Competing Reaction Mechanisms of Peptide Bond Formation in Water Revealed by Deep Potential Molecular Dynamics and Path Sampling

Rolf David, Iñaki Tuñón, Damien Laage

The competing reaction mechanisms of peptide bond formation in water revealed by deep potential molecular dynamics and path sampling

Rolf David,[†] Iñaki Tuñón,[‡] and Damien Laage[*,†]

†*PASTEUR, Department of Chemistry, École Normale Supérieure, PSL University, Sorbonne Université, CNRS, 75005 Paris, France*

‡*Departamento de Química Física, Universitat de Valencia, Burjassot, Valencia 46100, Spain*

E-mail: damien.laage@ens.psl.eu

Abstract

The formation of an amide bond is an essential step in the synthesis of materials and drugs, and in the assembly of amino-acids to form peptides. The mechanism of this reaction has been studied extensively, in particular to understand how it can be catalyzed, but a representation capable of explaining all the experimental data is still lacking. Numerical simulation should provide the necessary molecular description, but the solvent involvement poses a number of challenges. Here, we combine the efficiency and accuracy of neural network potential-based reactive molecular dynamics with the extensive and unbiased exploration of reaction pathways provided by transition path sampling. Using microsecond-scale simulations at the density functional theory level, we show that this method reveals the presence of two competing distinct mechanisms for

peptide bond formation between alanine esters in aqueous solution. We describe how both reaction pathways, via a general base catalysis mechanism and via direct cleavage of the tetrahedral intermediate respectively, change with pH. This result contrasts with the conventional mechanism involving a single pathway in which only the barrier heights are affected by pH. We show that this new proposal involving two competing mechanisms is consistent with the experimental data, and we discuss the implications for peptide bond formation under prebiotic conditions and in the ribosome. Our work shows that integrating deep potential molecular dynamics with path sampling provides a powerful approach for exploring complex chemical mechanisms.

Introduction

The amide bond is a ubiquitous chemical motif and its formation is of central importance in many areas of chemistry.[1] In materials chemistry, it links the building blocks of polyamide, which is used, for example, in synthetic fabrics such as nylon. In pharmaceutical chemistry, it is present in more than half of all new drug candidates, making amine acylation one of the most common reactions in the synthesis of pharmaceutical compounds.[2,3] In biochemistry, amide bonds are responsible for holding together the backbones of all proteins, since peptide bonds are amide bonds between amino acids.

The direct formation of an amide by condensing an amine with a carboxylic acid is well known to be extremely difficult, due to the very large energetic barrier. This is why, in synthetic chemistry the reaction is usually preceded by a first step in which the carboxylic acid is activated[1,4,5] and in living organisms the formation of peptide bonds is catalyzed, *e.g.*, by the ribosome.[6]

However, critical questions remain unanswered. These include, for example, guidelines for the design of industrial catalysts that provide atom-economic preparation routes for amides,[3,4,7,8] and a plausible reaction pathway for the formation of peptide bonds under prebiotic conditions in the absence of enzyme catalysts to elucidate the origins of life.[9] All

these questions require a detailed characterization of the amide bond formation mechanism.

The commonly accepted mechanism is based on the works of Jencks[10–13] and shown in figure 1a. It successively involves the nucleophilic attack of the amine on the carbonyl to yield a tetrahedral zwitterionic intermediate $\mathbf{T}^{\pm}$, followed by the amine deprotonation by a general base to yield the negatively charged tetrahedral intermediate $\mathbf{T}^{-}$, and finally the departure of the leaving group.

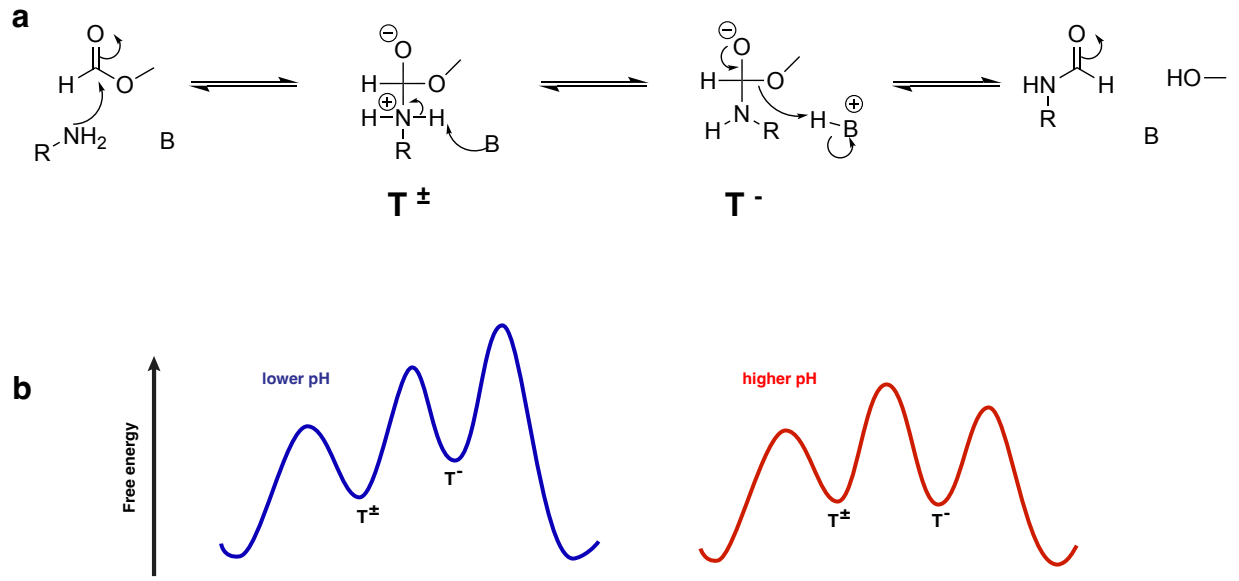


Figure 1: a) Conventional mechanism of amide bond formation[10–13] with general base catalysis. b) Schematic free energy profiles for the conventional mechanism at lower pH (pH=8) and at higher pH (pH=10).[6,11]

Breaks in the measured pH-rate profiles of amide formation revealed that the reaction rate-limiting step changes with pH.[10] It was proposed[11,12] that at high pH (pH=10) the rate-limiting step involves a tetrahedral transition state structure, consistent with the deprotonation step between $\mathbf{T}^{\pm}$ and $\mathbf{T}^{-}$ (figure 1b). In contrast, at lower pH (pH=8), it involves the breaking of the CO bond with the leaving group, the transition state is no longer tetrahedral and the rate-limiting step was suggested to be the breakdown of $\mathbf{T}^{-}$ (figure 1b).

However, while this mechanism is widely presented in chemistry textbooks, it has been questioned by kinetic isotope substitution measurements at different pH values. A landmark kinetic isotope effect (KIE) study[14] of the amide bond formation between hydrazine and

methyl formate raised questions regarding the stability of the tetrahedral intermediate.[14–16] Other experimental studies on acyl transfer have further suggested that depending on the nucleophile and nucleofuge structures the mechanism could become concerted.[17–20] The concerted or stepwise character of the nucleophilic attack and proton transfer steps therefore remains an open question.

An additional difficulty with the conventional mechanism (figure 1a) is that it leads to a chemically surprising picture of pH effects on the successive barriers (figure 1b). pH is indeed expected to affect the $\mathbf{T}^{\pm} \rightarrow \mathbf{T}^{-}$ proton rearrangement step much more than the $\mathbf{T}^{\pm}$-forming nucleophilic attack and than the $\mathbf{T}^{-}$ cleavage. In more basic pH conditions, the $\mathbf{T}^{\pm} \rightarrow \mathbf{T}^{-}$ deprotonation is expected to be increasingly facilitated. In contrast, the KIE study[14,16] concludes that $\mathbf{T}^{\pm} \rightarrow \mathbf{T}^{-}$ is rate-limiting in very basic conditions whereas it is not at almost neutral pH (figure 1b). This behavior is counter-intuitive and this suggests that fundamental questions about the mechanism of this widely used reaction remain to be answered.

Resolving these questions and obtaining a detailed characterization of the uncatalyzed reference mechanism is a prerequisite for guiding the catalysis of this key reaction. The molecular resolution provided by numerical simulations would therefore be extremely helpful. Pioneering electronic structure calculations[15,21–23] and molecular dynamics simulations[24–36] of amide formation and hydrolysis reactions have been performed and have for example confirmed the initial formation of the tetrahedral zwitterionic intermediate $\mathbf{T}^{\pm}$. However, molecular simulations face two major difficulties in the description of the amide formation reaction mechanism.

The first challenge arises from the active role played by the solvent. The rate-limiting steps probed in the KIE study[14] include proton transfers, which possibly involve the surrounding water molecules. While previous simulations[25,28,34] with a mixed quantum/classical (QM/MM) description have provided important insight in this reaction, this method requires an *a priori* partitioning of solvent molecules into reactive and non-reactive regions. This se-

lection can be ambiguous and often imposes that only a small number of water molecules be considered reactive.

The second challenge for simulations lies in the determination of a relevant reaction coordinate. The mechanism is known to involve multiple coordinates, including at least the CN amide bond that is formed, the CO ester bond that is cleaved by the leaving group departure, and the two proton rearrangements on the attacking amine nitrogen and on the leaving group oxygen atom. However, as emphasized in previous studies,[14–16] the stepwise or concerted character of these rearrangements and their specific order remain unknown. This information must therefore result from the analysis of the mechanism and not be postulated *a priori* to perform the analysis. While several techniques have been developed to determine transition pathways between initial and final states across complex energetic landscapes, their widespread application to the exploration of chemical reaction mechanisms has been so far limited by their requirement of long simulations, resulting in tremendous computational cost when used with ab initio or QM/MM molecular dynamics.

Here we solve both difficulties by taking advantage of recent developments in machine learning.[37,38] We design a reactive force field based on neural network potentials (NNPs) which reproduces a high-level electronic structure description of the system but at a fraction of the cost. The NNPs' very moderate computational cost has two major advantages: first the entire system can be described at the same electronic structure level so that any solvent molecule can be involved in the reaction, and second, NNPs can be combined with methods requiring extensive sampling to identify reaction pathways. Here we use transition path sampling (TPS)[39,40] and show how it can be advantageously integrated with NNPs to characterize the peptide bond formation mechanism. The TPS method has already achieved remarkable success for complex pathways in non-reactive physical and biophysical systems (see, *e.g.*, refs. 41–45), and although its potential for the study of chemical reactions has already been emphasized,[46] with pioneering works on, *e.g.*, water autodissociation,[47] enzyme catalysis[48,49] and electron transfer,[50] its general application to reactivity has so far been

limited by its computational cost. Here, this is solved by the efficiency of the NNP, while retaining a description of the energies and forces equivalent to that of accurate but costly density functional theory (DFT) calculations.

In this study, we focus on the formation mechanism of a peptide bond, due to the ubiquitous nature of this bond in biochemistry, and the important remaining questions regarding its formation in prebiotic conditions and its catalysis by the ribosome. We consider the peptide bond formation between two L-alanine methyl esters as a model system for peptide bond formation in aqueous solution. In the following, we first describe our simulation methodology: after an iterative training of the NNP at the DFT level to ensure that the system is properly described along the entire transformation, we use TPS to identify the reaction pathway. We then present our results: in contrast to the conventional picture, they reveal two competing reaction mechanisms for peptide bond formation. We characterize these mechanisms and show that they explain the experimental KIE measurements and elucidate the rate-limiting step changes with pH. We finally present the implications of our results for amide bond and peptide bond formation catalysis in the pharmaceutical and prebiotic chemistry contexts and provide some concluding remarks.

Methods

System. Our simulated system comprises two L-alanine methyl ester molecules solvated by 300 water molecules in a cubic box of side 21.046 Å which is periodically replicated in three dimensions, corresponding to an amino ester concentration of approximately 0.3 mol/L.

Neural network potential. We train a neural network potential to describe the potential energy surface of the entire system at the BLYP-D3 density functional level of theory (DFT), thus treating all solvent water molecules as potentially reactive. Reference DFT calculations are performed with CP2K[51] with a triple-ζ (TZV2P) basis set. NNP training and calculations employ the DeePMD-kit.[52] The exploration of the potential energy surface is

performed by short molecular dynamics trajectories propagated in the canonical ensemble at 300 K with LAMMPS[53] using the NNP forces and biased sampling techniques implemented in PLUMED.[54] The training set is generated iteratively by adding configurations until the NNP is able to correctly describe the part of the potential energy surface relevant for the chemical reaction. Particular attention is paid to considering all possible mechanisms and to sampling configurations close to the transition states. The training set thus combines configurations from unconstrained trajectories respectively started in the reactant state, in the product state and from guesses of the transition state, and configurations from biased trajectories propagated with several enhanced sampling techniques including steered molecular dynamics, metadynamics and umbrella sampling along coordinates which undergo an important change during the reaction (see SI for details). The full training set contains more than 76,000 configurations, covering every stage of the peptide bond formation. The final converged NNP reproduces the BLYP energies and forces with a computational cost reduced by approximately four orders of magnitude. It is validated on an independent set of close to 4,000 configurations collected at different stages of the bond formation. The root mean square error on the atomic forces is less than 0.04 eV·Å^{-1}, which supports the validity of the NNP description (see SI).

Transition path sampling. A number of techniques have been proposed to identify reaction coordinates (see, *e.g.*, refs. 55–59). When the reaction coordinate is expected to involve only a few relevant degrees of freedom, approaches like the string method[55] can successfully provide the reaction pathway.[60] However, these are not easily applied to the study of peptide bond formation, which can require a series of proton transfers involving many solvent molecules. We therefore use the transition path sampling (TPS) method[39,40] which generates the transition path ensemble via an iterative Monte Carlo scheme and which presents two key advantages. First, it includes the full set of system coordinates; it is free of any *a priori* choice of reaction coordinate to drive the reaction and it only requires the definition of reactant and product states. Its second important feature is that it provides an ensemble

of unbiased trajectories connecting reactants and products; it therefore naturally accounts for dynamical effects such as barrier recrossing, which have been suggested to be important in prior applications of TPS to enzymatic reactions[48,49] for example. We use the TPS implementation in OpenPathSampling[61,62] combined with the OpenMM[63] molecular dynamics engine, for which we wrote an interface to use the forces from the NNP calculated with DeePMD. Transition pathways were generated via TPS from seven independent initial reactive trajectories. The latter were obtained from steered molecular dynamics along different coordinates and utilized to conduct 74 independent TPS simulations initiated with different random seeds (see SI). More than 66,000 uncorrelated transition path trajectories have been generated, for a total simulated time of more than 0.5 μs. These trajectories have then been analyzed to determine the sequence of rearrangements and to identify the transition state location via a committor analysis (see, *e.g.*, refs.[39,64]).

Collective variables. A molecular interpretation of the reaction pathways is obtained from the *a posteriori* projection of TPS trajectories onto a selection of coordinates. We emphasize that, in contrast with many other methods, this selection is done after the trajectories are generated, and the trajectories are independent of the selected coordinates. We focus on five key coordinates. First, the progress of the nucleophilic attack is monitored by considering the difference between the lengths of the CN bond being formed and of the CO bond being cleaved, $\Delta d = d_{CN} - d_{CO}$. Second, the deprotonation of the attacking amine and the protonation of the leaving methoxy group are probed by the hydrogen coordination numbers of the amine nitrogen atom (C_N) and of the leaving group oxygen atom (C_O). Finally, proton transfers involving solvent water molecules and leading to the formation of hydronium or hydroxide ions in the solution are probed by the minimum and maximum hydrogen coordination numbers of all water oxygen atoms, $C_w^{\mathrm{min,max}}$ (see SI for details).

