Traversing the free-energy pathways of intricate biomolecular processes

Enhanced simulation development and applications

Pérez de Alba Ortíz, A.

Publication date
2021

4

The zipper-like helix handedness transition of polyproline chains

Caminante, no hay camino,
se hace camino al andar.

Antonio Machado y Joan Manuel Serrat

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 (PMD) enhanced sampling approach in combination with force-field MD simulations of conformational 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.

Based on: A. Pérez de Alba Ortíz, A. Tiwari, R.C. Puthenkalathil, and B. Ensing. *Advances in enhanced sampling along adaptive paths of collective variables.* J. Chem. Phys. **149**, 072320 (2018).

4.1. Introduction

Path-metadynamics (PMD) [1] 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. [2–6]), (2) those that focus on locating reactive trajectories or optimized transition paths (e.g. [7–13]), and (3) those that enhance the dynamics by raising the temperature of (part of) the system (e.g. [14–16]). PMD combines ideas from the first two classes.

The metadynamics method [4] belongs to the first class of enhanced sampling methods. During an MD simulation, it constructs a repulsive bias potential that "fills" free-energy minima, such that barriers in the free energy can be overcome and a free-energy profile 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 simplest of applications [18], 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 parametrized 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 PLUMED software [19], so that 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 an illustrative example.

In our 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.

Two additional case studies are presented in the publication on which this chapter is based [20]. In these examples, PMD is applied together with ab initio MD, based on density functional theory (DFT), to sample chemical transitions. In the first of these cases, we investigate the deprotonation of acetic acid in aqueous solution. PMD allows to converge a reaction mechanism and a free-energy profile in only a few tens of ps of simulation time. This protocol has been tested with other acids in [21]. In the second study, we focus on the formation of molecular hydrogen by proton reduction, catalyzed by a di-iron hydrogenase enzyme. In this case, apart from obtaining detailed mechanistic understanding and a free-energy profile, the path can be used to scrutinize the changes in the electronic structure during the reaction.

In the following, we present the application of multiple-walker PMD to the helix handedness transition of polyproline chains with increasing length. We rely on the PMD methodology as introduced in Chapter 2. After defining our simulation protocol, we report the results and conclusions derived from the case study.

4.2. System

Polyproline chains, Ace–$(\text{Pro})_n$–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 ($\phi = -75°, \psi = 160°, \omega = 0°$); and polyproline type II (PPII), a left-handed helix with all residues in the *trans* form and backbone dihedral angles ($\phi = -75°, \psi = 146°, \omega = \pm 180°$) [22]. 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 $\{\omega_i\}$ with $i = 1, ..., n-1$. 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 signaling [23], and their *cis-trans* isomerization is a rate-determining step in several protein folding processes [22]. 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 coworkers [22, 24–29] 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 (FESs) on such CVs. SMD was used to compute the work along different candidate pathways on the calculated surfaces, from which the transition

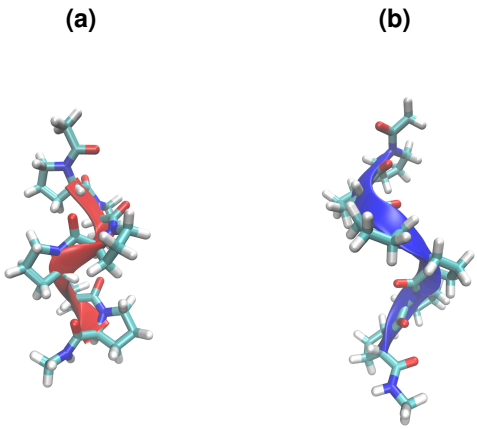

Figure 4.1: : (a) PPI right-handed helix structure with backbone dihedral angles ($\phi = -75°, \psi = 160°, \omega = 0°$). (b) PPII left-handed helix structure with backbone dihedral angles ($\phi = -75°, \psi = 146°, \omega = \pm 180°$)

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; 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 coworkers [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.

