fully transferred to the oxygen OP1 of the scissile phosphate (OP1−HW2 = 0.97 Å and OW−HW2 = 2.79 Å). Overall, the phosphorane-like TS geometry resembled the typical S_N2-like mechanism that characterizes the concerted A_ND_N associative mechanism, with no intermediate along the reaction path (Scheme 1). The activation energy was about 49 kcal/mol.

The A_ND_N pathway found on the PES was then used as a starting guess for the reaction mechanism for subsequent PCVs/ES calculations carried out on the free energy surface (FES). PCVs consist of two collective variables (S and Z) that describe the progression along (S), and the distance from (Z), the reaction pathway in terms of mean square deviations measured on a few key atoms involved in the reactivity.[21] These atoms are the most representative of the chemical reaction, while the remaining atoms are neglected. This choice was verified a posteriori by checking that the neglected atoms were correctly averaged out during the simulations (see the Supporting Information, SI). Then, steered-MD and US calculations, coupled with PCVs, were performed to identify the reaction pathways on the FES.

After a preliminary B3LYP/ADMP-based 1D-US run along S, we performed a more detailed bidimensional (2D) adaptive US to capture the difference between the starting IRC-based path and the simulated one (see SI). A total of 190 ps of sampling in both S and Z variables was carried out to determine a pathway of the reaction on the FES.[22] The reconstruction of the 2D FES (Figure 2) was performed using the fitting procedure originally developed by Maragliano and Vanden-Eijnden.[23] In particular, we adopted the version recently modified by Monteferrante et al.[24] Figure 2 includes the projection of the IRC pathway in S and Z (light green line), which lies on a narrow channel at low values of Z. Only a couple of kilocalories per mole separated IRC from a wider region at lower energy, which indicated alternative pathways on the FES. To further refine the FES, we performed steered-MD along the S variable. A confining wall over Z allowed the system relative freedom to relax onto the new pathway, while preventing the system from escaping the reactive region.[25] A set of frames was then extracted from the steered-MD trajectory. This set represented a novel guess path for subsequent PCVs/ES calculations. Further 2D US simulations were then performed, as previously described (see ref 22 and the SI). The newly sampled region is shaded gray in Figure 2 and magnified in the upper panel. This indicates the location of the FES associated with the novel reaction path with respect to the original IRC. The resulting free energy projected over S and Z showed almost vertical isolines (upper panel of Figure 2), which indicated that the new reference path was lying in a wider, low, free energy pathway, and allowed for a meaningful 1D projection (Figure 3A). Here, we point out that currently our approach does not account for quantum effects, although they are estimated to be negligible for the reaction under study.[10]