Free energy calculations. Free energy profiles are calculated along each of the two average pathways identified with TPS. Using the approach introduced in ref. 65, we define path collective variables for each pathway, in order to measure the progress (s) along the

curvilinear reaction coordinate and the deviation (z) away from the path (see SI). While the free energy landscape can be estimated from an analysis of TPS trajectories,[66] in order to systematically converge the reaction barriers, we calculate the free energy profiles for each mechanism from a weighted histogram analysis of a series of umbrella sampling simulations along the reaction progress coordinate s for a total simulation time of more than 0.1 μs (see SI).

Results and Discussion

We simulate the formation of a peptide bond between two methylated L-alanine amino acids in neat water. L-alanine is chosen because it is the smallest natural amino acid with a side chain. We consider the methylated form of the amino acids for two reasons. First, because it has been used in a number of experimental and simulation studies[9,14,22,28,67] of amide bond formation which showed that the mechanism is the same with and without methylation. Second, because methanolate is a better leaving group than hydroxide, making it interesting in the context of uncatalyzed prebiotic peptide bond formation.

We consider the protonation state corresponding to neutral or moderately basic pH conditions to connect to previous experimental studies.[9,14] There is no ambiguity regarding the protonation state of the carboxylic acids since the latter are methylated. Regarding the amine groups, their p$K_\mathrm{a}$is known from experiments to decrease by 2 units from 9.6 to 7.6[68,69] upon methylation of the carboxylic acid, which implies that the neutral and protonated forms have similar abundances at neutral pH conditions. Since the nucleophilic addition requires the neutral amine lone pair's availability and cannot take place with the protonated ammonium, we focus on the neutral amine form for the reactant state.

We explore the reaction pathways for the peptide bond formation via TPS. The reactant and product states are defined by specifying the difference $\Delta d = d_{CN} - d_{CO}$ (Δd >1.5 Å in the reactant where the CN bond is not formed, and Δd <-1.5 Å in the product where the CO

bond is cleaved), and the difference $\Delta C = C_N - C_O$ between the hydrogen coordinations of the attacking amine nitrogen atom C_N and of the leaving group oxygen atom C_O (ΔC >1.75 in the reactant where $C_N \simeq 2$ and $C_O \simeq 0$, and ΔC <O in the product where $C_N \simeq 1$ and $C_O \simeq 1$). After collecting a TPS ensemble of several tens of thousands of reactive trajectories, we project them on the $\Delta d, C_N, C_O$ coordinates to obtain a molecular picture of the reaction pathway. We determine the probability for each $(\Delta d, C_N, C_O)$ configuration of being visited by a reactive trajectory and we represent the contour surface of all points with a probability above a minimum threshold to highlight the most important reaction pathways (figure 2). We emphasize that while free energy profiles will be computed below, for now we only consider the probability to be part of the TPS ensemble.

The major result revealed by figure 2 is that peptide bond formation can follow two separate and very distinct reaction channels, each defined by different sequences of molecular rearrangements. These pathways appear in figure 2 with finite widths, reflecting the probability dispersion around minimum free energy paths due to thermal energy. These two distinct concurrent mechanisms cannot be easily distinguished when only the conventional nucleophilic attack coordinate Δd is used, as done in some previous simulation studies. For each pathway, TPS provides a straightforward identification of the transition states. Since TPS trajectories are generated with unbiased dynamics, reaction transition states are located by identifying the regions from which TPS trajectories have equal probabilities to end in the reactant and product states from a committor analysis.[39,64] Our results (colored symbols in figure 2) clearly show that the two pathways have different transition state structures and, thus, are expected to have different free energy barriers. Extending prior studies which had already used TPS to reveal competing pathways in other contexts,[42,43] our results therefore further emphasize that it is a powerful method to explore chemical reaction mechanisms without any *a priori* selection of coordinates.

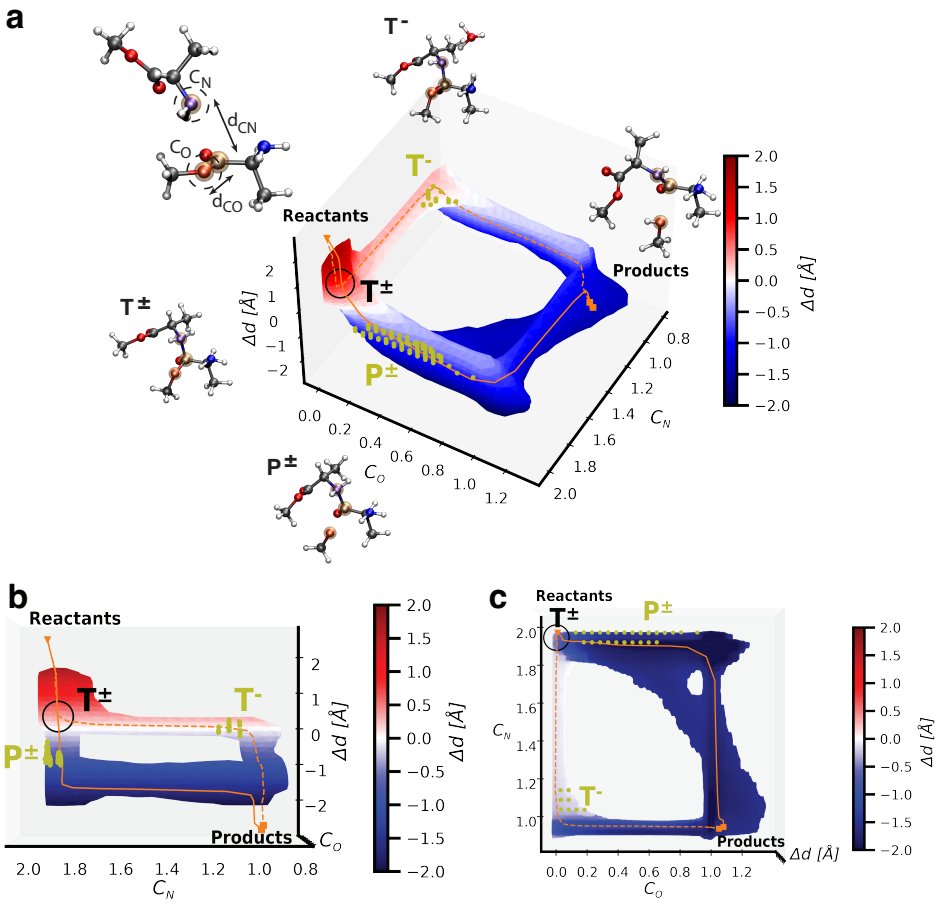


Figure 2: a) Contour surface of the most probable configurations within reactive trajectories collected by TPS, represented in perspective along the $\Delta d = d_{CN} - d_{CO}$ difference of distances and the oxygen C_O and nitrogen C_N coordination numbers (see definitions on reactant structure), together with complementary views focusing b) on the $(\Delta d, C_N)$ plane and c) on the (C_O, C_N) plane. Insets show key structures along the paths, yellow symbols mark the transition state locations obtained from a committor analysis, the orange dashed (mechanism **1**) and solid (mechanism **2**) lines show the most probable pathways, and the surface is colored according to the Δd value.

The two reaction schemes identified by the TPS study are summarized in figure 3. Both mechanisms initially involve the formation of the tetrahedral zwitterionic structure $\mathbf{T}^{\pm}$, in agreement with the conventional picture in figure 1 and with previous calculations.[28,34] However, the two mechanisms differ strongly in the order of the leaving group departure and of the proton rearrangements involving the attacking amine deprotonation and the methoxy leaving group protonation.

Mechanism **1** follows a pathway similar to that of the conventional mechanism proposed by Jencks[11–13] (but we already note that its free energy profile will be shown below to be very different from that depicted in Figure 1b). First the amine of the $\mathbf{T}^{\pm}$ intermediate is deprotonated by general base catalysis, here involving a water molecule as the proton acceptor in neutral pH conditions. This leads to the $\mathbf{T}^{-}$ anionic tetrahedral intermediate, which is the transition state structure for this mechanism (figure 3). This is finally followed by CO bond elongation, prompting the departure of the methoxy leaving group and its protonation by the solvent to yield the product.

Mechanism **2** revealed by our simulations had so far not been considered in the literature on amide bond formation in solution. It contrasts with mechanism **1** because after formation of the $\mathbf{T}^{\pm}$ intermediate, it almost exclusively involves the CO bond elongation to reach its transition state $\mathbf{P}^{\pm}$, without the proton transfer found in mechanism **1**. The transition state of mechanism **2** is structure $\mathbf{P}^{\pm}$, in which the carbonyl carbon atom is no longer tetrahedral and approaches the planar geometry found in the peptide bond. The CO bond elongation induces an increasingly negative charge on the leaving group oxygen atom, while the nucleophilic nitrogen atom is still positively charged, leading to a marked zwitterionic character for the transition state. The negative charge developed on the leaving group leads to a nascent hydrogen-bond accepted from water. The latter explains the fairly broad range of C_O values at the transition state, corresponding to different hydrogen-bond strengths. However, at the transition state C_O remains below 1, showing that no proton exchange occurs before the transition state is reached. Mechanism **2** therefore does not involve acid-

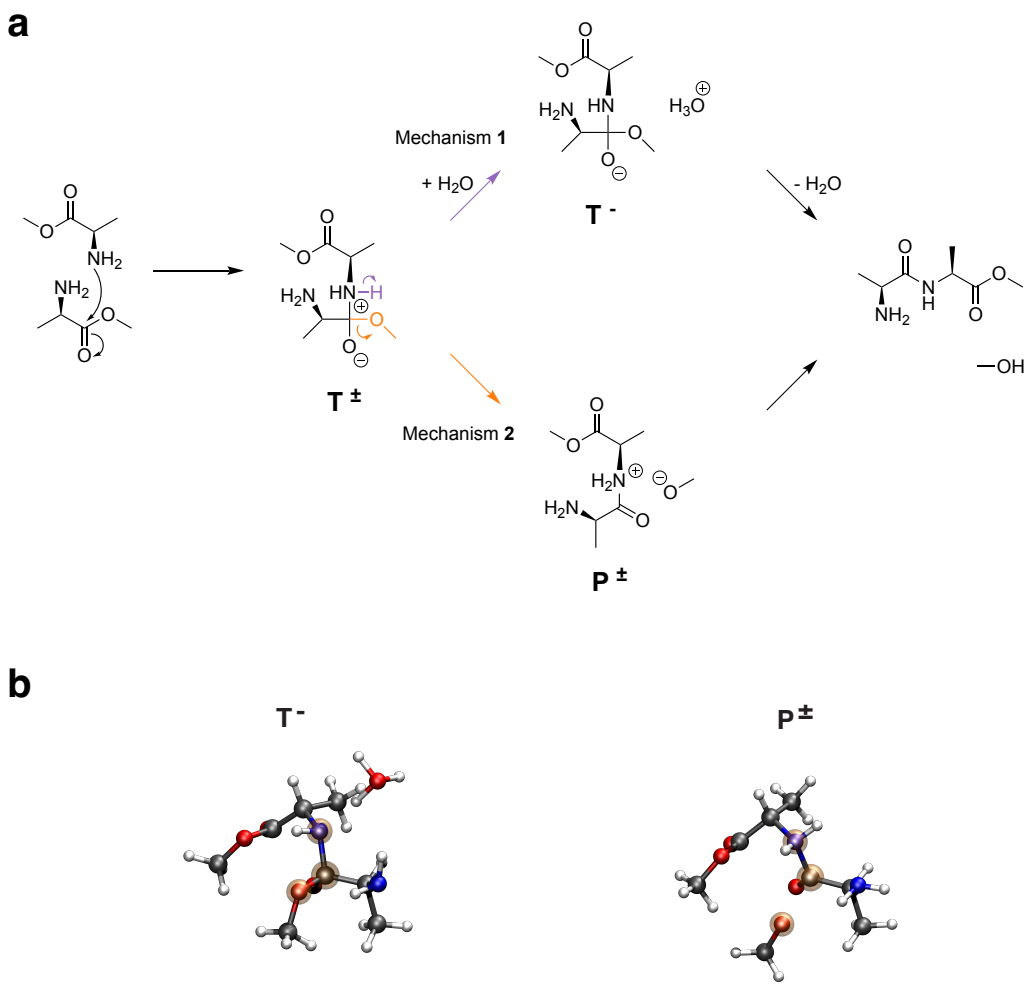


Figure 3: a) Schematic representation of the two reaction pathways obtained by TPS in figure 2, with mechanism **1** (top, violet) and mechanism **2** (bottom, orange). This representation only shows the pathways and the $\mathbf{T}^{-}, \mathbf{P}^{\pm}$ transition states obtained from TPS; free energies and stabilities are examined in figure 4. b) Typical transition state structures for both mechanisms, in which the N, C and O atoms involved in the CN bond formation and CO bond cleavage are highlighted; the electrophilic carbon environment is tetrahedral in $\mathbf{T}^{-}$ and almost planar in $\mathbf{P}^{\pm}$.

base catalysis, in contrast with the general base catalysis in mechanism **1**. Figure 2 shows that it is only after the transition state is crossed that mechanism **2** involves the protonation of the newly formed methoxide and the deprotonation of the amine. These two steps occur in a stepwise fashion for most reaction pathways, starting with methoxide protonation, while for a minority of trajectories, they take place in a quasi-concerted manner as shown by the

pathways progressing along the diagonal in the (C_N, C_O) plane in figure 2c.

Some previous studies[22,70,71] have proposed that peptide bond formation may proceed via a protonated carbonyl oxygen intermediate, in a stepwise mechanism. However, this is not supported by our simulations under neutral pH conditions. Our successful TPS pathways did not involve such intermediate (see SI). In addition, we specifically performed complementary simulations in which the carbonyl oxygen was prepared in its protonated form. They show that this form is metastable under neutral pH conditions and does not provide a low-energy pathway for peptide bond formation.

Our results emphasize the importance of proton transfers to and from the solvent water molecules in both mechanisms, involving hydronium or hydroxide ions in the solution. While QM/MM simulations on other systems[34] found that direct intramolecular proton transfers from the nucleophilic amine to the leaving group oxygen atom are the preferred pathway, in our simulations concerted proton transfers between amine and methoxide, occurring either directly via an intramolecular transfer or via a relay by one water molecule, represent only a minority reaction channel. This may arise from the differences between the simulated systems or from the description of all water molecules as reactive in our simulations.

We now characterize the free energy barriers associated with mechanisms **1** and **2**. To this end, we consider the average pathway of each mechanism identified by TPS, determine a collective progress coordinate s for each pathway, and calculate the free energy profile along each reaction coordinate (see Methods). For each mechanism, we determine the free energy barrier, identify the stabilities of the intermediates and monitor the evolution of some selected coordinates. The results are shown in Figure 4.

We first consider the reaction free energy barrier $\Delta G^{\ddagger}$ associated with each mechanism. $\Delta G^{\ddagger}$ governs the reaction rate constant and the relative importance of each pathway. Both barriers are very high with respect to $k_B T$, in agreement with the very slow kinetics of the uncatalyzed amide bond formation. An important result is that mechanism **2** (which had so far not been discussed in the literature) is kinetically much more favorable than mechanism

1 in the neutral pH conditions considered here, since the respective barriers are $\Delta G_1^{\ddagger}$=38 kcal/mol and $\Delta G_2^{\ddagger}$=32 kcal/mol (Figure 4). The reaction along mechanism **2** is thus approximately 20,000 times faster than that along mechanism **1**. The $\Delta G_2^{\ddagger}$ barrier is consistent with experimental rate constants for similar uncatalyzed amino acid condensations. For example, glycine dimerization in water at neutral pH and 1 mol/L concentration yields an approximate free energy barrier of 32 kcal/mol.[72] While this value cannot be directly compared to ours because of the different amino-acids, of the methylated form used in our study, and of the need to correct for the concentration effect in the thermodynamic activity, this shows that our simulations yield the right order of magnitude for the reaction barrier. Our simulated barriers are also in good agreement with calculations performed with a different electronic structure level for a similar system[34] (24.3 kcal/mol at the M06-2X/MM level).