4.3. Simulation protocol

4.3.1. Molecular dynamics

In our study, we construct proline oligomers using AmberTools17 [34], which produces PPII structures. Simulations are carried out using GROMACS 5.1.4 [35] patched with PLUMED 2.3.0 [19] with the added path-metadynamics code available on: `http://www.acmm.nl/ensing/software/PathCV.cpp`. We use the AMBER force field ff99SB [36] and no solvent. The time step size is 1 fs and a canonical

sampling through velocity rescaling (CSVR) thermostat [37] is set at 300 K with a time constant of 0.1 ps.

4.3.2. Path-metadynamics

The path for each proline n-mer is defined in a CV-space composed of all dihedrals $\{\omega_i\}$ with $i = 1, ..., n-1$. 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 $\{\omega_i\} = -90°$ to PPII $\{\omega_i\} = +90°$. For the counter-clockwise rotation (from 0° to ±180° through negative angle values) we add +90° to all $\{\omega_i\}$ and map the transition from PPI $\{\omega_i\} = +90°$ to PPII $\{\omega_i\} = -90°$. Whenever a particular transition direction is being sampled, the other one is blocked by harmonic walls with force constants of 1000 kcal/(mol rad^2) which restrain the sampling inside the $[-172°, +172°]$ 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 provided results consistent with previous literature [22, 24–26, 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 $s = 0$) to PPII ($s = 1$). 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 rad^2). Harmonic walls with force constants of 1000 kcal/mol per squared normalized path unit are set on s 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 s 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 in 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. 4.2). 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 metadynamics [38] 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

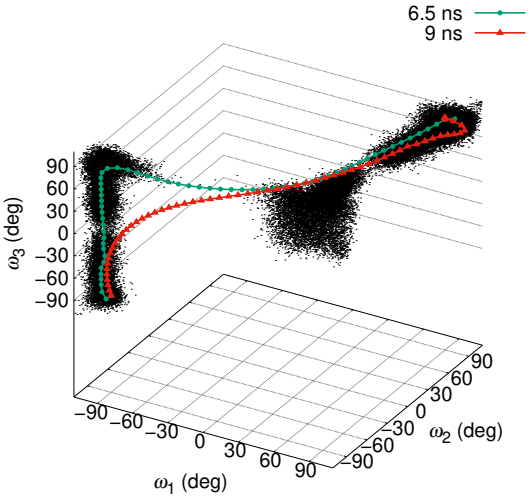

Figure 4.2: An evolving PPI-PPII path for a proline tetramer in $\omega_1\omega_2\omega_3$-space. As the single-walker sampling (black) drifts away from the $(\omega_1 = -90°, \omega_2 = -90°, \omega_3 = +90°)$ metastable state corner, the node reparametrization makes the path tend to a straight line (red) and undoes the previous optimization (green).

time is expected to be reduced.

4.3.3. Multiple-walker path-metadynamics

We setup 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 AmberTools17 [34] generates only PPII helices, the initial PPI configurations are obtained via SMD simulations (See Section 4.A). 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 every 0.5 ps. For the pentamer, the Gaussian height is reduced to 0.01 kcal/mol, and for the hexamer, to 0.025 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.

Data and example PLUMED input files required to reproduce the results reported in this chapter are available on PLUMED-NEST (`www.plumed-nest.org`), the public repository of the PLUMED consortium [39], as plumID:19.033.

4.4. Results and discussion

Our PMD results for the proline tetramer are in very good agreement with previous work [22, 24–26, 32, 33]. We observe a PPI-to-PPII zipper-like transition starting from the acetylated end following the order $\omega_1 \rightarrow \omega_2 \rightarrow \omega_3$ (See Fig. 4.3). This is equivalent to the PPI-to-PPII path with least average work in [24]. 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 MFEP for the counter-clockwise rotation.