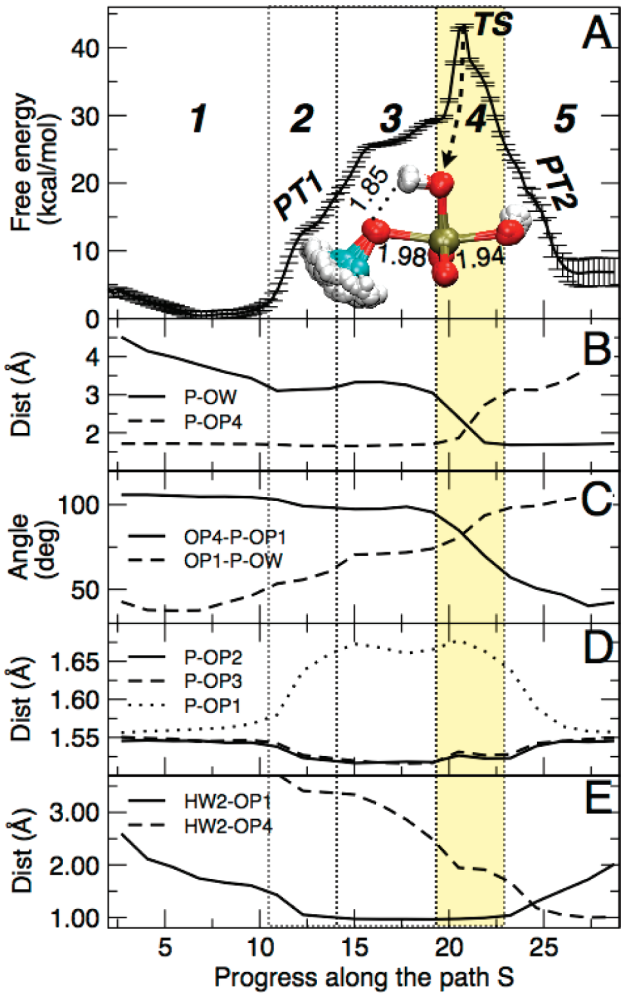


Figure 3. (A) 1D free energy profile over S (the progress of the reaction). The transition state is also displayed in ball-and-stick representation. (B−E) Behavior of the relevant structural features during the reaction. The TS region is highlighted in yellow.

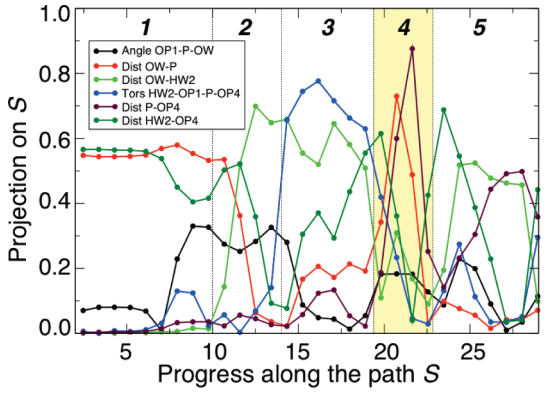


Figure 4. Absolute values of the projection of the gradients of relevant degrees of freedom over the gradient of the path variable S suitably normalized. Degrees of freedoms showing high values play a pivotal role in the reaction progress. This pattern was used to define the five regions as shown also in Figure 3. This partitioning scheme could be in principle applied to other chemical reactions.

The energy changes observed along the path could be ascribed to five consecutive structural events generated by different relevant components in the gradient plot (see Figure 4). When $S \in [1, 10]$ (region 1), the nucleophilic water approached the scissile phosphate, and the distance P−OW was markedly shortened without any apparent energetic penalty (Figure 3A, B). In the interval $S \in [10, 13]$ (region 2), proton HW2 was transferred from the water to OP1, which led to a barrier of ∼10 kcal/mol on the FES (Figure 3, PT1 in panel A and E). This barrier is in agreement with that reported by Florián and Warshel.[10] A decrease of the nonbridging oxygen lengths P−OP2 and P−OP3 (Figure 3D), together with a lengthening of the bridging P−OP1 bond, was also associated with this event. In principle, PT1 could also happen on OP2 or OP3. Therefore, we disclosed three degenerate pathways for this key event. Our calculations showed that this degeneracy did not significantly affect the free energy profile, since it accounted for a negligible change of ∼0.2 kcal/mol (see the SI). At $S \in [13, 19]$ (region 3), the slope of the free energy profile changed, showing a continuous increase of 15 kcal/mol. This was mainly due to the HW2−OP1 dipole reorientation with respect to the phosphate dipole, which facilitated the nucleophilic attack of the hydroxyl ion, otherwise unlikely. To further investigate the role of dipole reorientation, we carried out a gradient analysis versus the reaction variable S (see Figure 4).

This pointed to the driving components in the mean force, when moving from the reactants to the products. As far as the dipole reorientation was concerned, we could consistently observe that the torsion angle HW2−OP1−P−OP4 showed the highest contribution in region 3.

At this point, the phosphate inversion was due to the concerted change of P−OW and P−OP4 bond lengths, and OP4−P-OP1 and OP1−P−OW angles (Figure 3B,C). The associative TS geometry is represented by an "ensemble" of structures in Figure 3A. In the TS, the lengths of the forming (OW−P) and breaking (OP4−P) bonds were 1.94 Å and 1.98 Å, respectively. The geometrical features of the TS identified on the FES still pointed to a concerted associative mechanism. When $S \in [22.5, 29]$ (region 5), a late PT facilitated the leaving group departure (see Figure 3, PT2 in panels A and E).