We now consider the formation of the tetrahedral zwitterion $\mathbf{T}^{\pm}$ which is common to both mechanisms. Figure 4 shows that this species is unstable in both mechanisms and only appears as an inflection in the profiles (reaction coordinate $s \simeq 0.36$). This is in agreement with previous calculations[28,34] but this contrasts with the conventional mechanism of figure 1 that assumes a stable intermediate. Our results show that the free energy cost to reach this transient structure is approximately 22 kcal/mol, which thus represents more than half of the overall free energy barrier. The tetrahedral species includes an asymmetric carbon atom and can therefore exist as two different stereoisomers. In our simulations we have considered the nucleophilic attack of the carbonyl by the amine from both sides and we have confirmed that the resulting free energy profiles are identical (see SI). Figure 4 reports the results for the S stereoisomer. The geometry of $\mathbf{T}^{\pm}$ is characterized by the lengths of the newly formed CN bond d_{CN} and of the CO bond to be cleaved d_{CO} which are both smaller than 1.7 Å, while the carbonyl CO bond slightly elongates, in agreement with the transition to a single bond and a negative charge on the carbonyl oxygen (see figure 3a).

We then identify the rate-determining steps and transition state structures of mechanisms **1** and **2**. We first compare the transition state locations obtained from the dynamical

committor analysis of the TPS trajectories (symbols in figure 2 and green region in figure 4, see typical structures in Figure 3b) with the saddle points on the equilibrium free energy surfaces in figure 4. The transition state region is much narrower for mechanism **1** than for mechanism **2**. The agreement between the dynamical and free energy methods is excellent for both $\mathbf{T}^{-}$ and $\mathbf{P}^{\pm}$, which strongly supports our choice of coordinates for the free energy profile calculation.

The equilibrium mechanisms obtained from the minimum free energy paths are also in very good agreement with the dynamical mechanisms determined from the TPS trajectories. In mechanism **1** (figures 3,4), the transition state structure is $\mathbf{T}^{-}$ ($s = 0.5$), which results from the deprotonation of the amine in $\mathbf{T}^{\pm}$ by a proton transfer to a solvent water molecule forming a hydronium ion in solution. The rate-limiting step therefore involves general-base catalysis, here with water acting as the base. The transition state structure has a tetrahedral geometry, with both CN and CO bonds formed, the carbonyl CO bond transiently elongated, and with the central carbon atom in an sp^3 configuration (see Figure 3b). After the transition state is crossed, the methoxy group leaves (as seen by the increase in d_{CO}), and becomes protonated by the hydronium ion (as seen by the changes in the water and methoxy oxygen coordination numbers at $s = 0.66$) to form methanol. While our mechanism **1** shares several features with the conventional mechanism, we emphasize that our free energy profile shows that it occurs in a single concerted step, and that the $\mathbf{T}^{\pm}$ and $\mathbf{T}^{-}$ structures are not stable intermediates.

We now turn to mechanism **2**. In contrast with mechanism **1**, in the transition state geometry $\mathbf{P}^{\pm}$ the CO bond with the leaving group is already cleaved ($d_{CO} \simeq 2.5$ Å at $s = 0.5$) and the central carbon atom lies in a planar trivalent sp^2 configuration that already coincides with the characteristic arrangement of the product peptide bond. The charge separation between the positively charged moiety with the protonated amine and the negatively charged methoxy leaving group leads to a zwitterionic character for the transition state. Here the rate-limiting step involves only the nucleophilic attack and the departure of the leaving group,

but no proton rearrangement (see the constant $C_{O,N}$ between $\mathbf{T}^{\pm}$ and $\mathbf{P}^{\pm}$ in figure 4d) and therefore no acid-base catalysis. While the methoxide oxygen is not involved in any covalent proton rearrangement before and at the transition state (*i.e.*, that within a $\simeq$1.3 Å radius from the methoxide oxygen, no proton is transferred or shared by a donor), it undergoes changes in its hydrogen bonds with the nearby water molecules. Proton transfers only occur after the transition state is crossed. They successively involve the protonation of the leaving group and the deprotonation of the amine (see figure 4b). The free energy profile shows that mechanism **2**, like mechanism **1**, occurs in a single concerted step.

Our comparison of the two mechanisms highlights the importance of including the relevant coordinates to distinguish the different pathways. In addition to the nucleophilic attack coordinate Δd, it is essential to consider hydrogen coordination numbers. Because proton transfers involve the solvent, one should separately consider the C_N and C_O coordination numbers, and not only their difference or only one of them. We note that a recent study [73] of peptide bond formation at the water–silica interface identified mechanism **1** but not mechanism **2**, probably because the C_O coordination was not considered. Since proton exchanges between solute and solvent occur on a slow timescale relative to the accessible simulation times, their sampling is not spontaneous. Therefore, it is critical to include the slow coordination variable explicitly in the reaction path and free energy profile calculations.

We now show that our results are consistent with the large body of experimental data on amide bond formation from isotope effects measurements and from linear free-energy relationships (LFERs), but that they yield a picture of the reaction mechanism and of its change with pH that exhibits critical differences with respect to the conventional image of figure 1. In the conventional mechanism, the change in rate-limiting step with pH is interpreted as indicating the presence of a stable intermediate, proposed to be $\mathbf{T}^{\pm}$, *i.e.*, the mechanism involves the same intermediates at every pH but with successive barriers which depend on pH.[6] In contrast, our results show that it is the preferred reaction pathway and the nature of the intermediates that change with pH, and that each pathway is concerted

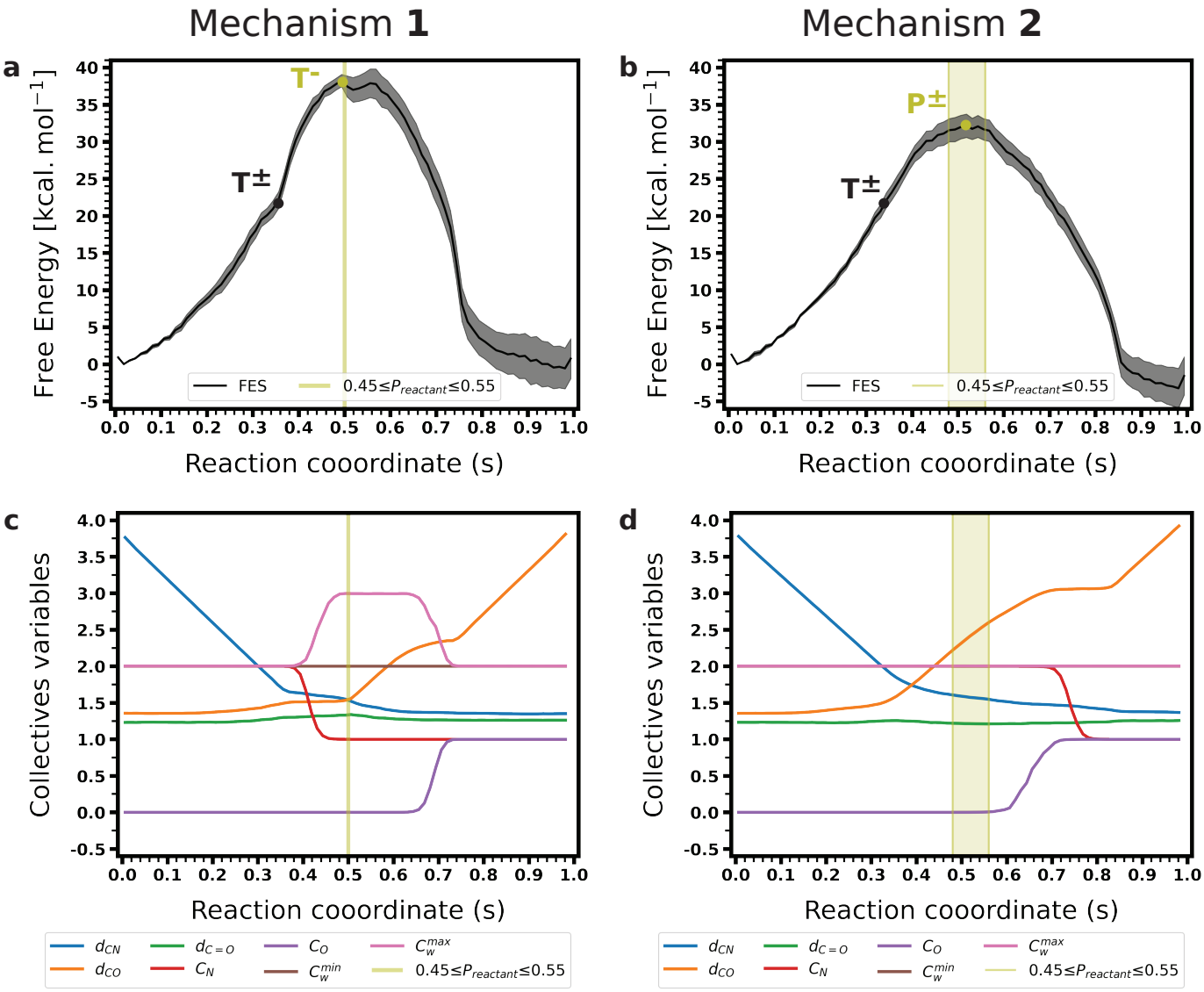


Figure 4: Free energy profiles along the path collective variable obtained via TPS, together with the transition state locations obtained from the committor analysis, respectively for a) mechanism **1** and b) mechanism **2**. Evolution of key distances and coordination numbers along the path collective variable, respectively for c) mechanism **1** and d) mechanism **2**. In addition to the coordinates defined in the text, the figures include the $d_{C=O}$ distance between the carbonyl carbon and oxygen atoms, and the $C_w^{\mathrm{min,max}}$ minimum and maximum hydrogen coordination numbers of all water oxygen atoms. The green vertical lines mark the regions where the reactant committor is 0.50±0.05 and show that this region is much narrower for mechanism **1** than for mechanism **2**.

(figure 5). In contrast to the conventional mechanism assuming a change in the rate-limiting step with pH, our results point to a change in the preferred mechanism with pH. They suggest that the pH-rate profile of peptide bond formation is expected to exhibit a flat region around

neutral pH where the pH-independent mechanism **2** is favored and a positive slope at more basic pH where the base-catalyzed mechanism **1** is dominant. We note that such a pH-rate profile has been measured experimentally for another ester aminolysis.[74]

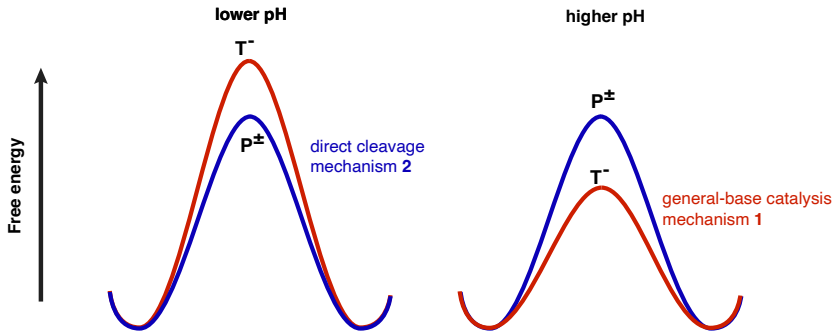


Figure 5: Schematic free energy profiles for the general-base catalysis mechanism **1** and the direct cleavage mechanism **2** at lower and higher pH.

When connecting the experimental results to the mechanisms in the two pH ranges, we particularly focus on two sets of KIE measurements. First, the KIE for the ester oxygen is larger than unity at pH=8 but close to unity at pH=10, which shows that the CO ester bond cleavage is involved in the rate-limiting step at pH=8 but not at pH=10. Second, the KIE for the carbonyl carbon has a large normal value (*i.e.*, >1) in both pH ranges, which prompted questions about the stability of the $\mathbf{T}^{\pm}$ intermediate.[14,16]

At high pH (pH=10), our simulations suggest that mechanism **1** is favored since it involves general-base catalysis (figure 5). Experimentally, LFERs[11] have shown that amine deprotonation is involved in the rate-limiting step. In addition, hydrogen KIE results[14,16] suggest a change of the carbonyl carbon atom hybridization from sp^2 in the reactant to sp^3 in the transition state, and the leaving group O KIE is very close to 1 since $k(^{16}O)/k(^{18}O)$ =1.0048,[75] which shows that cleavage of the CO ester bond occurs after the transition state. All these experimental observations match the rate-limiting $\mathbf{T}^{-}$ formation in our mechanism **1**: the $\mathbf{T}^{-}$ structure is tetrahedral (figure 3), its formation involves the amine deprotonation (figure 4) and the CO bond is still intact.

Regarding the large normal KIE measured[14] for the carbonyl carbon atom $k(^{12}C)/k(^{13}C)$ =

1.02, its origin and compatibility with the conventional mechanism have led to conflicting interpretations. It was first interpreted[14] as showing that $\mathbf{T}^-$ is formed by concerted nucleophilic attack and proton transfer, thus suggesting that $\mathbf{T}^\pm$ is unstable. A subsequent calculation study[15] offered a different interpretation based on CN bond weakening in $\mathbf{T}^\pm$ due to hyperconjugation between the ester oxygen lone pair and the CN^+ antibonding orbital, leading to the conclusion that the experimental KIE is consistent with the stable $\mathbf{T}^\pm$ intermediate proposed in the conventional mechanism. Here, our calculations align with the original interpretation of concerted $\mathbf{T}^-$ formation,[14] since our mechanism **1** involves concerted (but asynchronous) nucleophilic attack and proton transfer steps without going through a stable $\mathbf{T}^\pm$ intermediate. To explain the normal value of the carbonyl carbon KIE despite the change in hybridization from sp^2 to sp^3 between reactant and $\mathbf{T}^-$ transition state, calculations of isotope effects on the equilibrium constants of model reactions (see SI) concur with the hyperconjugation suggestion of ref.[15] and show the weakening of the bonds involving the carbon atom at the transition state. However, these considerations are here applied to the $\mathbf{T}^-$ transition state and not to the $\mathbf{T}^\pm$ intermediate.

At moderate pH (pH=8), KIE studies[75] on the leaving-group oxygen atom exhibit a large normal isotope effect of $k(^{16}O)/k(^{18}O)$ =1.06, suggesting that the CO bond cleavage is involved in the rate-determining step. This has so far been interpreted as arising from the cleavage of $\mathbf{T}^-$ becoming rate limiting. Our calculations suggest a different interpretation. They show that around neutral pH conditions the predominant mechanism is mechanism **2**, whose rate-determining step is indeed the cleavage of the CO bond, but from the sp2 zwitterionic intermediate $\mathbf{P}^\pm$, and not from $\mathbf{T}^-$ (figure 5). KIE measurements[14] reported that for the carbonyl carbon atom the KIE at pH=8 increases with respect to its value at pH=10 to reach $k(^{12}C)/k(^{13}C) = 1.038$. The normal KIE is due to the protonation of the N atom at the transition state that weakens the CN bond (see SI), while the increase observed with respect to pH=10 can be understood by the change in the transition state structure, which is now planar with an sp^2 hybridization of the carbon atom.