In [24], Moradi and coworkers 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 [24]. As a check, we performed a PMD run for the clockwise transition with a Gaussian height of 0.1 kcal/mol. This resulted in a center-to-ends PPI-to-PPII transition, ordered $\omega_2 \rightarrow \omega_3 \rightarrow \omega_1$ (See Section 4.A, Fig. 4.A.1), which is consistent with the least-work path reported in Ref. [24] 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. 4.4). This mechanism does not correspond to the optimal one reported in previous studies [24] (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 and the center). Given the path degeneracy for this case, it is not possible to confirm that the algorithm has converged to the (global) MFEP 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. 4.5). The free-energy difference and the free-energy barriers are similar to the ones in previous studies [22, 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 PMD protocol used for the hex-

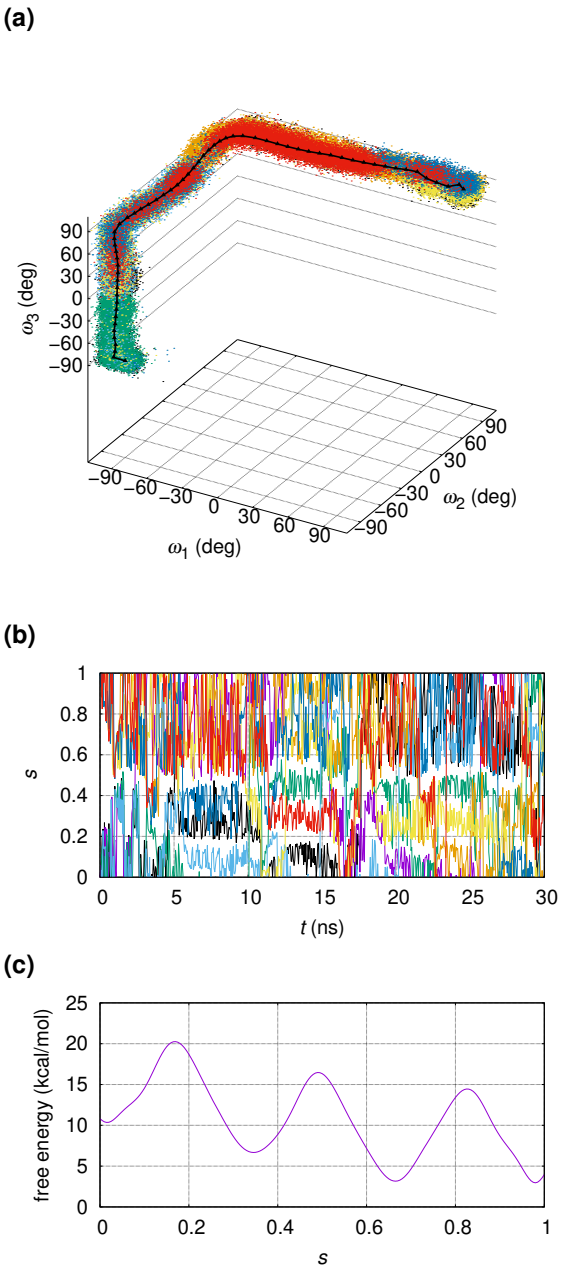

Figure 4.3: (a) Adapted PPI-PPII path for a proline tetramer in $\omega_1\omega_2\omega_3$-space after $t = 30$ ns (black) and sampling density for each of the eight walkers during the last 1 ns (colored dots). (b) Time evolution of the path progress parameter s for each walker (in colors). (c) Average free-energy profile from $t = 5$ ns to $t = 30$ ns.

