UvA-DARE (Digital Academic Repository)

Advances in enhanced sampling along adaptive paths of collective variables

Pérez de Alba Ortíz, A.; Tiwari, A.; Puthenkalathil, R.C.; Ensing, B.

DOI

[10.1063/1.5027392](https://doi.org/10.1063/1.5027392)

Publication date

2018

Document Version

Final published version

Published in

Journal of Chemical Physics

License

Article 25fa Dutch Copyright Act

[Link to publication](#)

Citation for published version (APA):

Pérez de Alba Ortíz, A., Tiwari, A., Puthenkalathil, R. C., & Ensing, B. (2018). Advances in enhanced sampling along adaptive paths of collective variables. *Journal of Chemical Physics*, 149(7), [072320]. <https://doi.org/10.1063/1.5027392>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Advances in enhanced sampling along adaptive paths of collective variables

Cite as: J. Chem. Phys. **149**, 072320 (2018); <https://doi.org/10.1063/1.5027392>

Submitted: 01 March 2018 . Accepted: 25 May 2018 . Published Online: 13 June 2018

A. Pérez de Alba Ortíz , A. Tiwari, R. C. Puthenkalathil, and B. Ensing

View Online

Export Citation

CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[Frequency adaptive metadynamics for the calculation of rare-event kinetics](#)

The Journal of Chemical Physics **149**, 072309 (2018); <https://doi.org/10.1063/1.5024679>

[Reweighted autoencoded variational Bayes for enhanced sampling \(RAVE\)](#)

The Journal of Chemical Physics **149**, 072301 (2018); <https://doi.org/10.1063/1.5025487>

[Collective variable discovery and enhanced sampling using autoencoders: Innovations in network architecture and error function design](#)

The Journal of Chemical Physics **149**, 072312 (2018); <https://doi.org/10.1063/1.5023804>

PHYSICS TODAY
WHITEPAPERS

ADVANCED LIGHT CURE ADHESIVES

Take a closer look at what these environmentally friendly adhesive systems can do

READ NOW

PRESENTED BY
MASTERBOND
ADHESIVES | SEALANTS | COATINGS

Advances in enhanced sampling along adaptive paths of collective variables

A. Pérez de Alba Ortíz,^{1,2,a)} A. Tiwari,^{1,2,b)} R. C. Puthenkalathil,^{1,2,c)} and B. Ensing^{1,2,d)}

¹*Van 't Hoff Institute for Molecular Sciences, Universiteit van Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands*

²*Amsterdam Center for Multiscale Modeling, Science Park 904, 1098 XH Amsterdam, The Netherlands*

(Received 1 March 2018; accepted 25 May 2018; published online 13 June 2018)

Study of complex activated molecular transitions by molecular dynamics (MD) simulation can be a daunting task, especially when little knowledge is available on the reaction coordinate describing the mechanism of the process. Here, we assess the path-metadynamics enhanced sampling approach in combination with force field and *ab initio* [density functional theory (DFT)] MD simulations of conformational and chemical transitions that require three or more collective variables (CVs) to describe the processes. We show that the method efficiently localizes the average transition path of each process and simultaneously obtains the free energy profile along the path. The new multiple-walker implementation greatly speeds-up the calculation, with an almost trivial scaling of the number of parallel replicas. Increasing the dimensionality by expanding the set of CVs leads to a less than linear increase in the computational cost, as shown by applying the method to a conformational change in increasingly longer polyproline peptides. Combined with DFT-MD to model acid (de-)protonation in explicit water solvent, the transition path and associated free energy profile were obtained in less than 100 ps of simulation. A final application to hydrogen fuel production catalyzed by a hydrogenase enzyme showcases the unique mechanistic insight and chemical understanding that can be obtained from the average transition path. *Published by AIP Publishing.* <https://doi.org/10.1063/1.5027392>

I. INTRODUCTION

Path-metadynamics¹ (PMD) is an effective means to enhance the sampling of activated molecular transitions in a molecular dynamics (MD) simulation. The probability to observe an activated process, such as a chemical reaction, a conformational change, or a phase transition, during an unbiased MD simulation, decreases exponentially with the height of the activation free energy, or transition state barrier, of the process. This so-called *rare event* problem has led to the development of a plethora of enhanced sampling techniques that can roughly be subdivided into three classes: (1) those that bias the dynamics with an external potential or force applied to a geometric descriptor, or collective variable (CV), of the process (e.g., Refs. 2–6), (2) those that focus on locating reactive trajectories or optimized transition paths (e.g., Refs. 7–13), and (3) those that enhance the dynamics by raising the temperature of (part of) the system (e.g., Refs. 14–16). PMD combines ideas from the first two classes.

The metadynamics method⁴ belongs to the first class of enhanced sampling methods. It constructs during an MD simulation a repulsive bias potential that “fills” free energy minima such that barriers in the free energy can be overcome and the reaction free energy and the free energy barrier can be computed *a posteriori* from the final metadynamics bias potential.

A difficulty with “class one” approaches can be the requirement of choosing an adequate CV that functions as the *reaction coordinate*. An imperfect choice for the CV is one that omits essential degrees of freedom for the description of the transition, which typically leads to poor convergence and hysteresis problems for the free energy estimation. Metadynamics has the advantage that it can handle several CVs simultaneously in a trivial manner, which alleviates somewhat the issue of knowing *a priori* the reaction coordinate; instead, only an appropriate set of CVs is needed. The result is then an insightful multidimensional free energy landscape, in which the (meta-)stable states and the connecting minimum free energy paths (MFEPs) can be located.¹⁷ In practice, however, the number of CVs is limited to about three for any but the simplest¹⁸ of applications due to the unfavorable scaling of the computational cost with the dimensionality.

To deal with complex transitions that require many CVs, the PMD method was developed. Here, an extra “path-CV” is introduced, which is a parameterized curve that connects the stable states and is defined in the space of the other CVs that describe the transition. Since *a priori* the transition path is not known, one starts with a guess path, which is then optimized during the PMD simulation to the average transition path. The key advantage is that the metadynamics biasing is now performed on the one-dimensional progress parameter along the path so that, in principle, a large set of CVs can be used.

Since its introduction, the PMD method has evolved with several new features and with the implementation of the algorithm in the freely available PLUMED¹⁹ software so that

^{a)}Electronic mail: a.perezdealbaortiz@uva.nl

^{b)}Electronic mail: a.tiwari@uva.nl

^{c)}Electronic mail: r.c.puthenkalathil@uva.nl

^{d)}Electronic mail: b.ensing@uva.nl

it can be used with various popular MD codes. The aim of this work is to assess the performance of PMD to find the average transition path and obtain the associated free energy profile for complex molecular transitions. We will do this at the hand of three illustrative case studies.

In the first case study, we introduce the multiple walker version of PMD, which parallelizes the path optimization algorithm, and apply it to simulate the conformational transition in polyproline peptides from a right-handed helix to a left-handed helix. By systematically increasing the polypeptide length, we investigate the PMD performance with increasing numbers of dihedral angle CVs.

In the second case study, we apply PMD to investigate the deprotonation of acetic acid in aqueous solution. This is a particularly challenging test for any enhanced sampling approach, not in the least because such a chemical rearrangement in explicit solvent requires in every step a computationally costly evaluation of the energy and forces. Here we use density functional theory (DFT) to obtain the forces on the atoms from an electronic structure calculation. This demands that the path and the free energy be converged on a simulation time scale of only several tens of picoseconds.