Our analysis further provides some indications regarding the reverse reaction, which is amide methanolysis. This reaction and the general case of amide hydrolysis have already been extensively studied by simulations[21,26,29–31,73,76] and by experiments.[77–80] Our calculated free energy profiles in figure 4 show that the peptide bond formation reaction is approximately thermoneutral ($\Delta_r G \simeq 0$) and therefore similar free energy barriers are expected for the condensation and hydrolysis reactions (we incidentally note that our simulations yield $\Delta_r G \simeq 0$ for both mechanisms, which indicates the good convergence of our calculations). Our results suggest that the two pathways revealed here for the condensation reaction will be important for the hydrolysis reaction and for understanding its changes with pH. We emphasize that kinetic analyses of amide hydrolysis[80] have typically assumed a stable tetrahedral intermediate, which is not supported by our calculations.

Finally, we discuss the implications of our results obtained in aqueous conditions for the comprehension of amide and peptide bond formation in other environments of importance. Our simulations show that under close to neutral pH conditions, mechanism **2** is prevalent. Because this mechanism does not involve any proton transfer to reach the transition state and is instead limited by the CO bond cleavage, its catalysis is expected to require the stabilization of the negative charge appearing on the leaving group oxygen atom at the $\mathbf{P}^{\pm}$ transition state. Our results thus suggest that Brønstedt acids are not necessary; instead, Lewis acids are good candidates, as proposed in the synthetic chemistry context to activate the leaving group.[3,4] While the chemical nature of the attacking amine and of the leaving group is expected to influence the barrier heights of each mechanism (*e.g.*, when the amino acid is not methylated and the leaving group is a water molecule instead of methanol), we expect that the presence of the two reaction pathways to be robust.

In the prebiotic context, our mechanism **2** further suggests that peptide bond formation could be catalyzed by non-protic Lewis acids such as cations. It is thus consistent with previous calculations showing the catalytic roles of surface Lewis sites on anatase TiO_2[81] and of Mg^{2+} ions at the surface of clay.[82] Futhermore, our results can explain the exper-

imental observation of salt-induced peptide bond formation[83,84] and the catalytic role of Cu^{2+} cations.[9,83] However, while the catalytic role of Cu^{2+} cations is often considered to arise from a chelation effect, our results suggest that these cations probably also have an important enthalpic impact by stabilizing the nascent negative charge on the leaving group.

For the peptide bond formation catalyzed by the ribosome, in addition to the catalytic effect arising from an entropic factor,[85,86] a particular attention has focused on a possible proton shuttle between the nucleophile proton donor and the leaving group proton acceptor, mediated by bases and water molecules, rather than the general base catalysis mechanism.[6,87–90] Our results are consistent with the key role played by the A76 O_2'H hydrogen-bond donor group identified in the ribosome.[6,87,88,90] Our mechanism **2** involving the $\mathbf{P}^{\pm}$ transition state shows why stabilizing the negative charge on the leaving group oxygen can lower the reaction barrier. More generally, the $\mathbf{P}^{\pm}$ transition state of this mechanism has a strong ion pair character that can be stabilized by electric fields created in the active site of ribosome.[90,91] In solution, the KIE for the reaction of ribosomal substrates[92] are similar to these of the amide bond discussed here, suggesting that they react via the same mechanism **2** in which no proton transfer occurs before the transition state is reached. In the ribosome, our results therefore suggest that in addition to proton relay considerations, the electrostatic stabilization of the leaving group negative charge should also play an important role.

Conclusion

Our study shows that the combination of neural network-based reactive potentials with transition path sampling provides a powerful method to determine reaction mechanisms. Transition path sampling[39,40] has already been successfully applied to a broad range of biophysical transformations and our results show that it is a promising approach to explore complex reactive pathways. Its application to the mechanism of peptide bond formation in aqueous solution reveals two distinct competing reaction mechanisms. The first mechanism

involves the conventional general-base catalysis pathway,[11–13] but our results show that a second mechanism is dominant at neutral pH conditions and that its rate-determining step is the cleavage of the leaving group from the zwitterionic planar intermediate. In contrast with the conventional picture which assumed that changing the pH does not change the reaction pathway but only affects the successive barriers, our results show that two distinct pathways coexist and that their relative importance depends on the pH. We have shown that our proposed mechanisms are consistent with experimental data from kinetics, kinetic isotope effects and linear free energy relationships. The nature of these two mechanisms provides a valuable reference to understand how peptide bond formation can be catalyzed in prebiotic conditions and in the ribosome. Finally, we anticipate that the method introduced here could provide valuable insight into the mechanisms of a wide range of complex organic reactions, particularly where important proton transfers with the solvent are involved (see, *e.g.*, refs. 93,94).

Acknowledgement

We thank Dr M. Bernal Uruchurtu (Universidad Nacional Autónoma de México) and Dr B. Athokpam (ENS) for preliminary studies, and Dr G. Stirnemann (ENS) for discussions. I.T. acknowledges the warm hospitality of the Département de Chimie (École Normale Supérieure, Paris) where he was on sabbatical leave thanks to a 'Salvador de Madariaga' grant of MCIN (Spain). This work was supported by a postdoctoral fellowship to R.D. from PSL OCAV (Idex ANR-10-IDEX-0001-02PSL) and by an HPC allocation from GENCI at IDRIS (Grant 2023-A0130707156).

Supporting Information Available

The following files are available free of charge. Training and validation of neural network potentials, simulation details, definitions of path collective variables, free energy profiles for

the (R)-$\mathbf{T}^{\pm}$ stereoisomer, complementary calculations. The NNP file is available from the authors upon reasonable request.

References

(1) de Figueiredo, R. M.; Suppo, J.-S.; Campagne, J.-M. Nonclassical Routes for Amide Bond Formation. *Chem Rev* **2016**, *116*, 12029–12122.

(2) Carey, J.; Laffan, D.; Thomson, C.; Williams, M. Analysis of the reactions used for the preparation of drug candidate molecules. *Org Biomol Chem* **2006**, *4*, 2337–2347.

(3) Wang, X. Challenges and outlook for catalytic direct amidation reactions. *Nat Catal* **2019**, *2*, 98–102.

(4) de Azambuja, F.; Loosen, A.; Conic, D.; van den Besselaar, M.; Harvey, J. N.; Parac-Vogt, T. N. En Route to a Heterogeneous Catalytic Direct Peptide Bond Formation by Zr-Based Metal–Organic Framework Catalysts. *ACS Catalysis* **2021**, *11*, 7647–7658.

(5) Yang, J.; Huang, H.; Zhao, J. Active Ester-Based Peptide Bond Formation and Its Application in Peptide Synthesis. *Org Chem Front* **2023**, *10*, 1817–1846.

(6) Leung, E.; Suslov, N.; Tuttle, N.; Sengupta, R.; Piccirilli, J. The mechanism of peptidyl transfer catalysis by the ribosome. *Annu Rev Biochem* **2011**, *80*, 527–555.

(7) Lundberg, H.; Tinnis, F.; Selander, N.; Adolfsson, H. Catalytic amide formation from non-activated carboxylic acids and amines. *Chem Soc Rev* **2014**, *43*, 2714–2742.

(8) Sabatini, M. T.; Boulton, Lee. T.; Sneddon, H. F.; Sheppard, T. D. A Green Chemistry Perspective on Catalytic Amide Bond Formation. *Nat Catal* **2019**, *2*, 10–17.

(9) Griffith, E. C.; Vaida, V. In situ observation of peptide bond formation at the water–air interface. *Proc Natl Acad Sci* **2012**, *109*, 15697–15701.

(10) Blackburn, G.; Jencks, W. The mechanism of the aminolysis of methyl formate. *J Am Chem Soc* **1968**, *90*, 2638–2645.

(11) Satterthwait, A. C.; Jencks, W. P. Mechanism of the aminolysis of acetate esters. *J Am Chem Soc* **1974**, *96*, 7018–7031.

(12) Satterthwait, A. C.; Jencks, W. P. Mechanism of partitioning of the intermediates formed in the hydrolysis of phenyl imidates. *J Am Chem Soc* **1974**, *96*, 7031–7044.

(13) Gresser, M.; Jencks, W. Ester aminolysis. Structure-reactivity relationships and the rate-determining step in the aminolysis of substituted diphenyl carbonates. *J Am Chem Soc* **1977**, *99*, 6963–6970.

(14) Marlier, J. F.; Haptonstall, B. A.; Johnson, A. J.; Sacksteder, K. A. Heavy-atom isotope effects on the hydrazinolysis of methyl formate. *J Am Chem Soc* **1997**, *119*, 8838–8842.

(15) Singleton, D. A.; Merrigan, S. R. Resolution of conflicting mechanistic observations in ester aminolysis. A warning on the qualitative prediction of isotope effects for reactive intermediates. *J Am Chem Soc* **2000**, *122*, 11035–11036.

(16) Marlier, J. Multiple isotope effects on the acyl group transfer reactions of amides and esters. *Acc Chem Res* **2001**, *34*, 283–290.

(17) DeTar, D. F. Tetrahedral intermediate in acyl transfer reactions. A revaluation of the significance of rate data used in deriving fundamental linear free energy relationships. *J Am Chem Soc* **1982**, *104*, 7205–7212.

(18) Williams, A. Concerted mechanisms of acyl group transfer reactions in solution. *Acc Chem Res* **1989**, *22*, 387–392.

(19) Stefanidis, D.; Cho, S.; Dhe-Paganon, S.; Jencks, W. P. Structure-reactivity correlations for reactions of substituted phenolate anions with acetate and formate esters. *J Am Chem Soc* **1993**, *115*, 1650–1656.

(20) Hengge, A. C.; Hess, R. A. Concerted or stepwise mechanisms for acyl transfer reactions of p-nitrophenyl acetate? Transition state structures from isotope effects. *J Am Chem Soc* **1994**, *116*, 11256–11263.

(21) Antonczak, S.; Ruiz-Lopez, M.; Rivail, J. Ab initio analysis of water-assisted reaction mechanisms in amide hydrolysis. *J Am Chem Soc* **1994**, *116*, 3912–3921.

(22) Ilieva, S.; Galabov, B.; Musaev, D.; Morokuma, K.; Schaefer, H. Computational study of the aminolysis of esters. The reaction of methylformate with ammonia. *J Org Chem* **2003**, *68*, 1496–1502.

(23) Wu, Z.; Ban, F.; Boyd, R. Modeling the reaction mechanisms of the amide hydrolysis in an N-(o-carboxybenzoyl)-L-amino acid. *J Am Chem Soc* **2003**, *125*, 6994–7000.

(24) Madura, J. D.; Jorgensen, W. L. Ab initio and Monte Carlo calculations for a nucleophilic addition reaction in the gas phase and in aqueous solution. *J Am Chem Soc* **1986**, *108*, 2517–2527.

(25) Antonczak, S.; Monard, G.; Ruiz-López, M. F.; Rivail, J.-L. Modeling of peptide hydrolysis by thermolysin. A semiempirical and QM/MM study. *J Am Chem Soc* **1998**, *120*, 8825–8833.

(26) Bakowies, D.; Kollman, P. A. Theoretical Study of Base-Catalyzed Amide Hydrolysis: Gas- and Aqueous-Phase Hydrolysis of Formamide. *J Am Chem Soc* **1999**, *121*, 5712–5726.

(27) Strajbl, M.; Florian, J.; Warshel, A. Ab Initio Evaluation of the Potential Surface for General Base-Catalyzed Methanolysis of Formamide: A Reference Solution Reaction for Studies of Serine Proteases. *J Am Chem Soc* **2000**, *122*, 5354–5366.

(28) Chalmet, S.; Harb, W.; Ruiz-López, M. F. Computer Simulation of Amide Bond Formation in Aqueous Solution. *J Phys Chem A* **2001**, *105*, 11574–11581.

(29) Zahn, D. On the Role of Water in Amide Hydrolysis. *Eur J Org Chem* **2004**, *2004*, 4020–4023.

(30) Gorb, L.; Asensio, A.; Tuñón, I.; Ruiz-López, M. The mechanism of formamide hydrolysis in water from ab initio calculations and simulations. *Chemistry* **2005**, *11*, 6743–6753.

(31) Blumberger, J.; Ensing, B.; Klein, M. L. Formamide Hydrolysis in Alkaline Aqueous Solution: Insight from Ab Initio Metadynamics Calculations. *Angew Chem Int Ed Engl* **2006**, *118*, 2959–2963.

(32) Schreiner, E.; Nair, N.; Marx, D. Peptide synthesis in aqueous environments: the role of extreme conditions on peptide bond formation and peptide hydrolysis. *J Am Chem Soc* **2009**, *131*, 13668–13675.

(33) Świderek, K.; Tuñón, I.; Martí, S.; Moliner, V.; Bertrán, J. Do zwitterionic species exist in the non-enzymatic peptide bond formation. *Chem Commun* **2012**, *48*, 11253.

(34) Świderek, K.; Tuñón, I.; Martí, S.; Moliner, V.; Bertrán, J. Role of solvent on nonenzymatic peptide bond formation mechanisms and kinetic isotope effects. *J Am Chem Soc* **2013**, *135*, 8708–8719.

(35) Kroonblawd, M.; Pietrucci, F.; Saitta, A.; Goldman, N. Generating Converged Accurate Free Energy Surfaces for Chemical Reactions with a Force-Matched Semiempirical Model. *J Chem Theory Comput* **2018**, *14*, 2207–2218.

(36) Fındık, V.; Ruiz-López, M.; Erdem, S. Mechanistic insights into lysine-targeting covalent inhibition through a theoretical study of ester aminolysis. *Org Biomol Chem* **2021**, *19*, 9996–10004.

(37) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys Rev Lett* **2007**, *98*, 146401.

(38) Wang, H.; Zhang, L.; Han, J.; E, W. DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics. *Comput Phys Commun* **2018**, *228*, 178–184.

(39) Dellago, C.; Bolhuis, P. G.; Geissler, P. L. Transition Path Sampling. *Adv Chem Phys* **2002**, *123*, 1–78.

(40) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu Rev Phys Chem* **2002**, *53*, 291–318.

(41) Bolhuis, P. G.; Swenson, D. W. H. Transition Path Sampling as Markov Chain Monte Carlo of Trajectories: Recent Algorithms, Software, Applications, and Future Outlook. *Adv Theory Simul* **2021**, *4*, 2000237.

(42) Arjun,; Berendsen, T.; Bolhuis, P. Unbiased atomistic insight in the competing nucleation mechanisms of methane hydrates. *Proc Natl Acad Sci U S A* **2019**, *116*, 19305–19310.

(43) Jung, H.; Covino, R.; Arjun, A.; Leitold, C.; Dellago, C.; Bolhuis, P. G.; Hummer, G. Machine-guided path sampling to discover mechanisms of molecular self-organization. *Nat Comput Sci* **2023**, *3*, 334–345.

(44) Dimelow, R.; Bryce, R.; Masters, A.; Hillier, I.; Burton, N. Exploring reaction pathways with transition path and umbrella sampling: application to methyl maltoside. *J Chem Phys* **2006**, *124*, 114113.

(45) Peters, B. Recent advances in transition path sampling: accurate reaction coordinates, likelihood maximisation and diffusive barrier-crossing dynamics. *Mol Sim* **2010**, *36*, 1265–1281.