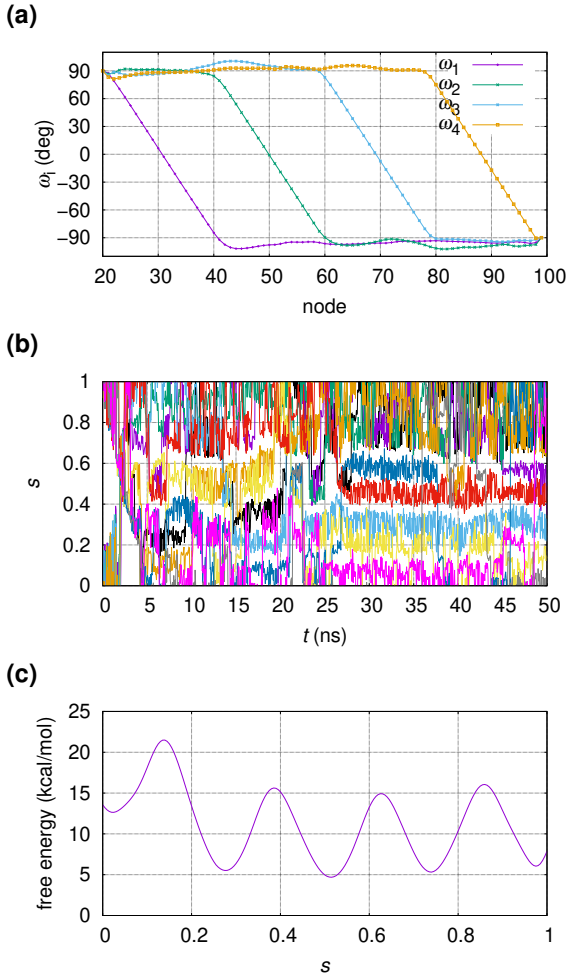

Figure 4.4: (a) Adapted PPI-PPII path representation for a proline pentamer in $\omega_1\omega_2\omega_3\omega_4$-space after $t = 50$ ns. (b) Time evolution of the path progress parameter s for each of the ten walkers (in colors). (c) Average free-energy profile from $t = 5$ ns to $t = 50$ ns.

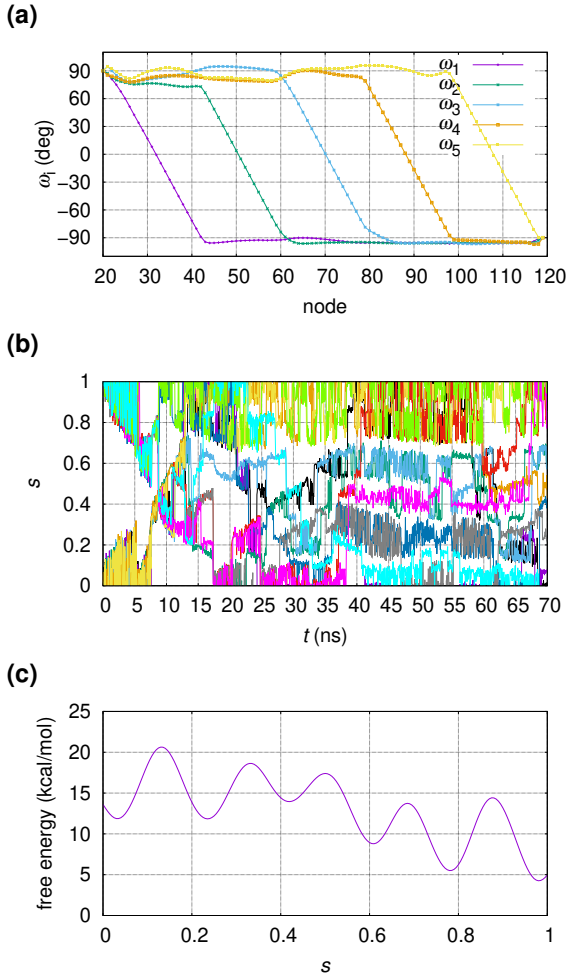

Figure 4.5: (a) Adapted PPI-PPII path representation for a proline hexamer in $\omega_1\omega_2\omega_3\omega_4\omega_5$-space after $t = 70$ ns. (b) Time evolution of the path progress parameter s for each of the twelve walkers (in colors). (c) Average free-energy profile from $t = 5$ ns to $t = 70$ ns.

amer (12 walkers, with 0.025 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. 4.6 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 per 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 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 polyproline system has allowed us to run a comprehensive exercise on path-metadynamics. 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 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], or to use multiple paths as done in Chapter 5.