In the last case study, we focus on the unique physical and chemical insight that can be obtained from the optimized transition path, at the hand of another chemical transition: formation of molecular hydrogen by proton reduction, catalyzed by a di-iron hydrogenase enzyme. Apart from obtaining detailed mechanistic understanding, we will use the path to scrutinize the changes in the electronic structure during the reaction.

In the following, we first briefly summarize the theory behind the PMD algorithm, before presenting the case studies. We end with a compilation of the main conclusions derived from the three case studies.

II. THEORETICAL BACKGROUND

In this section, we recapitulate the path finding method originally introduced in 2012.¹ Let us consider a molecular system whose free energy surface (FES) can be fully described by a set of N key descriptive degrees of freedom: the collective variables (CVs) $\{z_i(\vec{q})\}$ with $i = 1, \dots, N$, which are functions of the system's particle positions $\vec{q}(t)$. Let us also assume that these particle positions $\vec{q}(t)$ evolve in time with a canonical equilibrium distribution at temperature T under a potential $V(\vec{q})$. We are interested in finding the average transition path connecting two FES basins—the stable states A and B —in the space spanned by the CVs $\{z_i\}$. Given that the CVs are sufficiently good descriptors of the transition, iso-committor hyperplanes provide a continuous foliation of the CV-space between A and B . In this space, we define the transition flux density ρ as the number of A -to- B trajectories—initialized with random Maxwell-Boltzmann distributed velocities $\vec{v}(t)$ —going through the iso-committor surfaces per unit area. The flux density ρ peaks at the transition channel in the FES connecting A and B , resulting in the average transition path we aim to find. Three assumptions underlie our methodology:¹

1. The average transition path can be represented in the CV-space by a path-CV: a curve $\vec{s}(\sigma) : \mathbb{R} \rightarrow \mathbb{R}^N$, where the parameter $\sigma(\vec{z})|_s : \mathbb{R}^N \rightarrow [0, 1]$ yields the progress along the path from A to B , such that $\vec{s}(0) \in A$ and $\vec{s}(1) \in B$. This quantity is a well-defined reaction coordinate and can, in principle, be connected to the committor value.
2. In the vicinity of the path, the iso-committor planes S_σ are perpendicular to $\vec{s}(\sigma)$.
3. In the vicinity of the path, the normalized transition flux density ρ can be represented by the configurational probability $p(\vec{z}) = \exp(-F(\vec{z})/k_B T)$, where F is the free energy and k_B is the Boltzmann constant.

The first and second assumptions allow us to derive the progress parameter $\sigma(\vec{z})$ for any point $\vec{z}$ in CV-space by projecting it onto the closest point on the path $\vec{s}(\sigma)$. The third assumption enables us to adapt the path-CV $\vec{s}(\sigma)$ to the average CV density,

$$s(\sigma) = \int_{S_\sigma} dS_\sigma \vec{z}' p_\sigma(z'_1, \dots, z'_N),$$

with

$$p(z'_1, \dots, z'_N) = \frac{1}{Z} \int d\vec{q} e^{-\beta U(\vec{q})} \delta(z_1 - z'_1) \dots \delta(z_N - z'_N), \quad (1)$$

where $p_\sigma(z'_1, \dots, z'_N)$ is the probability density at the iso-committor surface perpendicular to $\vec{s}(\sigma)$, with the Dirac delta function δ and the partition function Z . In practice, the ensemble average of sampled points $\vec{z}$ at the hyperplanes is replaced by a time average,

$$\langle \vec{z} \rangle_{\sigma_g} = \lim_{t \rightarrow \infty} \int_0^t \int_{S_{\sigma_g}} \vec{z}(t') dS_{\sigma_g} dt'. \quad (2)$$

Since the CV density at each hyperplane peaks in the transition channel, we can converge the average transition path by iteratively adapting a guess path to the cumulative average density $\vec{s}_g(\sigma_g) = \langle \vec{z} \rangle_{\sigma_g}$ (see Fig. 1).

Numerically, the adaptive path is implemented as a set of M ordered nodes (coordinates in CV-space) $\vec{s}_g(\sigma_g, t) \rightarrow \vec{s}_j^{t_i}$ with $j = 1, \dots, M$, where t_i is the discrete time of path updates. The projection of any point $\vec{z}$ onto the path is done considering the closest two nodes, and σ is obtained by interpolation.¹ It should be noted that this approach requires equidistant nodes, which are imposed by a reparametrization step⁸ after each path update. Sec. III deals with the side-effects of this reparametrization and how to correct for them.

FIG. 1. Graphical representation of a path-CV converging to the average transition path between basins A and B . The path curve $\vec{s}_g(\sigma_g)$ is adapted to the cumulative average density $\langle \vec{z} \rangle_{\sigma_g}$ in CV-space, which peaks at the transition valley in the free energy F at each hyperplane S_{σ_g} .

The update step for the path nodes is done by

$$\vec{s}_j^{t_{i+1}} = \vec{s}_j^{t_i} + \frac{\sum_k w_k \cdot (\vec{z}_k - \vec{s}^{t_i}(\sigma(\vec{z}_k)))}{\sum_k \xi^{t_i-k} w_{j,k}}$$

with

$$w_k = \max \left[0, \left(1 - \frac{\|\vec{s}_j^{t_i} - \vec{s}^{t_i}(\sigma(\vec{z}_k))\|}{\|\vec{s}_j^{t_i} - \vec{s}_{j+1}^{t_i}\|} \right) \right], \quad (3)$$

where k is the current MD step and w is the weight of the adjustment, which is non-zero only for the two nodes closest to the average CV density. We also introduce the fade factor $\xi = \exp(-\ln(2)/\tau)$, with a half-life τ the number of MD steps after which the distances measured from the path weight only 50% of their original value. A short half-life increases path flexibility by allowing it to rapidly “forget” a bad initial guess, while a long half-life restricts the path fluctuations and leads eventually to optimal convergence. The first and last nodes, located at the stable states A and B , respectively, remain fixed. Extra trailing nodes can be placed at both ends of the path to better capture the free energy valleys. Wall potentials can be used to keep the sampling near the relevant $[0,1]$ σ -interval.

In principle, the presented recipe can localize the average transition path simply by sampling $\vec{z}$ during an MD run. However, since the transition is a rare event, the barrier region separating the A and B basins will be poorly sampled. This issue is typically addressed in enhanced sampling methods by biasing $\vec{z}$ directly, but such an approach is, in practice, limited to low dimensional CV-spaces. The path formalism frees from this restriction as we bias the one-dimensional progress parameter $\sigma_g = \sigma(\vec{z})|_{\vec{s}_g}$ along a guess transition path.

Here, metadynamics⁴ is our method of choice to sample the transition. We exert a history-dependent repulsive potential,

$$V_{bias}(\sigma_g, t) = \sum_t H(t) \exp \left(\frac{-(\sigma_g - \sigma_g(t))^2}{2W^2} \right), \quad (4)$$

with Gaussian height H and width W . This bias drives the system along the path, allowing for sufficient sampling of the CV transition density and localization of the average transition path. Simultaneously, the metadynamics algorithm continuously self-corrects¹⁷ and converges to the free energy along the average transition path $V_{bias}(\sigma, t \rightarrow \infty) = -F(\sigma)$.