(46) Zahn, D. Exploring the Mechanisms of Reactions in Solution from Transition Path Sampling Molecular Dynamics Simulations. *J Chem Theory Comput* **2005**, *2*, 107–114.

(47) Geissler, P. L.; Dellago, C.; Chandler, D.; Hutter, J.; Parrinello, M. Autoionization in Liquid Water. *Science* **2001**, *291*, 2121–2124.

(48) Basner, J. E.; Schwartz, S. D. How enzyme dynamics helps catalyze a reaction in atomic detail: a transition path sampling study. *J Am Chem Soc* **2005**, *127*, 13822–13831.

(49) Schwartz, S. D.; Schramm, V. L. Enzymatic transition states and dynamic motion in barrier crossing. *Nat Chem Biol* **2009**, *5*, 551–558.

(50) Tiwari, A.; Ensing, B. Reactive trajectories of the Ru2+/3+ self-exchange reaction and the connection to Marcus' theory. *Faraday Disc* **2016**, *195*, 291–310.

(51) Kühne, T. *et al.* CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J Chem Phys* **2020**, *152*, 194103.

(52) Zeng, J. *et al.* DeePMD-kit v2: A software package for deep potential models. *J Chem Phys* **2023**, *159*, 054801.

(53) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput Phys Commun* **2022**, *271*, 108171.

(54) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput Phys Commun* **2014**, *185*, 604–613.

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

(56) Ma, A.; Dinner, A. Automatic method for identifying reaction coordinates in complex systems. *J Phys Chem B* **2005**, *109*, 6769–6779.

(57) Li, W.; Ma, A. Recent Developments in Methods for Identifying Reaction Coordinates. *Mol Simul* **2014**, *40*, 784–793.

(58) Ketkaew, R.; Luber, S. DeepCV: A Deep Learning Framework for Blind Search of Collective Variables in Expanded Configurational Space. *J Chem Inf Model* **2022**, *62*, 6352–6364.

(59) Chen, H.; Roux, B.; Chipot, C. Discovering Reaction Pathways, Slow Variables, and Committor Probabilities with Machine Learning. *J Chem Theory Comput* **2023**, *19*, 4414–4426.

(60) Zinovjev, K.; Tuñón, I. Adaptive Finite Temperature String Method in Collective Variables. *J Phys Chem A* **2017**, *121*, 9764–9772.

(61) Swenson, D.; Prinz, J.; Noe, F.; Chodera, J.; Bolhuis, P. OpenPathSampling: A Python Framework for Path Sampling Simulations. 1. Basics. *J Chem Theory Comput* **2019**, *15*, 813–836.

(62) Swenson, D.; Prinz, J.; Noe, F.; Chodera, J.; Bolhuis, P. OpenPathSampling: A Python Framework for Path Sampling Simulations. 2. Building and Customizing Path Ensembles and Sample Schemes. *J Chem Theory Comput* **2019**, *15*, 837–856.

(63) Eastman, P.; Swails, J.; Chodera, J.; McGibbon, R.; Zhao, Y.; Beauchamp, K.; Wang, L.; Simmonett, A.; Harrigan, M.; Stern, C.; Wiewiora, R.; Brooks, B.; Pande, V. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* **2017**, *13*, e1005659.

(64) E, W.; Vanden-Eijnden, E. Transition-path theory and path-finding algorithms for the study of rare events. *Annu Rev Phys Chem* **2010**, *61*, 391–420.

(65) Branduardi, D.; Gervasio, F.; Parrinello, M. From A to B in free energy space. *J Chem Phys* **2007**, *126*, 054103.

(66) Brotzakis, Z.; Bolhuis, P. Approximating free energy and committor landscapes in standard transition path sampling using virtual interface exchange. *J Chem Phys* **2019**, *151*, 174111.

(67) Forsythe, J.; Yu, S.; Mamajanov, I.; Grover, M.; Krishnamurthy, R.; Fernández, F.; Hud, N. Ester-Mediated Amide Bond Formation Driven by Wet-Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth. *Angew Chem Int Ed Engl* **2015**, *54*, 9871–9875.

(68) Hay, R.; Porter, L.; Morris, P. The basic hydrolysis of amino acid esters. *Aust J Chem* **1966**, *19*, 1197.

(69) Martin, R. B. Free energies and equilibria of peptide bond hydrolysis and formation. *Biopolymers* **1998**, *45*, 351–353.

(70) Yang, W.; Drueckhammer, D. G. Computational Studies of the Aminolysis of Oxoesters and Thioesters in Aqueous Solution. *Org Lett* **2000**, *2*, 4133–4136.

(71) Charville, H.; Jackson, D. A.; Hodges, G.; Whiting, A.; Wilson, M. R. The Uncatalyzed Direct Amide Formation Reaction - Mechanism Studies and the Key Role of Carboxylic Acid H-Bonding. *Eur J Org Chem* **2011**, *2011*, 5981–5990.

(72) Sakata, K.; Kitadai, N.; Yokoyama, T. Effects of pH and temperature on dimerization rate of glycine: Evaluation of favorable environmental conditions for chemical evolution of life. *Geochim Cosmochim Acta* **2010**, *74*, 6841–6851.

(73) Brigiano, F. S.; Gierada, M.; Tielens, F.; Pietrucci, F. Mechanism and Free-Energy Landscape of Peptide Bond Formation at the Silica–Water Interface. *ACS Catalysis* **2022**, *12*, 2821–2830.

(74) Trmčić, M.; Hodgson, D. Kinetic studies and predictions on the hydrolysis and aminolysis of esters of 2-S-phosphorylacetates. *Beilstein J Org Chem* **2010**, *6*, 732–741.

(75) Sawyer, C. B.; Kirsch, J. F. Kinetic isotope effects for reactions of methyl formate-methoxyl-18O. *J Am Chem Soc* **1973**, *95*, 7375–7381.

(76) Pan, B.; Ricci, M.; Trout, B. A molecular mechanism of hydrolysis of peptide bonds at neutral pH using a model compound. *J Phys Chem B* **2011**, *115*, 5958–5970.

(77) Slebocka-Tilk, H.; Sauriol, F.; Monette, M.; Brown, R. S. Aspects of the hydrolysis of formamide: revisitation of the water reaction and determination of the solvent deuterium kinetic isotope effect in base. *Can J Chem* **2002**, *80*, 1343–1350.

(78) Slebocka-Tilk, H.; Neverov, A.; Brown, R. Proton inventory study of the base-catalyzed hydrolysis of formamide. Consideration of the nucleophilic and general base mechanisms. *J Am Chem Soc* **2003**, *125*, 1851–1858.

(79) Robins, L.; Fogle, E.; Marlier, J. Mechanistic investigations of the hydrolysis of amides, oxoesters and thioesters via kinetic isotope effects and positional isotope exchange. *Biochim Biophys Acta* **2015**, *1854*, 1756–1767.

(80) East, A. L. L. On the hydrolysis mechanisms of amides and peptides. *Int J Chem Kinet* **2018**, *50*, 705–709.

(81) Pantaleone, S.; Ugliengo, P.; Sodupe, M.; Rimola, A. When the Surface Matters: Prebiotic Peptide-Bond Formation on the TiO|2| (101) Anatase Surface through Periodic DFT-D2 Simulations. *Chemistry* **2018**, *24*, 16292–16301.

(82) Martínez-Bachs, B.; Rimola, A. Prebiotic Peptide Bond Formation Through Amino Acid Phosphorylation. Insights from Quantum Chemical Simulations. *Life* **2019**, *9*, 75.

(83) Rode, B. M.; Suwannachot, Y. The possible role of Cu (II) for the origin of life. *Coord Chem Rev* **1999**, *190*, 1085–1099.

(84) Rimola, A.; Rodríguez-Santiago, L.; Ugliengo, P.; Sodupe, M. Is the peptide bond formation activated by Cu(2+) interactions? Insights from density functional calculations. *J Phys Chem B* **2007**, *111*, 5740–5747.

(85) Sievers, A.; Beringer, M.; Rodnina, M. V.; Wolfenden, R. The ribosome as an entropy trap. *Proc Natl Acad Sci* **2004**, *101*, 7897–7901.

(86) Sharma, P.; Xiang, Y.; Kato, M.; Warshel, A. What are the roles of substrate-assisted catalysis and proximity effects in peptide bond formation by the ribosome. *Biochemistry* **2005**, *44*, 11307–11314.

(87) Trobro, S.; Åqvist, J. Mechanism of peptide bond synthesis on the ribosome. *Proc Natl Acad Sci* **2005**, *102*, 12395–12400.

(88) Wallin, G.; Åqvist, J. The transition state for peptide bond formation reveals the ribosome as a water trap. *Proc Natl Acad Sci U S A* **2010**, *107*, 1888–1893.

(89) Hiller, D. A.; Singh, V.; Zhong, M.; Strobel, S. A. A two-step chemical mechanism for ribosome-catalysed peptide bond formation. *Nature* **2011**, *476*, 236–239.

(90) Świderek, K.; Marti, S.; Tuñón, I.; Moliner, V.; Bertran, J. Peptide Bond Formation Mechanism Catalyzed by Ribosome. *J Am Chem Soc* **2015**, *137*, 12024–12034.

(91) Ruiz-Pernía, J. J.; Świderek, K.; Bertran, J.; Moliner, V.; Tuñón, I. Electrostatics as a Guiding Principle in Understanding and Designing Enzymes. *J Chem Theory Comput* **2024**, *20*, 1783–1795.

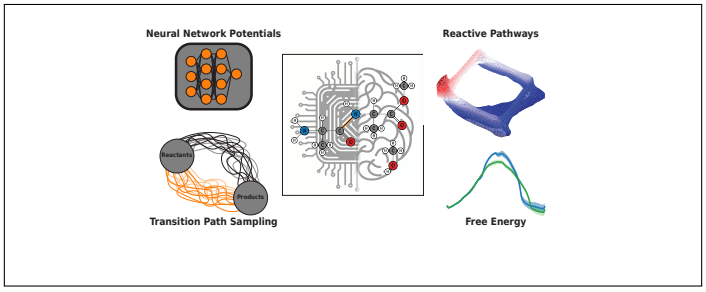
(92) Hiller, D.; Zhong, M.; Singh, V.; Strobel, S. Transition states of uncatalyzed hydrolysis and aminolysis reactions of a ribosomal P-site substrate determined by kinetic isotope effects. *Biochemistry* **2010**, *49*, 3868–3878.

(93) Plata, R.; Singleton, D. A case study of the mechanism of alcohol-mediated Morita

Baylis-Hillman reactions. The importance of experimental observations. *J Am Chem Soc* **2015**, *137*, 3811–3826.

(94) Liu, Z.; Patel, C.; Harvey, J.; Sunoj, R. Mechanism and reactivity in the Morita-Baylis-Hillman reaction: the challenge of accurate computations. *Phys Chem Chem Phys* **2017**, *19*, 30647–30657.

Graphical TOC Entry



Supplementary Information

The competing reaction mechanisms of peptide bond formation in water revealed by deep potential molecular dynamics and path sampling

Rolf David,[†] Iñaki Tuñón,[‡] and Damien Laage*,[†]

†*PASTEUR, Department of Chemistry, École Normale Supérieure, PSL University, Sorbonne Université, CNRS, 75005 Paris, France*

‡*Departamento de Química Física, Universitat de Valencia, Burjassot, Valencia 46100, Spain*

E-mail: damien.laage@ens.psl.eu

Contents

Simulated system

Our simulated system consists of two methyl L-alaninates solvated with 300 water in a $21.046\,\text{Å}^3$ cubic box in the reactant state, and of one methyl L-alanyl-L-alaninate and one methanol solvated with 300 water in a $21.046\,\text{Å}^3$ cubic box in the product state. Typical configurations in the reactant and product states are shown in Figure S1.

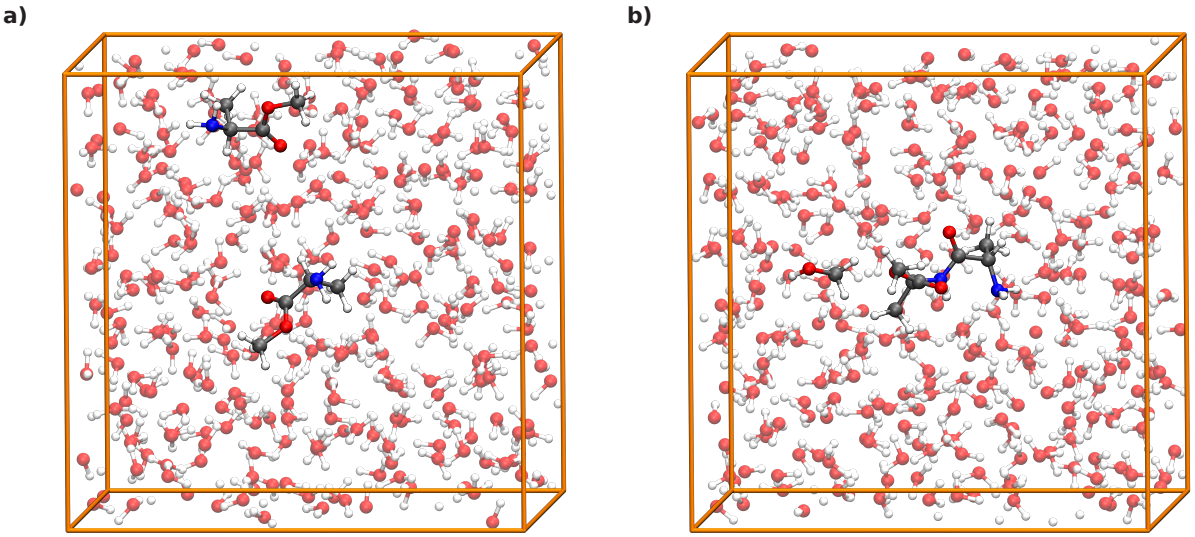


Figure S1: Simulation boxes where the molecules are represented in a ball-and-stick format where oxygen atoms are in red, nitrogen atoms in blue, carbon atoms in grey, and hydrogen atoms in white. a) Reactant state. b) Product state.

Collective variables

The system is described with the following collective variables (CVs).

The CVs first include a set of distances:

- d_{CN}: distance between the ester electrophilic carbon atom and the nucleophilic amine nitrogen atom
- d_{CO}: distance between the ester electrophilic carbon atom and the leaving methyl ester oxygen atom
- $d_{CO_{c=O}}$: distance between the ester electrophilic carbon atom (C) and the carbonyl oxygen atom ($O_{C=O}$)
- $\Delta d = d_{CN}$ - d_{CO}

They further include coordination numbers:

- C_N: coordination number in labile hydrogen around the nucleophilic amine nitrogen atom
- C_O: coordination number in labile hydrogen around the leaving methyl ester oxygen atom
- $C_{O_{C=O}}$: coordination number in labile hydrogen around the carbonyl oxygen atom
- $\Delta C = C_N$ - C_O
- $\sum C = C_N + C_O$
- C_w^{min}: minimum coordination in labile hydrogen around water oxygen atoms O_w, with $C_w^{min} \simeq 1$ when a hydroxide is present in solution and $C_w^{min} \simeq 2$ otherwise
- C_w^{max}: maximum coordination in labile hydrogen around water oxygen atoms O_w, with $C_w^{max} \simeq 3$ when a hydronium is present in solution and $C_w^{max} \simeq 2$ otherwise

The coordination number C between heavy atoms and labile hydrogens is defined as

$$\sum_A \sum_H C_A = \frac{1 - \left(\frac{r_{A-H}}{r_0}\right)^n}{1 - \left(\frac{r_{A-H}}{r_0}\right)^m},$$

where for solute heavy atoms ($A = N, O, O_{C=O}$), r_0=1.3 Å, $n = 14$, $m = 24$, and for water oxygen atoms ($A = O_w$), r_0 =1.3 Å, $n = 50$, $m = 10$.