4.5. Conclusions

At the hand of a case study of a complex activated molecular transition, we have shown that an adaptive path-CV, i.e. a flexible parametrized curve that is a function of a set of descriptive collective variables, 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 are 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 case study, in which we

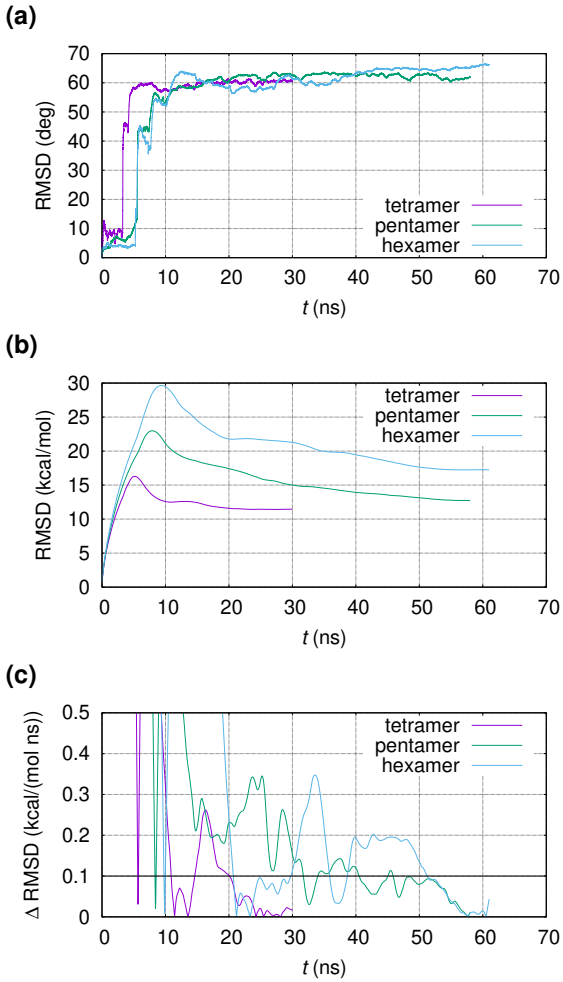

Figure 4.6: (a) RMSD time evolution of the PPI-to-PPII paths with respect to the initial straight paths. (b) RMSD time evolution of the running average free-energy profiles with respect to a flat profile at 0 kcal/mol. (c) Time evolution of the change in the RMSD of the running average free-energy profiles per ns.

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 cost-efficiency of PMD can be particularly beneficial in MD simulations where the calculation of each time step is highly expensive, i.e. when using DFT to calculate forces. For example, PMD has been successfully applied to chemical transitions such as the deprotonation of acetic acid in explicit water, or the production of molecular hydrogen from protons in the hydrogenase enzyme [20].

The polyproline 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 can potentially use OpenMP/MPI communication. Apart from the computational speed-up, the multiple-walker PMD has the advantage that the "corner-cutting" problem due to the reparametrization step in the path updates is largely avoided.

Finally, the polyproline study also raises the challenge of path degeneracy, i.e. having several feasible mechanisms with similar free-energy barriers connecting two states. The original PMD algorithm is designed to converge a single average transition path or MFEP. In landscapes with multiple transition channels, PMD is bounded to the closest low free-energy path to the initial guess. If the landscape is very diffusive, convergence can be elusive if the transition constantly switches from one path to another. In Chapter 5, we propose an extension of PMD to handle multiple transition pathways.