We refer to the use of metadynamics in combination with a path-CV as “path-metadynamics.” Our algorithm can benefit from most of the machinery already available for the well-established metadynamics sampling technique. For example, the well-tempered method²⁰ can aid in the convergence of free energy profiles by gradually reducing the Gaussian height. The multiple-walker implementation²¹ can accelerate the sampling by having multiple simulations that explore the same path and FES in parallel. In Sec. III, we showcase the benefits of the multiple walkers. Naturally, the path-CV framework is not exclusive to metadynamics. Other techniques—such as umbrella sampling (US), steered MD (SMD), or constrained MD (CMD)^{2,3,22}—can be straightforwardly used to sample along, and optimize, the path.

Finally, when exploring a FES with multiple transition channels, one generally aims to converge to the average transition path of a particular channel (connected to a specific reaction mechanism), rather than to the global average over all transition channels. In such a case, it can be advantageous to bias not only the sampling along the path but also in the directions perpendicular to it. A harmonic restraint set on $\|\vec{z}_k - \vec{s}^{t_i}(\sigma(\vec{z}_k))\|$ can help in converging to a specific transition path. We refer to this restraint as a “tube” potential. In the limit of an infinitesimally narrow tube potential, the path update step follows the local free energy gradient—similar to the string method⁸—and PMD converges to the MFEP instead of the average transition path (see also the supplementary material in Ref. 1 for an illustrative comparison). This control over the behavior of the algorithm is an advantageous feature of our method. In well-defined landscapes with a single channel, the path can adapt quickly and without restraints to follow the CV density through the transition channel. In more complex surfaces, a tube potential can be added and the path can be gradually optimized to the closest transition channel from the initial guess. We wish to highlight that the updating rules are the key distinguishing aspect between different path-CV approaches^{7,8,10–13} and that the use of the CV density in PMD provides the method with considerable versatility as it can perform either as a density-driven or as a gradient-driven scheme.

III. CASE STUDY I: PPI-TO-PPII TRANSITION IN PROLINE OLIGOMERS

Polyproline chains, Ace-(Pro) $_n$ -Nme, form two characteristic helical structures: polyproline type I (PPI), a right-handed helix with all residues in the *cis* form and backbone dihedral angles ($\phi = -75^\circ$, $\psi = 160^\circ$, $\omega = 0^\circ$) and polyproline type II (PPII), a left-handed helix with all residues in the *trans* form and backbone dihedral angles ($\phi = -75^\circ$, $\psi = 146^\circ$, $\omega = 180^\circ$). See Fig. 2. The proline’s cyclized side-chain considerably restricts the conformational space on the Ramachandran (ϕ , ψ) plane, such that stable and intermediate peptide conformations are mainly determined by the ω dihedral angles. Thus, a proline n -mer can be well described by a set of CVs $\{\omega_i\}$

FIG. 2. (a) PPI right-handed helix structure with backbone dihedral angles ($\phi = -75^\circ$, $\psi = 160^\circ$, $\omega = 0^\circ$). (b) PPII left-handed helix structure with backbone dihedral angles ($\phi = -75^\circ$, $\psi = 146^\circ$, $\omega = 180^\circ$).

with $i = 1, \dots, n - 1$. This provides an interesting test system for enhanced sampling techniques as the dimensionality of the problem can be tuned by adding or removing residues. Moreover, the system has biological relevance as proline switches are involved in protein signalling. We demonstrate the imperviousness of the path-CV to the increase in dimensionality by finding the average pathways and associated free energy profiles for the PPI-PPII transition for tetrameric, pentameric, and hexameric polyproline.

The transition between PPI and PPII has been extensively computationally studied by Moradi and co-workers^{23–29} via various enhanced sampling methods, including adaptively biased MD (ABMD),³⁰ Hamiltonian and temperature replica exchange MD (HT-REMD),³¹ and steered MD (SMD)³ from PPII to PPI and vice versa. As biasing all $\{\omega_i\}$ dihedrals proves inefficient, more compressed CVs were built based on summations and other functions of $\{\cos(\omega_i)\}$, as well as the helix handedness and radius of gyration. ABMD and HT-REMD were used to obtain free energy surfaces along such CVs. SMD was used to compute the work along different candidate pathways on the calculated surfaces, from which the reaction rates could be computed via an extension of Crook's transient fluctuation theorem. Simulations were carried out *in vacuo*, as well as in implicit and explicit water and other solvents. The preference for either PPI or PPII was observed to depend on the solvent environment and the chain length. Interestingly, the transition was found to occur following zipper-like mechanisms. That is, with adjacent residues rotating in sequential order, each representing one of $n - 1$ free energy barriers. Two main pathways were reported: (1) a zipper-like mechanism starting from the acetylated end when going from PPI to PPII, and from the amidated end when going from PPII to PPI, with intermediate states of the form Ace-TTT...CCC-Nme (C = *cis*, T = *trans*); and (2) a zipper-like mechanism starting from both ends of the chain when going from PPI to PPII, and from the center of the chain when going from PPII to PPI, with intermediate states of the form Ace-TTT...CCC...TTT-Nme. Oddly, some of the SMD results show a different preferred mechanism depending on the direction of the transition. Recently, the experimental studies of Shi and co-workers^{32,33} have confirmed the zipper-like mechanism starting from the acetylated end for the PPI-to-PPII transition. We refer to these previous results for comparison.

In our study, we construct proline oligomers using AmberTools17,³⁴ which produces PPII structures. Simulations are carried out using GROMACS 5.1.4³⁵ patched with PLUMED 2.3.0¹⁹ with the added PMD code available on: <http://www.acmm.nl/ensing/software/PathCV.cpp>. We use the AMBER force field ff99SB³⁶ and no solvent. The time step size is 1 fs and a canonical sampling through velocity rescaling (CSVR) thermostat³⁷ is set at 300 K with a time constant of 0.1 ps.

The path for each proline n -mer is defined in a CV-space composed of all dihedrals $\{\omega_i\}$ with $i = 1, \dots, n - 1$. Since our path implementation currently does not support periodic CVs, we consider only one direction for each residue rotation at a time. For the clockwise rotation (from 0° to 180° through positive angle values), we add -90° to all $\{\omega_i\}$ and map the transition from PPI $\{\omega_i\} = -90^\circ$ to PPII $\{\omega_i\}$

$= +90^\circ$. For the counter-clockwise rotation (from 0° to -180° through negative angle values), we add $+90^\circ$ to all $\{\omega_i\}$ and map the transition from PPI $\{\omega_i\} = +90^\circ$ to PPII $\{\omega_i\} = -90^\circ$. Whenever a particular transition direction is being sampled, the other one is blocked by harmonic walls with force constants of 1000 kcal/(mol rad²), which restrain the sampling inside the $[-172^\circ, +172^\circ]$ interval. In this work, we only test transitions in which all residues rotate in the same direction. We primarily focus on the counter-clockwise rotation, which provides results consistent with the previous literature.^{23–26,32,33}

Paths are described by 20 nodes per dihedral plus 20 trailing nodes at each end, giving a total of $20(n + 1)$ nodes for each proline n -mer. For all PMD simulations, the initial guess is a linear interpolation from PPI (path progress parameter $\sigma = 0$) to PPII ($\sigma = 1$). The path update pace is set to 0.5 ps and the half-life parameter to 2.0 ps. A tube potential is applied with a force constant of 50 kcal/(mol rad²). Harmonic walls with force constants of 1000 kcal/mol per squared normalized path unit are set on σ to limit the sampling inside the $[-0.1, +1.1]$ interval. Accordingly, the metadynamics force is set to zero outside the same interval. Since the range of σ is known *a priori*, it is easy and memory-efficient to use a grid to store the potential. The PLUMED framework facilitates all these measures.