The main difference with the FES-TS when compared to the PES-TS was the orientation of the proton HW2, which here pointed toward the leaving group. The HW2 PT1 and the subsequent rotation of HW2 around the P−OP1 bond is a necessary preliminary step; this key event was fundamental to inducing the nucleophilic attack of the hydroxide ion with a favorable dipole interaction. In addition, the forming and breaking bonds were ∼0.1 Å longer in the FES-TS when compared to those identified in PES-TS. This difference was well captured in the More O'Ferrall−Jencks (MOFJ) plot, which, in Figure 5, is reported for a set of configurations along the minimum free energy pathway obtained via US and steered-MD. Remarkably, both US and steered-MD showed a less associative character for the methyl phosphate hydrolysis on the FES, compared to that identified on the PES via IRC. In particular, as shown in Figure 5, steered-MD captured the concerted nature of the mechanism, although it overemphasized the less associative character of the nucleophilic attack to the phosphate. The estimated free energy barrier was 43 (±0.7) kcal/mol, in good agreement with the experimental value of 44−47 kcal/mol. The free energy associated with this barrier was about 6 kcal/mol less than the PES activation energy. With these simulations, we have shown that effects related to the intrinsic anharmonicity of the FES can

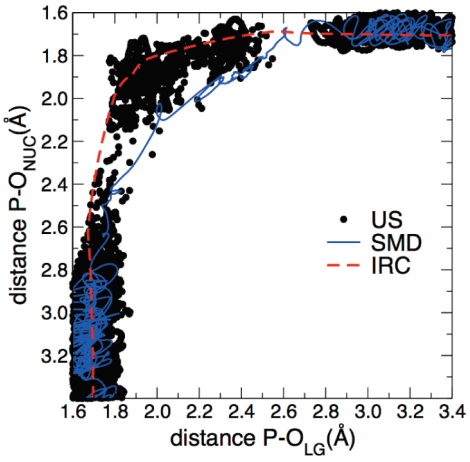


Figure 5. More O'Ferral−Jencks (MOFJ) plot for the IRC pathway. The configurations from the minimum free energy path obtained from US and steered-MD run over the low free energy pathway. P−O_{NUC} and P−O_{LG} are the distances between P and the nucleophile oxygen (OW) and the leaving group (OP4), respectively.

indeed change a reaction's TS in terms of structure and energetics with respect to a TS identified on the PES.

In summary, we have herein reported an innovative approach to the study of reaction mechanisms. Complementing traditional investigations on the PES, our approach focuses on the reaction's FES, highlighting and quantifying the smoothing effects of thermal energy on pathways identified initially on the PES. We have shown this to be true even for simple reactions, such as methyl phosphate hydrolysis. In this respect, the present methodology can be seen as extending the standard static calculations for a saddle point search toward methods that allow a direct sampling of the phase space of a chemical reaction, including its entropic effects. For instance, this approach could be very helpful in studying enzymatic reactions. In addition, it could be instrumental to better understanding linear free energy relationship experiments, which measure the change in reaction rates as a response of a different leaving or attacking group. Finally, from a mechanistic standpoint, the present study has offered a detailed understanding of the key factors that regulate the associative mechanism of a phosphoryl transfer reaction, including the proton shuttle required for a more proficient nucleophilic attack and departure of the leaving group.

■ ASSOCIATED CONTENT

Supporting Information. DFT B3LYP structure coordinates of reagents, TSs, and products; 1D umbrella sampling results; tests over the choice of the degrees of freedom involved; and a description of the 2D umbrella sampling procedure. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: andrea.cavalli@iit.it.

■ ACKNOWLEDGMENT

Giovanni Bussi is kindly acknowledged for providing the stochastic rescaling routines. We thank IIT Integrated Computational Multiscale Platform for the computational time.