4.A. Appendix

4.A.1. Path-steered molecular dynamics

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 $i = 1, ..., n-1$ 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].

4.A.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 where set as explained in Section 4.4. The resulting path, ordered $\omega_2 \to \omega_3 \to \omega_1$, corresponds to a center-to-ends PPI-to-PPII transition (Fig. 4.A.1). 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.

4.A.3. Selecting the number of path 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 Section 4.4, and only change the number of path nodes. Results are shown in Fig. 4.A.2. Too many nodes cause the path to "curl" near the stable states, as seen in the 40 nodes per dihedral path, at the $(\omega_1 = +90°, \omega_2 = +90°, \omega_3 = +90°)$ 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 $(\omega_1 = -90°, \omega_2 = -90°, \omega_3 = +90°)$ 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 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.

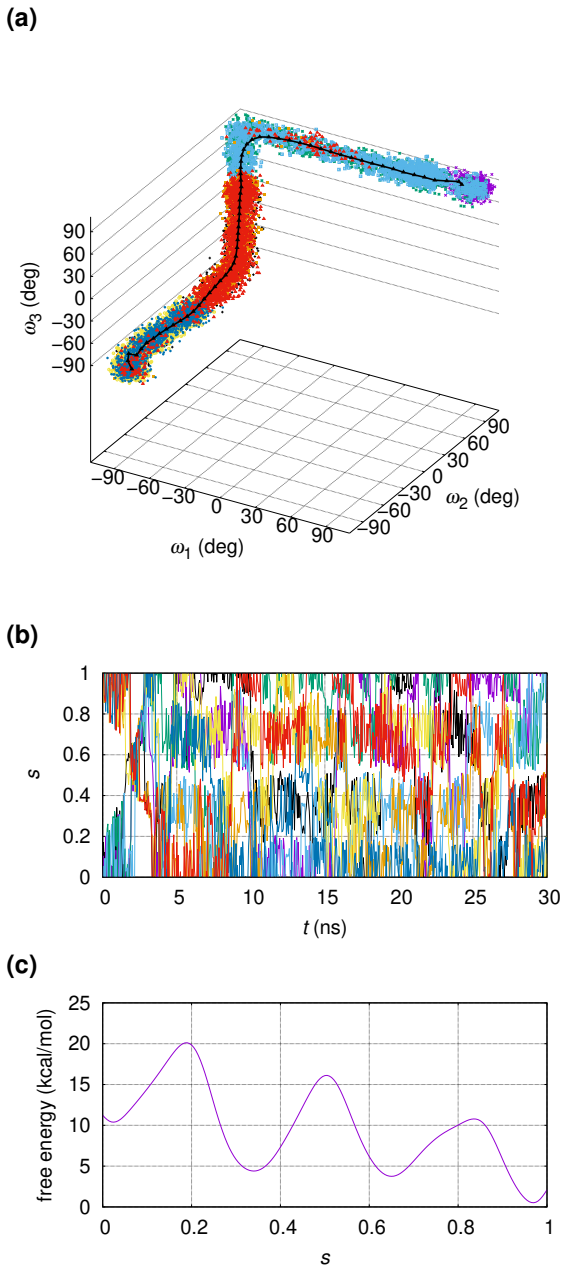

Figure 4.A.1: (a) Adapted PPI-PPII (clockwise) path for a proline tetramer in $\omega_1\omega_2\omega_3$-space after $t = 30$ ns (black) and sampling density for each of the eight walkers during the last 1 ns (colored dots). (b) Time evolution of the path progress parameter s for each walker (in colors).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