We start with a single walker PMD simulation of the proline tetramer. The path is set for the counter-clockwise transition. Gaussians with a height of 0.05 kcal/mol and a width of 0.05 normalized path units are deposited every 1.0 ps. The system escapes from the starting PPI minimum and follows a step-wise transition to PPII. As the residues rotate one at a time, the path consists of successive orthogonal segments in CV-space. Each section captures the crossing over a barrier and each “corner” the location of a metastable state. During the simulation, we observe problems while attempting to converge a path curve. As the system drifts away from each metastable state, leaving it momentarily unsampled, the reparametrization algorithm makes the nodes tend to a straight line and undoes the previous optimization (see Fig. 3). This “corner-cutting” problem persists until the metadynamics bias has filled all minima and the system diffuses quickly over every corner, such that all metastable states are visited frequently enough for the reparametrization not to undo the path. However, we may expect that this issue will aggravate as we move to larger systems with more residues and therefore, more path corners. Multiple-walker metadynamics²¹ presents a solution for this. By having several simulations updating the same path and the metadynamics bias, we can increase the probability of all metastable corners being frequently sampled. Additionally, the convergence time is expected to be reduced.

We set up the multiple-walker PMD simulations as follows. For each proline n -mer, we initialize $2(n + 1)$ walkers—two per expected valley in the FES—half of them starting from PPI and half from PPII, with random Maxwell-Boltzmann distributed velocities. Since AmberTools17³⁴ generates only PPII helices, the initial PPI configurations are obtained via SMD simulations (see Subsection 1 of the Appendix). We focus on the counter-clockwise transition for all chain lengths.

For the tetramer, we deposit Gaussians with a height of 0.02 kcal/mol and a width of 0.05 normalized path units

FIG. 3. An evolving PPI-PPII path for a proline tetramer in $\omega_1\omega_2\omega_3$ -space. As the single-walker sampling (black) drifts away from the ($\omega_1 = -90^\circ$, $\omega_2 = -90^\circ$, $\omega_3 = +90^\circ$) metastable state corner, the node reparametrization makes the path tend to a straight line (red) and undoes the previous optimization (green).

every 0.5 ps. For the pentamer, the Gaussian height is reduced to 0.01 kcal/mol and for the hexamer, to 0.0025 kcal/mol. The adjustment in Gaussian amplitude is for two reasons: First, to compensate the increase in total bias after adding more walkers; second, to favor convergence to the minimum energy channel in longer peptides, which present greater path degeneracy. Of course, runs with smaller Gaussians need a longer time to sample the path and compute the free energy along it.

Our PMD results for the proline tetramer are in very good agreement with previous work.^{23-26,32,33} We observe a PPI-to-PPII zipper-like transition starting from the acetylated end and following the order $\omega_1 \rightarrow \omega_2 \rightarrow \omega_3$ (see Fig. 4). This is equivalent to the PPI-to-PPII path with least average work in Ref. 23. The free energy profile shows barriers of around 10 kcal/mol for each rotated dihedral and a free energy difference of 7 kcal/mol between the two stable helical states. The same results were obtained with Gaussian heights of 0.01 and 0.005 kcal/mol. Thus, the zipper-like mechanism starting from the N-terminus appears to be the minimum free energy path for the counter-clockwise rotation.

In Ref. 23, Moradi and co-workers report different least work mechanisms for the PPI-to-PPII and the PPII-to-PPI transition. No evidence of this was observed in our PMD simulations. We hypothesize that the cosine-based CVs' inability to capture the direction of dihedral rotation might hamper the calculation of a reversible transition.²³ As a check, we performed a PMD run for the clockwise transition with a Gaussian height of 0.01 kcal/mol. This resulted in a center-to-ends PPI-to-PPII transition, ordered $\omega_2 \rightarrow \omega_3 \rightarrow \omega_1$ (see the Appendix, Fig. 15), which is consistent with the least work path reported in Ref. 23 for the PPII-to-PPI transition. This provides an explanation for the contradiction in reversibility for the reported result and points to a relation between the order and the direction in which the residues rotate.

For the proline pentamer, we again obtain a zipper-like PPI-to-PPII transition starting from the acetylated end, following $\omega_1 \rightarrow \omega_2 \rightarrow \omega_3 \rightarrow \omega_4$ (see Fig. 5). This mechanism does not

FIG. 4. (a) Adapted PPI-PPII path for a proline tetramer in $\omega_1\omega_2\omega_3$ -space after $t = 30$ ns (black) and sampling density for each of the eight walkers during the last 1 ns (colored dots). (b) Time evolution of the path progress parameter σ for each walker (in colors). (c) Average free energy profile from $t = 5$ ns to $t = 30$ ns.

correspond to the optimal one reported in previous studies²³ (ends-to-center), but it is included within the feasible pathways. The free energy profile also shows the expected barrier heights. When attempting to replicate the result with smaller or less frequent Gaussians, we observed different zipper mechanisms (starting from each end, both ends or the center). Given the path degeneracy for this case, it is not possible to confirm that the algorithm has converged to the (global) minimum free energy path after just one simulation. One could obtain more statistics either by optimizing the path several times using different parameters and initial velocities or by explicitly calculating the free energy along each of the possible $(n - 1)!$ transition paths.

The result is similar for the hexamer. We observe a PPI-to-PPII transition following the order $\omega_1 \rightarrow \omega_2 \rightarrow \omega_3 \rightarrow \omega_4 \rightarrow \omega_5$ starting from the acetylated end (see Fig. 6). The reaction free energy and the free energy barriers are similar

FIG. 5. (a) Adapted PPI-PPII path representation for a proline pentamer in $\omega_1\omega_2\omega_3\omega_4$ -space after $t = 50$ ns. (b) Time evolution of the path progress parameter σ for each of the ten walkers (in colors). (c) Average free energy profile from $t = 5$ ns to $t = 50$ ns.

to the ones in previous studies.^{23,24} Again, when modifying the Gaussian deposition parameters, we observe all different zipper-like mechanisms. The path degeneracy prevents us from finding the optimal path in a single simulation. However, the presented adaptive path framework can be used to extract enough statistics for this, or even longer peptides.

In order to evaluate the path-CV's handling of the increasing $\{\omega_i\}$ -space dimensionality, we analyze the convergence of the PPI-to-PPII path and the free energy profile for the different oligomers. To have comparable results, we run the tetramer and the pentamer with the same multiple-walker metadynamics protocol used for the hexamer (12 walkers, with 0.0025 kcal/mol Gaussian height). The tetramer run produces the same Ace-to-Nme zipper-like mechanism, while the pentamer produces a $\omega_1 \rightarrow \omega_3 \rightarrow \omega_4 \rightarrow \omega_2$ mechanism. We assume that the rate of convergence is independent of the captured pathway. In Fig. 7, we depict the root-mean-square deviation (RMSD) time evolution of the paths with respect to the initial guess (i.e., the linear interpolation) and of the running average free energy profiles with respect to a flat profile at 0 kcal/mol. We stop the runs as soon as the RMSD of the running average free energy profile does not fluctuate more than 0.1 kcal/(mol ns) for a duration of 10 ns. We observe that the convergence time does not increase exponentially—not even linearly—with the dimension of CV-space. We argue that convergence is reached in two stages. In the first stage, the adaptive path searches for

FIG. 6. (a) Adapted PPI-PPII path representation for a proline hexamer in $\omega_1\omega_2\omega_3\omega_4\omega_5$ -space after $t = 70$ ns. (b) Time evolution of the path progress parameter σ for each of the twelve walkers (in colors). (c) Average free energy profile from $t = 5$ ns to $t = 70$ ns.

a transition channel, while at the same time the minima in the free energy profile are filled. This process is mediated by the flexibility of the path, the strength of the bias, the features of the FES, and the dynamics of the system. In the second stage, when the metadynamics walker(s) has (have) visited the full length of the path, the average transition path and the free energy are further converged while sampling the orthogonal degrees of freedom. The complexity for this task depends on the Gaussian size and deposition frequency, as well as on the features of the one-dimensional free energy and the diffusivity of the system along the path. We reason that the same performance persists for higher-dimensional systems and for non-step-wise transitions.