Finally, our CVs include two angles

- ω: peptide bond torsion angle in the product state defined between the alpha carbon (C^α), the carbonyl carbon (C), the nitrogen of the nucleophile (N_{nuc}), and the alpha carbon (C^α) of the nucleophilic amino acid; ω discriminates between the *trans* and *cis* forms of the dipeptide

- θ: nucleophilic attack angle for the nucleophilic amine group on the electrophilic ester carbon (see figure S2); θ is the angle between the normal vector of the plane ($O_{C=O}CO$) formed by the $CO_{C=O}$ bond and the CO_{me} bond and the vector formed by the $C^\alpha N_{nuc}$ bond; θ discriminates between the two stereoisomers of the tetrahedral $\mathbf{T}^\pm$ zwitterion, with $-\frac{\pi}{2}$ rad$< \theta <\frac{\pi}{2}$ rad for *(R)*-$\mathbf{T}^\pm$ and $\frac{\pi}{2}$ rad$< \theta <\frac{3\pi}{2}$ rad for *(S)*-$\mathbf{T}^\pm$

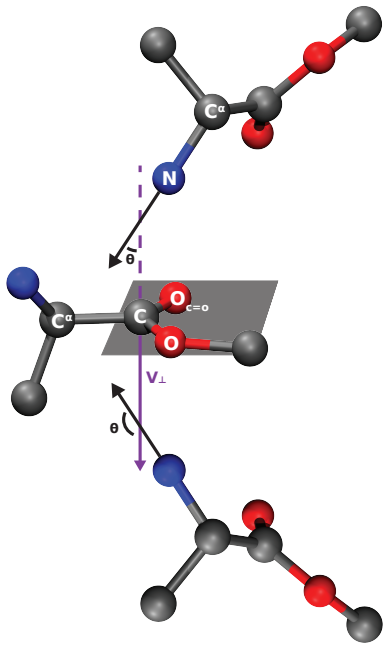


Figure S2: Representation of the θ angle and of the two possible angles of attack of the nucleophile amine group on the ester carbon. Top: θ close to 0 rad, indicating an attack of the nucleophile forming the R tetrahedral ($T^{\pm}$) zwitterion. Bottom: θ close to π rad indicating an attack of the nucleophile forming the S tetrahedral ($T^{\pm}$) zwitterion. The violet arrow represents the normal vector to the $O_{C=O}CO$ plane and the black arrow represents the vector formed by the $C^{\alpha}N_{nuc}$ bond.

Neural Network Potentials

Preparation of the initial training dataset

To obtain an initial dataset including positions, energies and forces for training the NNPs, several unbiased molecular dynamics simulations were performed at the DFTB2/mio-1.1 [1,2] level of theory using CP2K.[3] The trajectories were propagated in the NVT ensemble, with a 0.5 fs timestep and the temperature set to 300 K by a CSVR thermostat [4] with a damping parameter of $100\,\mathrm{ps}^{-1}$. Starting in the reactant state and in the product state arranged in a $21.046\,\text{Å}^3$ cubic box, trajectories were propagated for 10 to 400 ps. Furthermore, multiple exploratory metadynamics [5] were performed to obtain reactive structures (*e.g.*, transition states and intermediates). The systems, starting from both reactant and product states, were biased using the Δd collective variable. The hills were added every 25 fs with a height of $0.5\,\mathrm{kcal}\cdot\mathrm{mol}^{-1}$ and a width of 0.05 Å to ensure fast exploration of the energy lanscape from reactants to products and from products to reactants. From all these simulations, 802 structures were extracted from the unbiased reactant trajectories, another 802 from the unbiased product trajectories, and 1004 from the biased reactive trajectories. These 2602 structures were labeled with the method of reference, namely BLYP [6,7] with dispertion corrections D3,[8] using TZV2P basis set [9] and GTH pseudo-potentials,[10,11] to create the initial training dataset (*i.e.* to get the total energy and atomic forces for each structure). (While the BLYP-D3 description remains approximate, it offers a compromise beytween computational cost and chemical accuracy and it has been extensively used in ab initio molecular dynamics; it would be straightforward to relabel our training set at a higher level like the hybrid B3LYP-D3 functional.)

Training of the Neural Network Potentials

The initial dataset was used to train three NNPs using DeePMD-kit,[12] version 2.0.3, with different random seeds, to obtain a committee of NNPs for concurrent learning. The DeepPot-

SE scheme[13] was employed and the cutoff for the radial and angular information was set to 6 Å with a cosine weight function for atoms beyond 0.5 Å. The energy was minimized using the Adam optimizer[14] with an initial learning rate (α_0) equal to 0.001 and a final learning rate (α_{final}) equal to 1×10^{-6} over 600 000 steps. The full training parameters are summarized in Table S1.

Table S1: Training Parameters

Parameters	Iterative Training	Final (Production) Training
Initial learning rate (α_0)	0.001	0.001
Final learning rate (α_{final})	1e-6	1e-8
Number of training steps (t_{final})	600000	5700000
Number of neurons in the embedding network	[25, 50, 100]	[25, 50, 100]
Cutoff (Å)	6	6
Smoothing cutoff (Å)	0.5	0.5
Initial energy loss prefactor (p_e^0)	0.02	0.02
Final energy loss prefactor (p_e^∞)	1	1
Initial force loss prefactor (p_f^0)	1000	1000
Final force loss prefactor (p_f^∞)	1	1
Number of neurons in the fitting network	[240, 240, 240, 240]	[240, 240, 240, 240]

The training set was expanded iteratively, following an approach inspired by the DP-GEN procedure.[15] Sequences of training, exploration, and labeling phases were performed until very few candidates were added. The initial exploration phase consisted of two sets of molecular dynamics simulations, starting respectively from the reactant state and from the product state. For each set, each of the three NNPs was used as the reactive force field, culminating in six simulations. They were propagated in the NVT ensemble, each for 10 ps with a temperature of 300 K using the CVSR thermostat and a 0.5 fs timestep. For each trajectory, a query-by-committee selection was performed: candidate geometries were selected based on an error indicator, ϵ_t, defined as the maximal standard deviation of the atomic force predicted by the NNPs. A lower bound, σ_{low}, and an upper bound, σ_{high}, were set respectively to $0.15\,\mathrm{eV} \cdot \mathrm{Å}^{-1}$ and $0.80\,\mathrm{eV} \cdot \mathrm{Å}^{-1}$. Structures for which $\epsilon_t < \sigma_{low}$ were considered as adequately described by the NNPs and those where $\epsilon_t > \sigma_{high}$ were considered as unphysical and were discarded. The remaining structures, where $0.15 < \epsilon_t <$

0.80 eV/Å were labeled with the method of reference (as per the initial training dataset), BLYP-D3/TZV2P, hereby enriching the training dataset.

In iteration 4, the initial set (obtained from DFTB2 geometries) was removed from the training dataset due to differences between the DFTB2 and BLYP-D3 equilibrium distributions. In iteration 6, several biased molecular dynamics were performed using PLUMED[16] to enhance the exploration of the phase space and sample the transition region between the reactant and product basins. In iteration 14, two-dimensions umbrella sampling simulations along Δd and ΔC were performed and the query-by-committee was performed on each window trajectories.

In total, MD explorations were propageted for 10 ps to 1000 ps and biased MD explorations were propagated for 10 ps to 1000 ps.

To avoid any bias in the training favoring a particular mechanism, a large number of different parameters (*e.g.*, strength of restraints for SMD, Gaussian deposition speed in metadynamics), and various collective variables (*e.g.*, distances, coordination numbers, combinations of multiple CVs) were used. Furthermore, from approximate transition states estimated from these biased molecular dynamics (as metadynamics and umbrella sampling can be readily used to calculate the underlying free energy surface), multiple committor-type calculations were performed. Approximate transition state geometries were selected as starting points, and for each of them 5 sets of initial velocities drawn from a Maxwell-Boltzmann distribution at 300 K were created, together with 5 additional ones with the opposite velocities. These were used to propagate 10 unbiased molecular dynamics simulations in the NVT ensemble from each starting point, generating more candidate structures to label and to add to the training dataset. After the total number of candidates was less than ten for unbiased exploration and less than two hundred for biased explorations, the iterative training was considered to be completed.

It was after a total of 19 iterations with a final set of 76 656 structures, out of which 7693 were obtained from unbiased exploration (*i.e.* reactant and product states) and 68 963 were

obtained from biased explorations. This final set was used as the training set and used to train three new NNPs using the production parameters (see Table S1).

Validation of the Neural Network Potentials

A validation dataset was constructed by randomly selecting 450 structures in the reactant state, 450 structures in the product state, and 2940 structures obtained from Umbrella Sampling windows. This validation dataset, which consisted of a total of 3840 structures with their energies and atomic forces, was used to test the accuracy and reliability of the trained NNPs. After calculating the energy and forces of each structure in the training and validation datasets with the three NNPs and with our reference method, the best NNP was selected based on the lowest deviation in energy and forces on both sets, and kept as the production NNP. It is worth mentioning that the other two NNPs showed highly comparable deviations, indicating the completeness of the training by the fact that they all converged to the same parameters. Figure S3 shows the correlation plots between the predicted forces (NNPs) and the reference forces (BLYP-D3) for the [-2.0 eV/Å, +2.0 eV/Å] range for each dataset.

Accuracy of the Neural Network Potentials along the reaction paths

After the pathways for mechanisms **1** and **2** were identified, we *a posterio* verified the accuracy of the NNPs along the paths. For both paths 10 structures uniformely selected within each umbrella sampling window, totaling 1290 structures for mechanism 1 and 1340 structures for mechanism 2.

The atomic forces were calculated using the NNPs and the reference method (BLYP-D3) for each structure. The Root Mean Square Error (RMSE) of the predicted atomic forces relative to the standard deviation of the reference atomic forces is reported in Figure S4. It is $0.038\,87\,\mathrm{eV}\cdot\mathrm{Å}^{-1}$ for mechanism **1** and $0.039\,21\,\mathrm{eV}\cdot\mathrm{Å}^{-1}$ for mechanism **2**, thus confirming

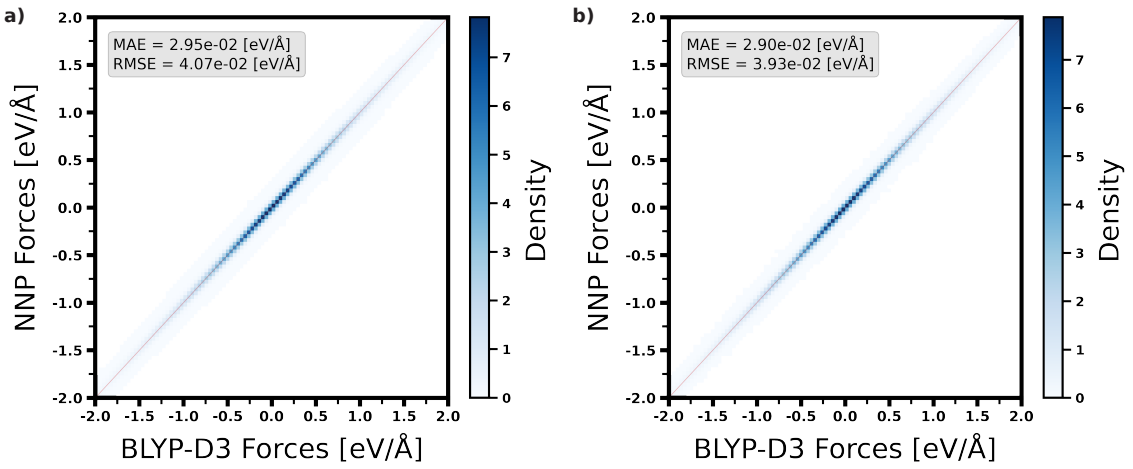


Figure S3: a) Correlations between the predicted forces (NNP) and reference forces (BLYP-D3) for the training set, zoomed over the $[-2.0, 2.0]$ eV/Å range. b) Correlations between the predicted forces (NNP) and reference forces (BLYP-D3) for the validation set, zoomed over the $[-2.0, 2.0]$ eV/Å range.

that the atomic forces are very well predicted by the NNPs for both mechanisms.

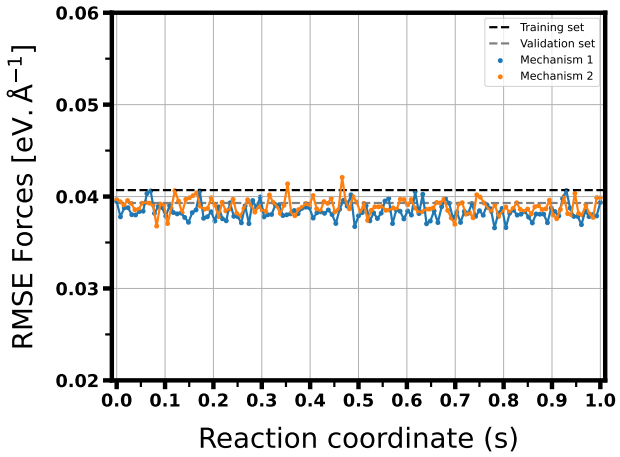


Figure S4: Root Mean Square Error (RMSE) on the predicted atomic forces relative to the standard deviation of the reference atomic forces for each window along the path for both mechanisms.

Transition Path Sampling

Initial reactive trajectories

To generate the initial reactive trajectories required by Transition Path Sampling[17–19] (TPS), several steered molecular dynamics simulations (SMDs) were performed. These simulations were started from the reactant and steered to the product, using a range of collective variables. They covered different possible mechanisms (mechanisms **1** and **2** notably, but also the concerted mechanism and the stepwise mechanism through the protonation of the $O_{C=O}$ suggested in ref.[20]) which had been explored *via* the biased simulations during the NNPs training. From these SMD trajectories, a committor-type analysis was conducted on configurations that were geometrically close to putative transition states. Subsequently, non biased NVT molecular dynamics simulations were carried out at 300 K using a CVSR thermostat and a timestep of 0.5 fs. These simulations involved random initial velocities drawn from a Maxwell-Boltzmann distribution, along with their opposite velocity counterparts, for a committor-type analysis. Candidates were selected from the configurations in which half of the trajectories led to the reactant well and the other half to the product well. From these, seven full trajectories were constructed, each comprising segments from the aforementioned half-trajectories and were then used as input for the TPS simulations.

Transition Path Sampling simulations

The TPS simulations were performed using the OpenPathSampling (OPS) program,[21,22] in conjunction with OpenMM,[23] which was integrated with an in-house modified plugin, enabling the calculation of forces using the NNPs within OpenMM, and with the PLUMED package, facilitating the use of a wide range of collective variables.

The reactant and product state basins were defined with the Δd and ΔC collective variables. The reactant well was defined by $\Delta d > 1.5$ Å and $\Delta C > 1.75$. The product well was defined by $\Delta d < -1.5$ Å and $\Delta C < 0$.