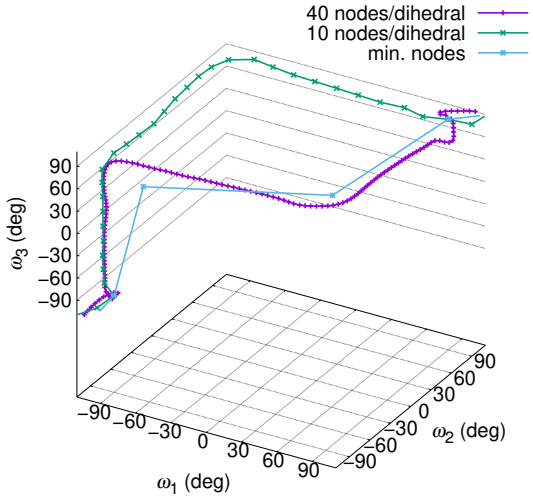

Figure 4.A.2: Adapted PPI-PPII paths for a proline tetramer in $\omega_1\omega_2\omega_3$-space with varying number of nodes.

References

[1] G. Díaz Leines and B. Ensing, *Path finding on high-dimensional free energy landscapes,* Phys. Rev. Lett **109**, 020601 (2012).

[2] G. M. Torrie and J. P. Valleau, *Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling,* J. Comput. Phys. **23**, 187 (1977).

[3] C. Jarzynski, *Nonequilibrium equality for free energy differences,* Phys. Rev. Lett **78**, 2690 (1997).

[4] A. Laio and M. Parrinello, *Escaping free-energy minima,* Proc. Natl. Acad. Sci. U.S.A. **99**, 12562 (2002).

[5] E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, *Constrained reaction coordinate dynamics for the simulation of rare events,* Chem. Phys. Lett **156**, 472 (1989).

[6] D. Frenkel and B. Smit, *Understanding Molecular Simulations: From algorithms to applications* (Academic Press, 2002).

[7] H. Jónsson, G. Mills, and K. W. Jacobsen, *Nudged elastic band method for finding minimum energy paths of transitions,* in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. Berne, G. Ciccotti, and D. F. Coker (World Scientific, 1998) pp. 385–404.

[8] L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, *String method in collective variables: Minimum free energy paths and isocommittor surfaces,* J. Chem. Phys. **125**, 024106 (2006).

[9] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, *Transition path sampling and the calculation of rate constants,* J. Chem. Phys. **108**, 1964 (1998).

[10] S. Park, M. K. Sener, D. Lu, and K. Schulten, *Reaction paths based on mean first-passage times,* J. Chem. Phys. **119**, 1313 (2003).

[11] P. Y. Ayala and H. B. Schlegel, *A combined method for determining reaction paths, minima, and transition state geometries,* J. Chem. Phys. **107**, 375 (1997).

[12] D. Branduardi, F. L. Gervasio, and M. Parrinello, *From A to B in free energy space,* J. Chem. Phys. **126**, 054103 (2007).

[13] A. C. Pan, D. Sezer, and B. Roux, *Finding transition pathways using the string method with swarms of trajectories,* J. Phys. Chem. B **112**, 3432 (2008).

[14] B. Berg and T. Neuhaus, *Multicanonical ensemble: A new approach to simulate first-order phase transitions,* Phys. Rev. Lett. **68**, 9 (1992).

[15] Y. Sugita and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding,* Chem. Phys. Lett. **314**, 141–151 (1999).

[16] L. Maragliano and E. Vanden-Eijnden, *A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations,* Chem. Phys. Lett. **426**, 168 (2006).

[17] B. Ensing, A. Laio, M. Parrinello, and M. L. Klein, *A recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions,* J. Phys. Chem. B **109**, 6676 (2005).

[18] B. Ensing and M. L. Klein, *Perspective on the reactions between F- and CH3CH2F: the free energy landscape of the E2 and SN2 reaction channels.* Proc. Natl. Acad. Sci. USA **102**, 6755 (2005).

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

[20] A. Pérez de Alba Ortíz, A. Tiwari, R. Puthenkalathil, and B. Ensing, *Advances in enhanced sampling along adaptive paths of collective variables,* J. Chem. Phys. **149**, 072320 (2018).