The tunable polypoline system has allowed us to run a comprehensive exercise on PMD. The capabilities of the path-CV to handle increasing dimensionality were analyzed and a two-stage convergence is proposed. Additionally, we demonstrated how explicitly considering multiple CVs can impact the specificity of the sampled configurations, which is a key aspect to tackle more complex systems. Multiple-walker simulations were presented not only as an accelerating tool but also as a relevant complement to sample step-wise transitions while reparametrizing the path. We also observed the challenges associated with path degeneracy as computing each of the $(n - 1)!$ transition channels is a quickly growing task for

FIG. 7. (a) RMSD time evolution of the PPI-to-PPII paths with respect to the initial straight paths. (b) RMSD time evolution of the running average free energy profiles with respect to a flat profile at 0 kcal/mol. (c) Time evolution of the change in the RMSD of the running average free energy profiles per ns.

longer chains. An interesting approach to study such multiple-pathway transitions is to sample in the direction perpendicular to the path to find alternative transition channels.¹²

IV. CASE STUDY II: ACIDITY CONSTANT CALCULATION OF ACETIC ACID IN AQUEOUS SOLUTION

In our second case study, we apply the PMD method to study acid dissociation in aqueous solution. The acid molecule donates a proton to the nearest water molecule, which in turn donates a proton to the next neighboring water, and so forth, following the well-known Grotthuss mechanism.³⁸ Proton abstraction is not only important in acid-base chemistry but also plays an important role in various catalytic reactions. This makes it a very interesting case to investigate using atomistic modeling. Acid dissociation has already been studied theoretically,^{39–41} including that of acetic acid^{42,43} due to its biological and industrial relevance. Here we use PMD to compute the reaction free energy for the deprotonation and, from that, the acidity constant, pK_a . Choosing appropriate CVs and calculating the free energy for acetic acid dissociation is challenging for several reasons. First, due to the Grotthuss mechanism, the atom index of the excess proton changes after every transfer. Second, proton transfer is governed by a complex solvent reorganization. A third challenge is the degeneracy of the

reactant state as the proton can sit at either of the two carboxylic acid oxygens in the reactant state. In the product state, both carboxylic oxygens change their coordination number to zero as the proton is donated to the solvent water molecule. To deal with the proton hopping process, we will introduce a special proton transfer CV. Moreover, proton transfer involves the breaking and forming of covalent bonds, which we will model here with computationally demanding density functional theory based molecular dynamics (DFT-MD) simulations. We use the path-CV to capture additional local solvent changes as discussed in detail hereafter.

The DFT-MD simulations were performed with the combined Gaussian and plane-wave implementation in the QUICKSTEP module⁴⁴ of the CP2K package.⁴⁵ We used the TZV2P Gaussian basis set⁴⁶ and a 300 Ry planewave basis. A high SCF energy cutoff of 5×10^{-7} a.u. was used. We employed the BLYP⁴⁷ exchange-correlation functional in conjunction with Grimme's D3 van der Waals correction.⁴⁸ A special smoothing function was used to smoothen the grid to prevent artifacts leading to frozen atoms on the grid as implemented in the CP2K package. A cubic box of edge length 12.43 Å was used, containing one solute acetic acid and 62 water molecules. Core electrons were represented by the norm conserving pseudopotentials of Goedecker *et al.* (GTH).⁴⁹ The NVT ensemble was used to perform the DFT-MD simulations with a time step of 0.5 fs. A CSVR thermostat³⁷ maintained a constant temperature of 300 K, with a coupling constant of 500 steps. The trajectory was saved every 5 fs for further analysis and visualization. After an initial equilibration of 20 ps, a production run was performed for 100 ps.

For the initial guess path, we take a linear interpolation between the reactant and product states in a three-dimensional CV-space. The first CV (n_{OH}) is the coordination number of carboxylic oxygens with respect to all non-aliphatic H-atoms in the system as defined by Eq. (5) with $x = O$ and $y = H$,

$$n_{xy} = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \frac{1 - (r_{ij}/r_0)^{nm}}{1 - (r_{ij}/r_0)^{mm}}. \quad (5)$$

The second CV (n_{OO}) is the coordination number of the carboxylic oxygens with respect to water oxygen atoms. The third CV, which describes the proton hopping in the solvent, is the distance, r_{PO} , between the proton and the carboxylic C atom, in which the proton position, q_P is calculated using the following equation:

$$q_P = \frac{\sum e^{\gamma n_{OH}} \cdot q_O}{\sum e^{\gamma n_{OH}}}. \quad (6)$$

The position q_P is thus obtained as an exponentially weighted average of all proton acceptor positions, q_O , which are here the water and acid oxygen atoms. With a positive number for the parameter γ , the weighting selects the oxygen(s) with the highest coordination number with respect to H-atoms, n_{OH} . The q_P position thus fluctuates between the central oxygen position in an Eigen (H_3O^+) hydronium ion or in between the two oxygens in a Zundel ($H_5O_2^+$) hydronium complex. This approach has some similarity with an earlier CV for proton transfer proposed by Parrinello *et al.*⁴²

The path is comprised of 30 nodes. The end nodes are fixed. The value of the progress parameter along the path is

FIG. 8. Free energy profile for the acetic acid dissociation reaction, showing three (meta-)stable states.

$\sigma = 0$ in the reactant state, where the proton is on the acid and $\sigma \geq 0.6$ in the product state, where the proton is fully solvated as a hydronium ion. No tube potential is applied. This path is updated every 0.5 ps with a half-life of 0.5 ps. The metadynamics Gaussian potentials have a height of 0.05 kcal/mol and a width of 0.045 in normalized path units and are added every 25 fs. To restrict the proton from interacting with the periodic copies of the carboxylic oxygens, a restraint potential is applied just before the transition to the third hydration shell.

All PMD simulations were performed using the single walker formulation. The initial frame is taken from the end of an equilibration run. Since acetic acid is a weak acid, no deprotonation event was observed during the 20 ps equilibration run. The first deprotonation event after starting the PMD simulation is seen after 3 ps. This leads to the formation of the metastable contact ion pair as shown in Fig. 8 after about 6 ps. Next, several back and forth transitions are observed between the contact ion pair and the protonated acetic acid. After 7 ps, the proton transfers to the second hydration shell. After shuttling numerous times between the first and the second hydration shell, the path is converged after the total simulation time of 20 ps as shown in Fig. 9. The whole free energy landscape is explored within the first 20 ps with several recrossings observed. Since the

FIG. 9. Path evolution for the acetic acid dissociation shown in a two-dimensional projection of the CV-space, starting from a linear initial guess path (black). The path converges in the first 20 ps.