For the TPS simulations, the VVVR-Langevin integrator[24] with a collision rate of $20\,\mathrm{ps}^{-1}$ was used, alongside a timestep of 0.5 fs and a target temperature of 300 K. Starting from the seven initial trajectories, 74 TPS equilibration runs were conducted using the One-Way Shooting (OWS) move.[25] Equilibration was deemed successful when a new trajectory added to the transition path ensemble had no frames in common with the initial trajectory, i.e., was fully decorrelated. 70 of the 74 equilibration runs were successful and the last 4 were discarded due to the lack of decorrelation within 400 Monte-Carlo steps. The last trajectories (70 in total) from each TPS equilibrated run, now independent from the initial ones, along with their corresponding velocities, were used as input for the production phase of the TPS simulations. The same integrator parameters were maintained, but with the collision rate adjusted to $1\,\mathrm{ps}^{-1}$. From these, a total of 30 independent runs were performed using the OWS move, another 26 using the Two-Way Shooting (TWS) move,[17,18] and finally, 14 using the Spring Shooting (SS) move,[26] each consisting on average of 8800 Monte-Carlo steps. MC trials were halted if they did not reach any basin within 5 ps. For the TWS simulations, random velocities were generated from a Maxwell-Boltzmann distribution at 300 K, and for the SS simulations, the parameters were set as $k = 1.0$ and $\delta = 10$.

Analysis of the Transition Path Sampling simulations

Table S2: TPS Simulations (production runs)

Details	
MC Steps	616003
Total time [ns]	543
Accepted Steps	268106
Acceptance Rate %	43.5
Decorrelated Steps	65602
Rejected Steps	347897
Rejected Steps Truncated	35043

Statistics of the TPS simulations are shown in Table S2. From this entire ensemble of trajectories, since both the OWS move and the SS move provide only half the trajectory,

subsequents analysis were only done on decorrelated trajectories (*i.e.*, on trajectories which did not share any frame with each other) for a total of 65 602 independent trajectories. The distribution of the path lengths for the decorrelated trajectories is reported in Figure S5. They show that the two mechanisms have similar path lengths distributions, with a mean of 0.88 ps for mechanism **1** and 0.60 ps for mechanism **2**. The slightly longer mean path length for mechanism **1** is due to the involvement of a H_3O^+ ion which can diffuse away before protonating the leaving group, while in mechanism **2** protonation of the leaving group is typically done by a water molecule, which relays the proton from the amine group to the methoxide.

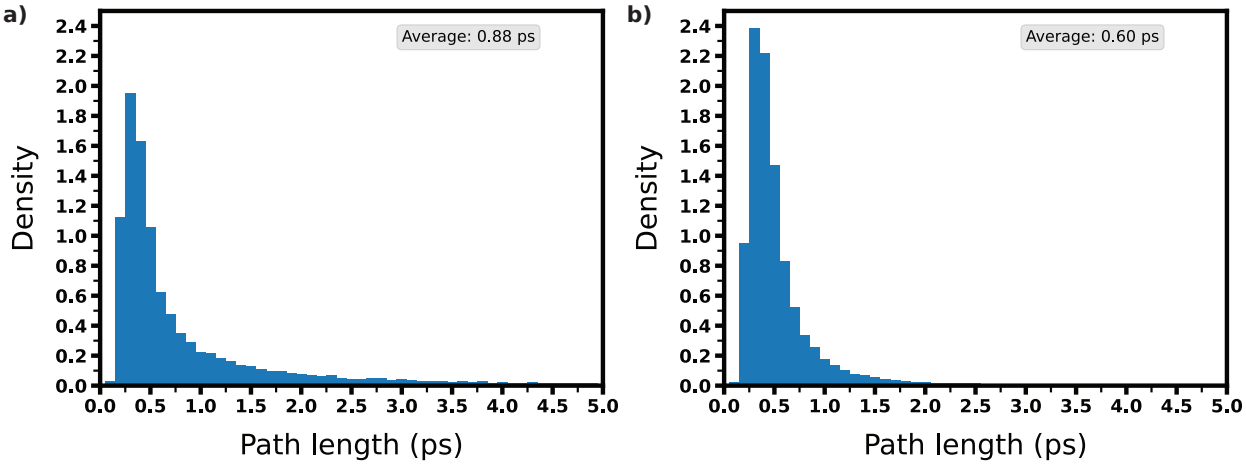


Figure S5: Distributions of the path lengths for the decorrelated trajectories obtained from the TPS simulations. a) Mechanism **1**. b) Mechanism **2**.

The transition path ensemble was extracted along three main coordinates: Δd, $C_{N_{nuc}}$, and $C_{O_{me}}$. To achieve this, the CV space was divided into bins, and a count was added each time a bin is visited by a trajectory. If a trajectory visited a bin more than once or if it stayed in a bin for multiple timesteps, the count for this trajectory and for this bin was incremented only once. If the trajectory hops between non-adjacent bins (due to our finite time step), interpolation is used to increment the count to the intermediate bins. In figure 2 in the main text, the transition path ensemble for each mechanism is represented in the Δd,

C_N, and C_O space, showing the contour surface of the 0.0015 probability. Although other mechanisms from literature (e.g. from ref.[20]) were considered as starting trajectories, they quickly evolved into the two mechanisms that we propose in this work.

Figure S6 shows the projected transition path density on the $(\Delta d, C_N)$ plane and on the $(\Delta d, C_O)$ plane for mechanism **1** and for mechanism **2**. It shows that for mechanism **1**, the reaction path first evolves along Δd and C_N, then along Δd and C_O. In contrast, for mechanism **2**, the reaction path first evolves along Δd and C_O, then along Δd and C_N.

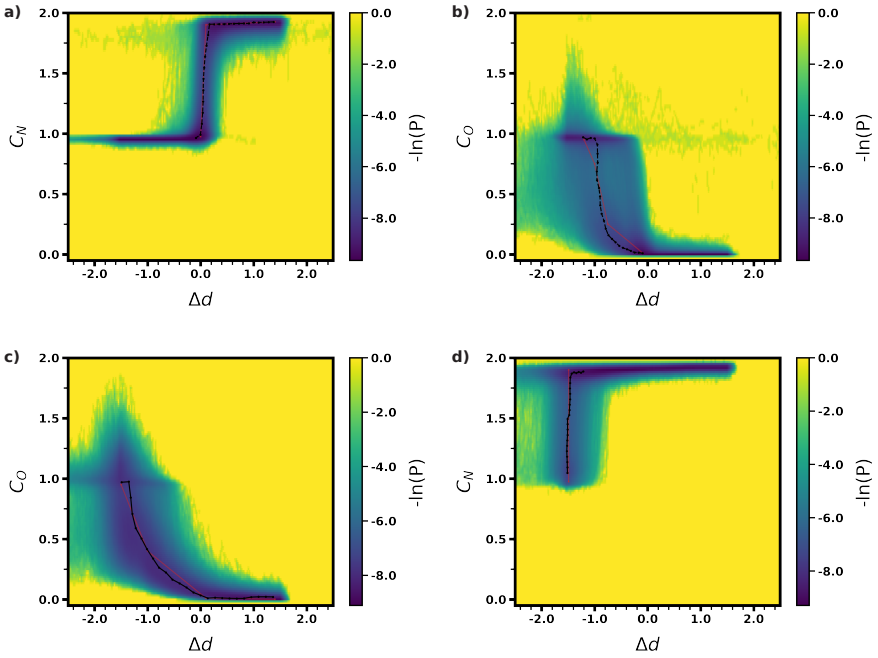


Figure S6: Negative log density $-\ln(p)$ of the transition path ensemble projected on two collective variables with the initial guess string (red solid line) and the final string (black solid line) for each mechanism. Trajectories of mechanism **1** a) in the $(\Delta d, C_N)$ plane and b) in the $(\Delta d, C_O)$ plane. Trajectories of mechanism **2** c) in the $(\Delta d, C_N)$ plane and d) in the $(\Delta d, C_O)$ plane.

For each half-path, the string method[27,28] (using 25 nodes) was used within the (Δd,C_N, C_O) space to extract the most probable path using the gradient descent method on the negative log density. An initial string was created with equally spaced nodes and each node was moved by following the gradient descent of the negative log density with a step of 0.01. Then, the nodes were reparametrized by equal arc lengths. The process was repeated until a converged string was obtained. Trailing points were added (30 for mechanism **1**,

and 23 for mechanism **2**) to extend the Δd range from $-2.5\,\text{Å}$ to $2.5\,\text{Å}$. Both paths were then reparametrized using equal arc-length parametrization. For both mechanisms, the 20 closest configurations to each node were extracted from the transition path ensemble, with each configurations coming from independent TPS production runs. A new path, comprising 80 nodes, was then constructed from these configurations using four CVs averaged over the 20 configurations: $d_{CN_{nuc}}$, $d_{CO_{me}}$, $C_{N_{nuc}}$, and $C_{O_{me}}$. Then again, equal arc-length parametrization was used for both mechanisms to have 40 nodes. For both paths (consisting of 4 CVs), a path CV[29] was constructed to represent the progress along the path in s space (represented in Figure S7).

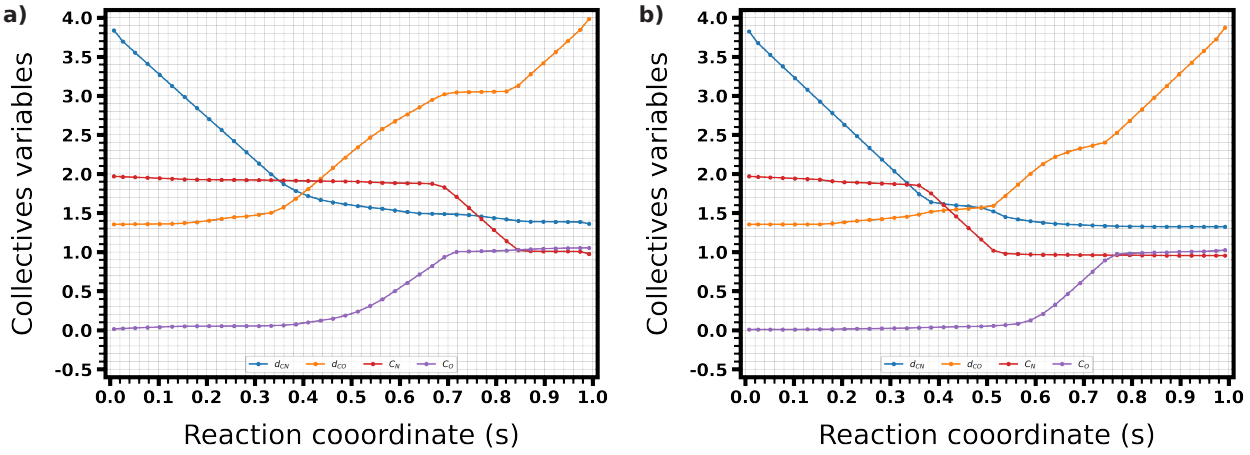


Figure S7: The 4 CVs (d_{CN}, d_{CO}, C_N and C_O) of the two mechanisms as a function of s. a) Mechanism **1**. b) Mechanism **2**.

The collective variables were analyzed from the TPS simulations for each mechanism as follows. First, the s values along each decorrelated trajectory were calculated. The time origin was fixed to the point in each trajectory where the s value is nearest to 0.5. Given the varying lengths of the trajectories, they were uniformly extended to cover a range from -9 ps to 9 ps by duplicating the end values of the CVs at both extremes. Subsequently, for every time step, the average and the 95% confidence interval were computed for each CV, for both mechanisms and are represented in Figure S8. In both mechanisms, we can see the importance of the solvent, as for mechanism **1**, the increase of the $C_w^{\max}$ coordination number

from 2 to 3 is a clear indication of the proton transfer via a long-lived hydronium ion. For mechanism **2**, the small decrease in the $C_w^{\min}$ coordination number from 2 to 1.8 indicates a proton transfer via a water molecule, forming a transient hydroxide ion.

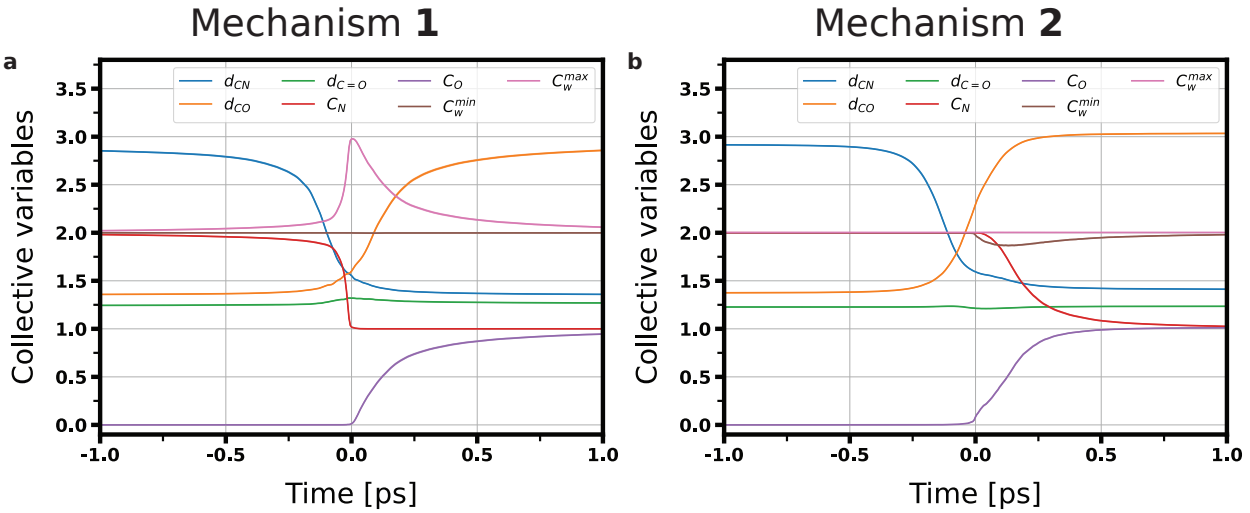


Figure S8: Average of the CVs for the decorrelated trajectories obtained from the TPS simulations. a) Mechanism **1**. b) Mechanism **2**.

Free Energy calculations

Four initial trajectories were selected from the TPS ensemble, one for each mechanism and each stereochemistry (R or S) of the tetrahedral $\mathbf{T}^{\pm}$. For both mechanisms, umbrella sampling along s was performed at a temperature of 300 K using a CSVR thermostat. For mechanism **1**, the path was linearly spaced into 119 windows from $s = 0.0074$ to $s = 0.9926$. An additional 5 windows were linearly added between $s = 0.7233$ and $s = 0.7640$, and 5 more between $s = 0.3711$ and $s = 0.4057$, totaling 129 windows. For mechanism **2**, the path was linearly spaced into 119 windows from $s = 0.0074$ to s =0.9923. An additional 4 windows were linearly added between $s = 0.7297$ and $s = 0.7563$, 6 between $s = 0.3632$ and $s = 0.4059$, and 6 between $s = 0.8000$ and $s = 0.8416$, totaling 134 windows.

For each window, the starting points were brought to the corresponding values of the CVs (d_{CN}, d_{CO}, C_N, and C_O) *via* Steered MD with a harmonic constraint with a force constant of 500 kcal $\cdot$ mol^{-1} and a timestep of 0.25 fs for 50 ps. Then the force constant was linearly increased from 500 kcal $\cdot$ mol^{-1} to 1000 kcal $\cdot$ mol^{-1} over another 50 ps. Afterward, equilibration was performed by switching the constraint to the corresponding s value with a 50 kcal $\cdot$ mol^{-1} force constant, a timestep of 0.10 fs. The z transverse coordinate was constrained to a value of -0.0012 with a force constant of 10 000 kcal $\cdot$ mol^{-1}. These equilibration trajectories were propagated for 20 ps. For the production, two sets of constraints were used: one with a force constant of 10 kcal $\cdot$ mol^{-1} and a second one with a force constant of 50 kcal $\cdot$ mol^{-1}. Each one was propagated for 40 ps. In Figure S9, the s distributions of the windows for each mechanism and each stereochemistry show their good overlap.