[21] A. Tiwari, *Toward accurate simulation of electrocatalyzed water splitting: Enhanced quantum chemical dynamics simulations of proton and electron transfer reactions*, Ph.D. thesis, University of Amsterdam (2019).

[22] M. Moradi, V. Babin, C. Roland, T. A. Darden, and C. Sagui, *Conformations and free energy landscapes of polyproline peptides,* Proc. Natl. Acad. Sci. USA **106**, 20746 (2009).

[23] T. J. Narwani, H. Santuz, N. Shinada, A. M. Vattekatte, Y. Ghouzam, N. Srinivasan, J.-C. Gelly, and A. G. de Brevern, *Recent advances on polyproline II,* J. Amino Acids **49**, 705 (2017).

[24] M. Moradi, V. Babin, C. Roland, and C. Sagui, *A classical molecular dynamics investigation of the free energy and structure of short polyproline conformers,* J. Chem. Phys. **133**, 125104 (2010).

[25] M. Moradi, J. G. Lee, V. Babin, C. Roland, and C. Sagui, *Free energy and structure of polyproline peptides: an ab initio and classical molecular dynamics investigation,* Int. J. Quantum Chem. **110**, 2865 (2010).

[26] M. Moradi and E. Tajkhorshid, *Driven metadynamics: Reconstructing equilibrium free energies from driven adaptive-bias simulations,* J. Phys. Chem. Lett. **4**, 1882 (2013).

[27] M. Moradi, C. Sagui, and C. Roland, *Calculating relative transition rates with driven nonequilibrium simulations,* Chem. Phys. Lett. **518**, 109 (2011).

[28] M. Moradi, C. Sagui, and C. Roland, *Investigating rare events with nonequilibrium work measurements. I. Nonequilibrium transition path probabilities,* J. Chem. Phys. **140**, 034114 (2014).

[29] M. Moradi, C. Sagui, and C. Roland, *Investigating rare events with nonequilibrium work measurements. II. Transition and reaction rates,* J. Chem. Phys. **140**, 034115 (2014).

[30] V. Babin, C. Roland, and C. Sagui, *Adaptively biased molecular dynamics for free energy calculations,* J. Chem. Phys. **128**, 134101 (2008).

[31] V. Babin and C. Sagui, *Conformational free energies of methyl-α-l-iduronic and methyl-β-d-glucuronic acids in water,* J. Chem. Phys. **132**, 104108 (2010).

[32] L. Shi, A. E. Holliday, H. Shi, F. Zhu, M. A. Ewing, D. H. Russell, and D. E. Clemmer, *Characterizing intermediates along the transition from polyproline I to polyproline II using ion mobility spectrometry-mass spectrometry,* J. Am. Chem. Soc. **136**, 12702 (2014).

[33] L. Shi, A. E. Holliday, B. C. Bohrer, D. Kim, K. A. Servage, D. H. Russell, and D. E. Clemmer, *"wet" versus "dry" folding of polyproline,* J. Am. Soc. Mass Spectrom. **27**, 1037 (2016).

[34] D. 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 (2017).

[35] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, *GROMACS: a message-passing parallel molecular dynamics implementation,* Comput. Phys. Commun. **91**, 43 (1995).

[36] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, *Comparison of multiple Amber force fields and development of improved protein backbone parameters,* Proteins **65**, 712 (2006).

[37] G. Bussi, D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling,* J. Chem. Phys. **126**, 014101 (2007).

[38] P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti, and M. Parrinello, *Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics,* J. Phys. Chem. B **110**, 3533 (2006).

[39] M. Bonomi, G. Bussi, C. Camilloni, G. A. Tribello, P. Banáš, A. Barducci, M. Bernetti, P. G. Bolhuis, S. Bottaro, D. Branduardi, *et al.*, *Promoting transparency and reproducibility in enhanced molecular simulations,* Nat. Methods **16**, 670 (2019).