FIG. 10. The evolution of the CVs for the acetic acid dissociation. The three (meta-)stable states can be clearly identified. The color palette shows the value of the proton distance CV (r_{PO}).

half-life is quite short, the path can still evolve over time in the rest of the simulation.

The evolution of the three CVs is shown in Fig. 10. As the reaction proceeds, n_{OH} goes to zero. The r_{PO} distance increases from 1 Å in the undissociated reactant state to 3 Å when the contact ion pair is formed and further increases when the proton hops further into the solvent. The mechanistic details will be further discussed elsewhere as here we focus on the assessment of the convergence of the free energy and the path using PMD for a complex chemical process within the DFT-MD framework.

The free energy profile along the path shows three (meta-)stable states (see Fig. 8). The first minimum belongs to the reactant state, the second minimum corresponds to the formation of the contact ion pair, and the last minimum is associated with the proton solvated in the second or third hydration shell. The barrier height is about 9.0 kcal/mol and the free energy difference between the undissociated acid state and the solvated hydronium ion state is approximately 7.0 kcal/mol. The pK_a of the acetic acid can be obtained from the reaction free energy difference using the following equation:

$$pK_a = \frac{\Delta F}{k_B T \ln[10]}. \quad (7)$$

We obtain a pK_a of 5.1, which matches well with the previous theoretical studies of acetic acid, as well as with the experimental value of 4.75, considering that our DFT-MD approach neglects nuclear quantum effects (in particular, zero-point energy differences). To conclude, we have shown that PMD works well with the DFT-MD simulations. Both the path and the free energy can be converged within the short accessible computational time scale of 100 ps.

V. CASE STUDY III: H_2 FORMATION CATALYZED BY A HYDROGENASE ENZYME

In the third case study, we focus our attention on the molecular insights that can be obtained from the optimized transition path. For that, we consider here a chemical reaction at the active site of the [FeFe] hydrogenase enzyme. This enzyme can reversibly reduce protons to molecular hydrogen with remarkable efficiency. The active site of the hydrogenase enzyme is commonly known as the H-cluster as illustrated in Fig. 11. The H-cluster consists of a $[Fe_4S_4]$ cubane cluster that is connected to a $[Fe_2S_2]$ cluster through a bridging cysteine ligand. The two S atoms of the Fe_2S_2 unit are connected by

FIG. 11. H-cluster and the residues considered in this study. The atoms with the star are fixed at their crystallographic positions. The transfer of protons is shown by brown arrows and the electron transfer from the cubane cluster to the 2Fe subcluster is indicated with orange arrows.

an azadithiolate ligand. The Fe atom connected to the $[\text{Fe}_4\text{S}_4]$ cubane via cysteine is denoted as proximal iron (Fe_p) and the other Fe atom, which has a vacant site, as distal iron (Fe_d). We employ PMD to calculate the free energy for the final hydrogen production reaction step, in which Fe_d is already protonated and the second proton has to transfer from the azadithiolate nitrogen. In addition to rationalizing the reaction mechanism from the resulting average transition path, we will use the path to obtain insight into the changing electronic structure during the reaction.⁵⁰

For the setup of the system, the crystal structure of the H-cluster in the $[\text{FeFe}]$ hydrogenase is taken from *Clostridium pasteurianum* (PDB 3C8Y). The CN^- groups of the H-cluster interact strongly with the residues of the protein (Lys358, Ser232) through hydrogen bonds.^{51,52} This interaction restricts the rotation of the CN ligands and maintains the structural integrity of the H-cluster. We fixed the N atom of the CN^- group at its crystallographic position based on this observation. The cysteine ligands of the Fe_4S_4 clusters are terminated at the CH_3 groups. The total charge of the cluster (-3) is consistent with the previous studies.^{51–53} In the study of the dynamics of the H-cluster, two important residues, Cys299 and H_2O -612, have been included. Studies on proton transfer in the enzyme have shown that these residues play a crucial role in the transport of the proton from the outside of the enzyme to the active site.⁵⁴ The CP2K⁴⁴ program is used for the simulation. DFT-MD simulations are carried out with the Perdew-Burke-Ernzerhof (PBE)⁵⁵ functional and a DZVP-MOLOPT⁴⁶ basis set and a plane wave cutoff of 300 Ry. Valence-core interactions are represented by GTH-type pseudopotentials.^{49,56} Grimme's D3 pseudopotentials⁴⁸ is used to account for dispersion interaction. All calculations are carried out in the NVT ensemble using the CSVR³⁷ thermostat to maintain a temperature of 298.15 K. The Martyna-Tuckerman (MT) Poisson solver⁵⁷ is used to remove spurious periodic interactions from the cubic super cell, which has an edge of 25 Å.

Three CVs are used for the PMD free energy calculations: (1) The coordination number of N with respect to the protons H1 and H2 (n_{NH}). (2) The coordination number of Fe with respect to the protons H1 and H2 (n_{FeH}). (3) The coordination number of H1 and H2 (n_{HH}). The initial path is obtained by a linear interpolation between the reactant and product states. The path consists of 25 nodes between the reactant and product

states and 10 trailing nodes on each side. The path is updated every 0.25 ps and the half-life parameter is 1.0 ps. We used wall potentials with a force constant of 500 kcal/mol per squared normalized path unit on the path progress parameter σ at -0.05 and 1.05 . The initial and final paths are shown in Fig. 12. The positions marked in the path are used for calculating electronic properties (Fig. 14).

The formation of H_2 by the reaction of the terminal hydride and a proton is a downhill process, but it has to overcome a barrier of 7 kcal/mol [Fig. 13(b)]. The optimized path provides the reaction mechanism in terms of the relevant CVs. A straight linear interpolation between the reactant and product states (as we took for the initial guess) would indicate a perfectly concerted mechanism. Figure 12 shows that the final path deviates significantly from the initial straight line. It is interesting to note that the reaction has two overlapping pathways. In one pathway, the proton directly reacts with the hydride to form H_2 . In the other pathway, the proton transfers to the terminal iron first and forms an intermediate structure, where both the hydrogen atoms are on the terminal iron, and then reacts further to form H_2 .

In addition to the CVs used to bias the system, we can also project the trajectory and the path onto other insightful CVs, such as relevant distances. This is illustrated in Fig. 13(c): the distance between N–H2 and Fe–H2 oscillates around a constant value along the path. The Fe–H1 distance decreases once the H1 is transferred from the nitrogen to the Fe atom. Both Fe–H1 and Fe–H2 distances are similar once H1 is transferred to the distal iron.

PMD can be used to investigate electronic properties along the different pathways, which may help in understanding the observed chemical reactivity. The difference of the total electron density of the complex with respect to the separated enzyme complex and a H_2 molecule [$\rho(\text{Hcluster-H}_2) - \rho(\text{Hcluster}) - \rho(\text{H}_2)$] is shown in Fig. 14(a) for selected configurations along the path. The negative (magenta colored) electron density difference located between the H atoms in the reactant state ($\sigma = 0$) shows that initially no covalent bond exists between the hydrogens, whereas the positive (green)

FIG. 12. The final path (orange) for the H_2 formation reaction starting from an initial linearly interpolated path (green). Purple dots show the sampled configurations in the CV-space. The five labeled red points along the transition path are further analyzed (see also Fig. 14).

FIG. 13. (a) Molecular structures of the reactant and product states. (b) Free energy profile for the formation of the H_2 molecule. (c) Selected relevant distances along the path.