For each system (*i.e.*, each mechanism and each stereochemistry), the free energy was calculated using the weighted histogram analysis method (WHAM).[30,31] Each 40 ps-long production run was divided into 4 blocks of 10 ps and the free energy was calculated for each block. 95% confidence intervals on the computed free energy were calculated for each block by setting the free energy at 0 kcal $\cdot$ mol^{-1} at $s = 0.018\,75$. The free energy profile for a value of θ=0 rad was calculated and is reported in Figure S10.

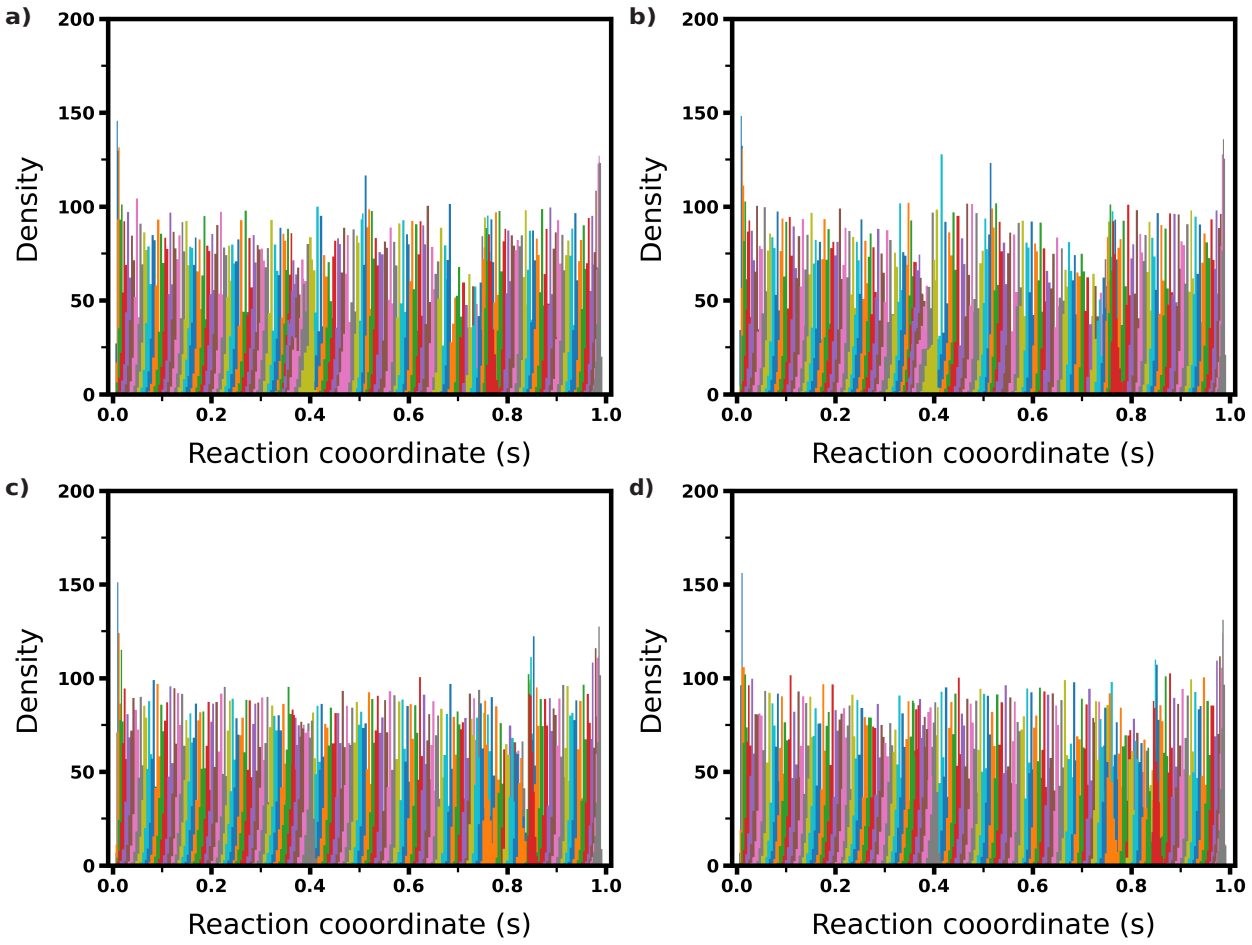


Figure S9: s distributions of the windows for each mechanism and each stereochemistry. a) Mechanism **1** and R-stereochemistry. b) Mechanism **1** and S-stereochemistry. c) Mechanism **2** and R-stereochemistry. d) Mechanism **2** and S-stereochemistry.

In table S3, the reaction free energy between the reactant and the product state and the free energy barrier between the reactant and the transition state are reported for each mechanism and each stereochemistry.

The free energy barrier for mechanism **1** is higher than that for mechanism **2**. The

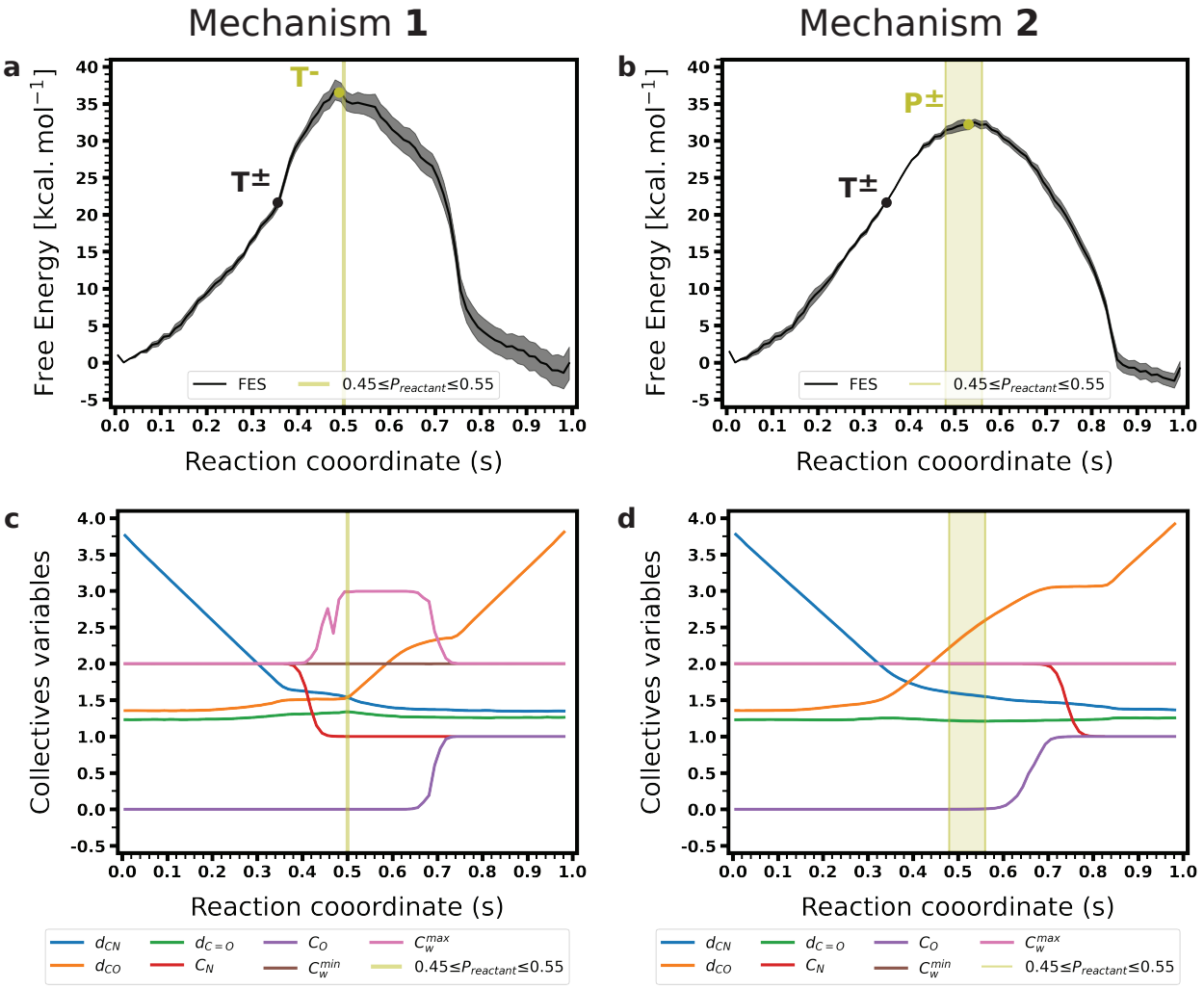


Figure S10: Free energy profiles along the path collective variable obtained via TPS, together with the transition state locations obtained from the committor analysis, respectively for a) mechanism **1** and b) mechanism **2** obtained for the (S)-$\mathbf{T}^{\pm}$ stereoisomer (θ=0). Evolution of key distances and coordination numbers along the path collective variable, respectively for c) mechanism **1** and d) mechanism **2**. In addition to the coordinates defined in the text, the figures include the $d_{C=O}$ distance between the carbonyl carbon and oxygen atoms, and the $C_w^{\mathrm{min,max}}$ minimum and maximum hydrogen coordination numbers of all water oxygen atoms. The green vertical lines show the regions where the reactant committor is 0.50±0.05.

results in Table S3 further confirm the good convergence of our calculations. As expected, the reaction free energy is the same for both mechanisms within the confidence interval, and the free energy barrier is independent of the stereochemistry.

Table S3: Reaction free energy and free energy barrier.

Mechanism	Stereochemistry	$\Delta_r G$ [kcal/mol]	$\Delta G^{\ddagger}$ [kcal/mol]
1	R	-1.39 ± 2.17	36.90 ± 1.36
1	S	-0.57 ± 2.76	38.43 ± 1.08
2	R	-2.47 ± 0.98	32.49 ± 0.45
2	S	-3.25 ± 2.68	32.16 ± 1.58

Isotope Effects

Table S4 reports the calculated carbon isotope effect on the equilibrium constant of a series of reactions that model the change between the peptide bond formation reactants and $\mathbf{T}^-$: the carbonyl carbon atom changes from sp^2 to sp^3 hybridizations and the products exhibit the type of hyperconjugation found in $\mathbf{T}^-$. The results show that the change from sp^2 to sp^3 yields the expected inverse isotope effect, except when the carbonyl atom is negatively charged and a normal isotope effect is obtained. This can be understood by the weakening and lengthening of the carbon bonds. Calculations were done with Gaussian at the B3LYP/6-31G* level at T=298.15 K and P=1 atm.

Table S4: Isotope effect on the equilibrium constant $K_L/K_H = \exp\left[-(\Delta G_L - \Delta G_H)/RT\right]$ for a carbon atom whose hybridization changes from sp^2 to sp^3, with and without a charged carbonyl oxygen atom.

Reaction	$K(^{12}C)/K(^{13}C)$	length of new C-X bond in products (Å)
$H_2C{=}O + OH^- \rightarrow H_2C(O^-)\text{-}OH$	1.004	1.554
$H_2C{=}O + H_2O \rightarrow H_2C(OH)\text{-}OH$	0.992	1.407
$H_2C{=}O + NH_4^+ \rightarrow H_2C(OH)\text{-}NH_3^+$	1.000	1.567
$H_2C{=}O + NH_3 \rightarrow H_2C(OH)\text{-}NH_2$	0.994	1.444

Table S5 reports the calculated carbon isotope effect on the equilibrium constant of a series of reactions that model the change between the peptide bond formation reactants and $\mathbf{P}^{\pm}$: the carbonyl carbon atom keeps its sp^2 hybridization and the products exhibit the type of hyperconjugation found in $\mathbf{P}^{\pm}$. The results show that when the N atom is not charged in the product the isotope effect is very close to 1, whereas when the N atom is positively charged in the product, a normal isotope effect is obtained, which can be understood by the weakening and lengthening of the C-N bond. Calculations were done with Gaussian at the B3LYP/6-31G* level at T=298.15 K and P=1 atm.

Table S5: Isotope effect on the equilibrium constant $K_L/K_H = \exp\left[-(\Delta G_L - \Delta G_H)/RT\right]$ for a carbon atom which retains its sp^2 hybridization, with different charges on the N atom.

Reaction	$K(^{12}C)/K(^{13}C)$	length of new C-N bond in products (Å)
HCO-OH + NH_3 → H_2CO-NH_2 + H_2O	1.001	1.362
HCO-OH + NH_3 → H_2CO-NH_3^+ + OH^-	1.024	1.584

References

(1) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Suhai, S.; Seifert, G.; Frauenheim, Th. *Phys. Rev. B* **1998**, *58*, 7260–7268.

(2) Elstner, M.; Seifert, G. *Phil. Trans. R. Soc. A.* **2014**, *372*, 20120483.

(3) Kühne, T. D. et al. *J. Chem. Phys.* **2020**, *152*, 194103.

(4) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

(5) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.

(6) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(7) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

(8) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

(9) VandeVondele, J.; Hutter, J. *J. Chem. Phys.* **2007**, *127*, 114105.

(10) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703–1710.

(11) Hartwigsen, C.; Goedecker, S.; Hutter, J. *Phys. Rev. B* **1998**, *58*, 3641–3662.

(12) Wang, H.; Zhang, L.; Han, J.; E, W. *Comput. Phys. Commun.* **2018**, *228*, 178–184.

(13) Zhang, L.; Han, J.; Wang, H.; Saidi, W. A.; Car, R.; E, W. *NeurIPS* **2018**, *2018-Decem*, 4436–4446.

(14) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. 2017.

(15) Zhang, Y.; Wang, H.; Chen, W.; Zeng, J.; Zhang, L.; Wang, H.; E, W. *Comput. Phys. Commun.* **2020**, *253*, 107206.

(16) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

(17) Bolhuis, P. G.; Dellago, C.; Chandler, D. *Faraday Disc.* **1998**, *110*, 421–436.

(18) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. *J. Chem. Phys.* **1998**, *108*, 1964–1977.

(19) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.

(20) Świderek, K.; Tuñón, I.; Martí, S.; Moliner, V.; Bertrán, J. *J. Am. Chem. Soc.* **2013**, *135*, 8708–8719.

(21) Swenson, D. W. H.; Prinz, J.-H.; Noe, F.; Chodera, J. D.; Bolhuis, P. G. *J. Chem. Theory Comput.* **2019**, *15*, 813–836.

(22) Swenson, D. W. H.; Prinz, J.-H.; Noe, F.; Chodera, J. D.; Bolhuis, P. G. *J. Chem. Theory Comput.* **2019**, *15*, 837–856.

(23) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. *PLoS Comput. Biol.* **2017**, *13*, e1005659.

(24) Sivak, D. A.; Chodera, J. D.; Crooks, G. E. *J. Phys. Chem. B* **2014**, *118*, 6466–6474.

(25) Bolhuis, P. G. *J. Phys.: Condens. Matter* **2003**, *15*, S113–S120.

(26) Brotzakis, Z. F.; Bolhuis, P. G. *J. Chem. Phys.* **2016**, *145*, 164112.

(27) E, W.; Ren, W.; Vanden-Eijnden, E. *Phys. Rev. B* **2002**, *66*, 052301.

(28) E, W.; Ren, W.; Vanden-Eijnden, E. *J. Chem. Phys.* **2007**, *126*, 164103.

(29) Branduardi, D.; Gervasio, F. L.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 054103.

(30) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(31) Grossfield, A. WHAM: The Weighted Histogram Analysis Method. 2023.