■ REFERENCES

(1) Thrular, D. G.; Pliego, J. R., Jr. In *Continuum Solvation Models in Chemical Physics: From Theory to Application*; Mennucci, B., Cammi, R., Eds.; Wiley: Chichester, U. K., 2008; pp 338−365.
(2) Barone, V. *J. Chem. Phys.* **2005**, *122*, 014108.
(3) Johnson, L. N.; Lewis, R. J. *Chem. Rev.* **2001**, *101*, 2209–42.
(4) Knowles, J. R. *Annu. Rev. Biochem.* **1980**, *49*, 877–919.
(5) Aqvist, J.; Kolmodin, K.; Florián, J.; Warshel, A. *Chem. Biol.* **1999**, *6*, R71–R80.
(6) Klähn, M.; Rosta, E.; Warshel, A. *J. Am. Chem. Soc.* **2006**, *128*, 15310–15323.
(7) De Vivo, M.; Dal Peraro, M.; Klein, M. L. *J. Am. Chem. Soc.* **2008**, *130*, 10955–10962.
(8) De Vivo, M.; Ensing, B.; Dal Peraro, M.; Gomez, G. A.; Christianson, D. W.; Klein, M. L. *J. Am. Chem. Soc.* **2007**, *129*, 387–394.
(9) Florián, J.; Goodman, M. F.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 8163–8177.
(10) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 719–734.
(11) Kamerlin, S. C. L.; Florián, J.; Warshel, A. *ChemPhysChem* **2008**, *9*, 1767–1773.
(12) Branduardi, D.; Gervasio, F. L.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 054103.
(13) (a) Schlegel, H. B.; Millam, J. M.; Iyengar, G. A.; Voth, G. A.; Scuseria, G. E.; Daniels, A. D.; Frisch, M. J. *J. Chem. Phys.* **2001**, *114*, 9758–9763. (b) Iyengar, S. S.; Schlegel, H. B.; Millam, I. M.; Voth, G. A.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **2001**, *115*, 10291–10302.
(14) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision A.2; Gaussian, Inc., Wallingford CT, 2009.
(15) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471–2474.
(16) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
(17) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669–681.
(18) Kamerlin, S. C. L.; Haranczyk, M; Warshel, A. *Chem. Phys. Chem.* **2009**, *10*, 1125–1134.
(19) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, F. A.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
(20) (a) Hratchian, H. P.; Schlegel, H. B. In *Theory and Applications of Computational Chemistry: the First 40 Years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005, pp 195−249. (b) A 6-31++G** basis set was adopted in all of the present calculations. The IRC provided 200 steps in both forward and backward directions.
(21) The reaction in question comprises three distinct events: (1) PT1 from the water molecule to the phosphate dianion, (2) rotation of the proton HW2 around the oxygen OP1, and (3) transfer of the proton HW2 to the leaving group (PT2). Therefore, we considered in PCVs representation the following key atoms: OW, HW2, HW1, P, OP1, and OP4. From the IRC trajectory, 30 equally spaced frames were selected, with an average root-mean-square deviation of 0.12 Å. This allowed us to choose a value of λ = 110.0 $Å^{-2}$. λ is the smoothing parameter for running PCVs calculations.

(22) Umbrella potentials on both S and Z spaces for a total of 506 grid points. The spring constant on the S was 1.2×10^3 kcal/mol, while that chosen for Z was 5×10^5 kcal/mol Å^{-4}.

(23) Maragliano, L.; Vanden-Eijnden, E. *J. Chem. Phys.* **2008**, *128*, 184110.

(24) Monteferrante, M; Bonella, S.; Meloni, S.; Ciccotti, G. *Mol. Simul.* **2009**, *35*, 1116–1129.

(25) A harmonic wall over the Z variable was imposed for values higher than 0.04 Å^2. A relative soft spring constant of 3.1 kcal/mol Å^{-4} with a scaling factor of 0.04 Å^2 was chosen.

(26) Lad, C.; Williams, N. H.; Wolfenden, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5607–5610.