FIG. 14. (a) Electron density difference (isocontour value ± 0.02 e upon H_2 formation). Gain in electron density is shown as the green isosurface and loss in the electron density is shown as the magenta isosurface. (b) Wannier center displacement along the path. Selected Wannier centers are shown in orange. Hydrogen atoms are shown in white transparent spheres. Only the Fe_2S_2 subcluster is shown for clarity.

density shows the covalent Fe_d-H_2 and $N-H_1$ bonds. Interestingly, also the Fe_4S_4 cluster exhibits some charge transfer with respect to the empty (i.e., deprotonated) enzyme. Along the reaction path, the $N-H$ bond is broken, the Fe_d-H_2 weakens, and some electron density appears between the H atoms, although significantly less than the reference isolated H_2 molecule.

Figure 14(b) shows selected centers of maximally localized Wannier orbitals⁵⁸ for the same configurations. These Wannier centers are typically localized on chemically intuitive positions, such as at the nuclei, at chemical bonds, or at electron lone pair positions. Here, we see in the reactant state a single Wannier center on the H_2 (rather than on the $Fe-H$ bond), indicating that H_2 is an anionic hydride ligand, whereas the amine group carries three centers, located at the bonds close to the N atom. Note that subsequently, H_1 transfers without a Wannier center, thus being a proton, to H_2 , after which the Wannier center on the hydride moves in between H_1 and H_2 , establishing the bond when the hydrogen molecule forms.

VI. CONCLUSIONS

At the hand of three case studies of complex activated molecular transitions, we have shown that an adaptive path-CV, i.e., a flexible parameterized curve that is a function of a set of descriptive CVs is a powerful means to describe the transition and to enhance the sampling of the process. Combined with metadynamics and given an appropriate set of CVs, the path-CV transforms on-the-fly to the mean transition flux density and is simultaneously used as a reaction coordinate along which the dynamics is biased and the free energy profile of the process is obtained.

By applying the bias onto the path-CV, instead of directly on each of the descriptive CVs in the set, the restriction on the number of CVs, which is about three in “normal” metadynamics, is essentially removed. The first case study, in which we applied PMD to model the conformational switch from a right-handed helix to a left-handed helix in polyproline oligomers, showed that the computational cost to converge the transition path and the free energy profile grows less than linearly with

the number of dihedral angles in the set of CVs. The observed increase in the required simulation time was mainly due to the increasing number of intermediate meta-stable states and barriers that need to be “filled” with metadynamics Gaussians in the first part of the PMD simulation. But once the path-CV has found the reaction mechanism and the full length of the path has been sampled, the second part of the convergence is even less dependent on the dimensionality of the CV set.

The polypyrrole study also showcased the multiple-walker implementation of PMD, in which parallel replicas of the simulation simultaneously sample and update the path-CV, as well as construct the metadynamics bias potential. The multiple-walker parallelization has a trivial scaling with the number of walkers, perhaps only limited by the communication of the transition flux density information, which in the current implementation in the PLUMED software is handled through reading and writing the data files of each replica, but in a later version will use MPI communication. Apart from the computational speedup, the multiple-walker PMD has the advantage that the “corner-cutting” problem due to the reparametrization step in the path updates is largely avoided.

At the DFT-MD level of theory, PMD was able to obtain the reaction mechanism and the associated free energy profile of acetic acid deprotonation in explicit water, presented in the second case-study. It shows that PMD is efficient enough to study complex chemical processes, despite the computationally costly quantum chemical calculation of the electronic structure every MD step, which limits the total simulation time with current high-performance computing (HPC) resources to several tens to hundreds of picoseconds. For this application, different types of CVs were used, two coordination numbers and one distance. Internally, the PMD implementation can use scaling factors to make the CV-space dimensionless and, if needed, stretch or compress the CV-space in certain directions, although here the scaling factors were set to unity.

The third case study focused on the mechanistic insight that can be obtained from the optimized transition path. Here, we performed a PMD simulation of the final chemical reaction step in the hydrogenase enzyme catalyzed production of molecular hydrogen from protons. The resulting path showed that the second proton can either bind directly to the first, iron coordinated hydride or first coordinate to the same iron ion before forming molecular hydrogen. By analyzing the electronic structure along the transition path, here by means of the electronic density difference and the centers of maximally localized Wannier orbitals, we obtain further insight into the reactivity of the enzyme.

ACKNOWLEDGMENTS

A. Pérez de Alba Ortiz thanks the Mexican National Council for Science and Technology (CONACYT) for financial support. This work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative and is part of the Industrial Partnership Programme (IPP) “Computational sciences for energy research” of the NWO-I, which is part of the Netherlands Organisation for Scientific Research (NWO). This research programme is co-financed by Shell Global Solutions International B.V.

APPENDIX: PPI-TO-PPII TRANSITION IN PROLINE OLIGOMERS

1. Steered MD protocol

Steered MD (SMD) was used to obtain PPI structures parting from the default PPII produced by AmberTools17.³⁴ We set straight paths from -1 to 1 in the space of $\{\cos(\omega_i)\}$ with $i = 1, \dots, n - 1$ for each proline n -mer. The path update pace is set to 0.1 ps and the half-life to 0.025 ps. In all cases, the SMD spring constant is of 1000 kcal/mol per squared normalized path unit. We only use a tube potential for the hexamer, with a force constant of 30 kcal/mol. We start the simulation with a 10 ns equilibration, drive the transition in 50 ns, and finish with another 10 ns equilibration and a visual inspection of the structure. It is possible to repeat this process recursively—using optimized paths as initial guesses—to converge a path and a free energy profile.³

FIG. 15. (a) Adapted PPI-PPII (clockwise) path for a proline tetramer in $\omega_1\omega_2\omega_3$ -space after $t = 30$ ns (black) and sampling density for each of the eight walkers during the last 1 ns (colored dots). (b) Time evolution of the path progress parameter σ for each walker (in colors). (c) Average free energy profile from $t = 5$ ns to $t = 30$ ns.

FIG. 16. Adapted PPI-PPII paths for a proline tetramer in $\omega_1\omega_2\omega_3$ -space with varying number of nodes.

2. PPI-to-PPII clockwise transition for the proline tetramer

PMD was applied to the PPI-to-PPII transition via clockwise rotations of the proline tetramer. The metadynamics and path-CV parameters were set as explained in Sec. III. The resulting path, ordered $\omega_2 \rightarrow \omega_3 \rightarrow \omega_1$, corresponds to a center-to-ends PPI-to-PPII transition (Fig. 15). Comparing this to the counter-clockwise Ace-to-Nme PPI-to-PPII transition points to a relation between the direction and the order in which the residues rotate.

3. Selecting the number of nodes

We have no *a priori* rule to find the required number of nodes for a given transition. However, we can identify specific signs of either lack or excess of nodes during simulations. We run multiple-walker simulations for the tetramer transition as described in Sec. III and only change the number of path nodes. Results are shown in Fig. 16. Too many nodes cause the path to “curl” near the stable states, as seen in the 40 nodes per dihedral path, at the $(\omega_1 = +90^\circ, \omega_2 = +90^\circ, \omega_3 = +90^\circ)$ corner. This occurs as numerous nodes start to capture small fluctuations around the FES basins. Too few nodes generate an effect as seen in the minimal nodes path—that is, 1 node per metastable state—at the $(\omega_1 = -90^\circ, \omega_2 = -90^\circ, \omega_3 = +90^\circ)$ corner. As only one node locates the metastable state, any sampling fluctuation from it will cause the path to drift away. It should also be noted that changing the number of nodes leads to a different pathways. This is a consequence of the changes in path flexibility. Still, the algorithm has a degree of robustness as the path with 10 nodes per dihedral produces the same result as the original path with 20 nodes per dihedral.

- ⁷K. J. H. Jonsson and G. Mills, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by F. D. Coker, B. J. Berne, and G. Ciccotti (World Scientific, 1998), p. 385.
- ⁸L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, *J. Chem. Phys.* **125**, 024106 (2006).
- ⁹C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, *J. Chem. Phys.* **108**, 1964 (1998).
- ¹⁰S. Park, M. K. Sener, D. Lu, and K. Schulten, *J. Chem. Phys.* **119**, 1313 (2003).
- ¹¹P. Y. Ayala and H. B. Schlegel, *J. Chem. Phys.* **107**, 375 (1997).
- ¹²D. Branduardi, F. L. Gervasio, and M. Parrinello, *J. Chem. Phys.* **126**, 054103 (2007).
- ¹³A. C. Pan, D. Sezer, and B. Roux, *J. Phys. Chem. B* **112**, 3432 (2008).
- ¹⁴B. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1992).
- ¹⁵Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141–151 (1999).
- ¹⁶L. Maragliano and E. Vanden-Eijnden, *Chem. Phys. Lett.* **426**, 168 (2006).
- ¹⁷B. Ensing, A. Laio, M. Parrinello, and M. L. Klein, *J. Phys. Chem. B* **109**, 6676 (2005).
- ¹⁸B. Ensing and M. L. Klein, *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6755 (2005).
- ¹⁹G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, *Comput. Phys. Commun.* **185**, 604 (2014).
- ²⁰A. Barducci, G. Bussi, and M. Parrinello, *Phys. Rev. Lett.* **100**, 020603 (2008).
- ²¹P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti, and M. Parrinello, *J. Phys. Chem. B* **110**, 3533 (2006).
- ²²G. Ciccotti, M. Ferrario, J. T. Hynes, and R. Kapral, *Chem. Phys.* **129**, 241 (1989).
- ²³M. Moradi, V. Babin, C. Roland, and C. Sagui, *J. Chem. Phys.* **133**, 125104 (2010).
- ²⁴M. Moradi, V. Babin, C. Roland, T. A. Darden, and C. Sagui, *Proc. Natl. Acad. Sci. U. S. A.* **106**, 20746 (2009).
- ²⁵M. Moradi, J. G. Lee, V. Babin, C. Roland, and C. Sagui, *Int. J. Quantum Chem.* **110**, 2865 (2010).
- ²⁶M. Moradi and E. Tajkhorshid, *J. Phys. Chem. Lett.* **4**, 1882 (2013).
- ²⁷M. Moradi, C. Sagui, and C. Roland, *Chem. Phys. Lett.* **518**, 109 (2011).
- ²⁸M. Moradi, C. Sagui, and C. Roland, *J. Chem. Phys.* **140**, 034114 (2014).
- ²⁹M. Moradi, C. Sagui, and C. Roland, *J. Chem. Phys.* **140**, 034115 (2014).
- ³⁰V. Babin, C. Roland, and C. Sagui, *J. Chem. Phys.* **128**, 134101 (2008).
- ³¹V. Babin and C. Sagui, *J. Chem. Phys.* **132**, 104108 (2010).
- ³²L. Shi, A. E. Holliday, H. Shi, F. Zhu, M. A. Ewing, D. H. Russell, and D. E. Clemmer, *J. Am. Chem. Soc.* **136**, 12702 (2014).
- ³³L. Shi, A. E. Holliday, B. C. Bohrer, D. Kim, K. A. Servage, D. H. Russell, and D. E. Clemmer, *J. Am. Soc. Mass Spectrom.* **27**, 1037 (2016).
- ³⁴D. A. Case, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, D. M. York, and P. A. Kollman, AMBER, 2017.
- ³⁵H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, *Comput. Phys. Commun.* **91**, 43 (1995).
- ³⁶V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, *Proteins: Struct., Funct., Bioinf.* **65**, 712 (2006).
- ³⁷G. Bussi, D. Donadio, and M. Parrinello, *J. Chem. Phys.* **126**, 014101 (2007).
- ³⁸S. Cukierman, *Biochim. Biophys. Acta, Bioenerg.* **1757**, 876 (2006).
- ³⁹B. L. Trout and M. Parrinello, *Chem. Phys. Lett.* **288**, 343 (1998).
- ⁴⁰M. Sprik, *Chem. Phys.* **258**, 139 (2000).
- ⁴¹P. L. Geissler, C. Dellago, D. Chandler, J. Hutter, and M. Parrinello, *Science* **291**, 2121 (2001).
- ⁴²J. M. Park, A. Laio, M. Iannuzzi, and M. Parrinello, *J. Am. Chem. Soc.* **128**, 11318 (2006).
- ⁴³Z. Liu, C. Siu, and S. T. John, *Chem. Phys. Lett.* **314**, 317 (1999).
- ⁴⁴J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, *Comput. Phys. Commun.* **167**, 103 (2005).
- ⁴⁵The CP2K developers group, <http://www.cp2k.org>.
- ⁴⁶J. VandeVondele and J. Hutter, *J. Chem. Phys.* **127**, 114105 (2007).
- ⁴⁷C. T. Lee, W. T. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- ⁴⁸S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, *J. Comput. Phys.* **132**, 154104 (2010).
- ⁴⁹S. Goedecker, M. Teter, and J. Hutter, *Phys. Rev. B* **54**, 1703 (1996).
- ⁵⁰L. Liu, B. Lukose, and B. Ensing, *J. Phys. Chem. C* **121**, 2046 (2017).
- ⁵¹A. R. Finkelmann, M. T. Stiebritz, and M. Reiher, *Chem. Sci.* **5**, 215 (2014).

¹G. Díaz Leines and B. Ensing, *Phys. Rev. Lett.* **109**, 020601 (2012).

²G. M. Torrie and J. P. Valleau, *J. Comput. Phys.* **23**, 187 (1977).

³C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).

⁴A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12562 (2002).

⁵E. A. Carter and G. Ciccotti, *Chem. Phys. Lett.* **156**, 472 (1989).

⁶D. Frenkel and B. Smith, *Understanding Molecular Simulations: From Algorithms to Applications* (Academic Press, 2002).

- ⁵²M. Bruschi, C. Greco, M. Kaukonen, P. Fantucci, U. Ryde, and L. De Gioia, *Angew. Chem., Int. Ed.* **48**, 3503 (2009).
- ⁵³S. Mebs, R. Kositzki, J. Duan, M. Senger, F. Wittkamp, U.-P. Apfel, T. Happe, S. T. Stripp, M. Winkler, and M. Haumann, *Biochim. Biophys. Acta, Bioenerg.* **1859**, 28 (2017).
- ⁵⁴B. Ginovska-Pangovska, M.-H. Ho, J. C. Linehan, Y. Cheng, M. Dupuis, S. Rauegi, and W. J. Shaw, *Biochim. Biophys. Acta, Bioenerg.* **1837**, 131 (2014).
- ⁵⁵J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- ⁵⁶C. Hartwigsen, S. Gødecke, and J. Hutter, *Phys. Rev. B* **58**, 3641 (1998).
- ⁵⁷G. J. Martyna and M. E. Tuckerman, *J. Chem. Phys.* **110**, 2810 (1999).
- ⁵⁸N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, *Rev. Mod. Phys.* **84**, 1419 (2012).